

Link to the manuscript pdf: [akanna15.pdf](#)

## Response Letter

Dear Editor and Reviewers,

Thank you for taking the time to review the proposal. Here is a list of our responses to your feedback and how we will address it.

[R1]: [https://drive.google.com/file/d/1iW8f7YH6SrHZ-KHI-0jcw-pfQJKsKX15/view?usp=drive\\_link](https://drive.google.com/file/d/1iW8f7YH6SrHZ-KHI-0jcw-pfQJKsKX15/view?usp=drive_link)

[R1.1]: PI clearly states they will compare their results with "known fire events". This is good empirical evidence to have, but how will you measure similarity? Who produced the "known fire events" dataset - is it public and trusted?

**Response:** The "known fire events" will be coming from historical data sources such as MODIS (NASA), FIRMS (Fire Information for Resource Management System), or the Global Fire Atlas — which provide time-stamped satellite detections and burn perimeters. In order to measure similarity, we are considering using spatiotemporal metrics, such as using visual similarity score using SSIM (Structural Similarity Index) if using video generation outputs.

[R2]: [https://drive.google.com/file/d/1mMFNtdojtoKiSbmStxfh-0sOZsnoTAZZ/view?usp=drive\\_link](https://drive.google.com/file/d/1mMFNtdojtoKiSbmStxfh-0sOZsnoTAZZ/view?usp=drive_link)

[R2.1]: The proposal is scientifically grounded, with strong motivation for incorporating oxygen as a variable and a solid literature foundation. However, details on evaluating predictive performance (e.g., what constitutes "good" performance, which fire events will serve as validation, and what kind of error analysis will be conducted) are not fully developed, which limits the rigor.

**Response:** Predictive performance will be evaluated using spatial overlap (IoU), spread rate error (<10% deviation), and video similarity (SSIM/FVD) metrics. Error analysis will involve sensitivity tests to identify which environmental variables most strongly affect predictive deviation

[R2.2]: The proposal's timeline is ambitious if the author pursues a full 3D simulation.

**Response:** The reviewer's concerns are valid and the ambitiousness of the project is a factor we have considered as well. In order to stay grounded, the project will first implement a 2D spatiotemporal simulation to validate model performance. The 3D video generation phase will build on pre-trained architectures (e.g., Video Diffusion Models, NeRF-based methods).

[R3]: [https://drive.google.com/file/d/1hVcB6HemR5TrB2FwXC\\_p4QpsnV\\_QZeGa/view?usp=drive\\_link](https://drive.google.com/file/d/1hVcB6HemR5TrB2FwXC_p4QpsnV_QZeGa/view?usp=drive_link)

[R3.1]: The plan mentions collecting historical data in Weeks 1-2, but doesn't state whether this data has already been located or if access has been secured, which could be a potential bottleneck

**Response:** Our project looks to avoid these bottlenecks by using open-source datasets, including NASA FIRMS for fire events and ERA5/NOAA reanalysis data for atmospheric and oxygen parameters. Data retrieval pipelines have been identified and will be implemented in Python to ensure timely access during Week 1.

[R4]: [https://drive.google.com/file/d/1qLSWyhS03RRuEeJxLiwg37LVcn1ZPzSI/view?usp=drive\\_link](https://drive.google.com/file/d/1qLSWyhS03RRuEeJxLiwg37LVcn1ZPzSI/view?usp=drive_link)

[R4.1]: The proposal lays out how it will use oxygen data, but is there anything visually novel about its output? In 2D, what would this output look like? Would it be overlaid on a map? The proposal also mentions 3D simulations. How would these simulations be rendered?

**Response:** For the 2D visualization, the outputs will be overlaid on geospatial maps with heat-based color coding of oxygen and temperature fields. The novel 3D representation will have the generative video model will produce volumetric fire spread animations, with interpretable visuals of the ecological factors

[R4.2]: How would this tool be evaluated? How could it be compared with past fire events? Is there a way to quantify improvements in user experience?

**Response:** We will be conducting a preliminary evaluation with environmental scientists to assess visual clarity, interpretability, and usability. The users will rate the system on visual realism, data transparency, and ease of pattern recognition, providing qualitative feedback for iterative refinement.



# Visual Analysis of an Ensemble Forest Fire Simulations

Aditi Kannan

PI

aditi\_kannan@brown.edu

Olivia Heng

Co-PI

olivia\_heng@brown.edu

Divya Banesh

Collaborator

dbanesh@lanl.gov

Jim Ahrens

Collaborator

ahrens@lanl.gov

October 14, 2025

## Abstract

Forest fires can be extremely complex and difficult to simulate. In this project, we propose a novel method to extract climate patterns and oxygen interactions with climate factors from historical data, and create a video generation tool to simulate future conditions which may lead to fires. Our work links climate–oxygen interactions to predictive fire spread modelling, while providing intuitive dynamic visualizations.

# 1 Aims

Forest fires are highly dynamic and influenced by complex interactions between climate, ecology, and atmospheric conditions. The aim of this project is to develop a visual analysis framework that links historical climate trends with oxygen availability and other ecological factors to improve fire spread simulations. The framework will generate dynamic video-based visualizations of simulated fire events, enabling interpretable spatiotemporal understanding and supporting proactive management strategies.

The aim of this project is to develop a deep learning-based 3D video generation framework for forest fire simulations that explicitly integrates oxygen flux / atmospheric oxygen concentration and ecologically-relevant microclimatic and soil variables, to improve both the predictive accuracy and interpretability of fire spread models. Specifically, the project will:

- Extract climate, atmospheric (especially oxygen), ecological, and soil-moisture patterns from historical data, quantifying how oxygen availability (e.g. oxygen flux or concentration) modulates burning dynamics in forest fuel beds, building on experimental work by Thomas et al. (“Experimental investigation of the impact of oxygen flux. . .”).
- Condition simulation models on those extracted patterns, then translate outputs into 3D dynamic videos via deep learning (e.g. conditional video/diffusion/GANs or neural rendering) that reveal spatiotemporal fire spread under different oxygen/climate/ecological regimes.
- Identify and rank which variables (oxygen flux, microclimate, soil moisture, fuel properties, etc.) most strongly drive fire behaviour, especially under non-standard atmospheric oxygen conditions (e.g. higher oxygen concentrations), filling gaps where current fire models typically assume “normal” atmospheric oxygen or neglect oxygen as an explicit input.
- Use the visualized simulations to support testing of intervention strategies (e.g. modifying fuel moisture, controlling oxygen flux, etc.), providing a tool for exploring “what if” scenarios under future climate or atmospheric oxygen shifts.

The oxygen flux- fire relationship is a gap in which current research hasn’t filled. It is important to explore this angle for several reasons:

- The Thomas et al. experiments show that oxygen flux has strong effects on burning dynamics: increasing O flux increases peak heat release rate, reduces flaming duration, changes combustion efficiency and CO/CO ratios. This implies that oxygen is a controlling variable in how fires burn, not just temperature / moisture / wind.
- Vitali et al. show that under elevated atmospheric oxygen concentration, fire activity increases, but forests may still persist under certain ecological feedback. That suggests that future atmospheric conditions (or microenvironments with altered oxygen) could shift fire behavior significantly.
- Yet most simulation and visualization models ignore oxygen (or assume standard atmospheric O), which may lead to underestimation or mis-estimation of fire spread, heat and burning rates, or mis-ranking of feature importance.

So, investigating oxygen (both concentration and flux) along with climate and ecological variables fills a systemic gap: it may improve both the physics of simulation (make them more accurate under broader conditions) and the interpretability of models (which factors are really crucial under different oxygen regimes).

## 2 Significance

Forest fires are among the most destructive natural hazards, yet current simulation models often oversimplify or exclude critical drivers such as microclimatic conditions, oxygen interactions, and soil moisture. By systematically identifying and ranking the most influential variables and translating these insights into a visualization pipeline, this project offers both scientific significance—improving predictive accuracy and interpretability—and practical significance—providing decision-makers with tools to test ecological interventions virtually.

**Intellectual Merit:** The project introduces a novel coupling of climate–oxygen interactions with fire simulations, supported by data-driven feature importance analysis. It also advances methodological frontiers by adapting video generation models for spatiotemporal ecological visualization, a task not yet widely attempted.

**Broader Impact:** Beyond forest fire research, the proposed pipeline has potential for generalizable use in other climate hazard domains (e.g., flood or hurricane modeling). It can improve public communication of fire risk through accessible visualizations and provide educational value in climate science courses. Ultimately, it contributes to proactive fire prevention and mitigation strategies that safeguard ecosystems and communities.

Additionally, dynamic, deep learning-based 3D video generation conditioned on those interactions will enhance interpretability: stakeholders can see how changes in oxygen along with climate or fuel moisture affect spread patterns, fire front intensities, etc., rather than just numeric outputs. This has both scientific impact (improving model fidelity, new knowledge about what drives fire dynamics) and practical impact (decision support, policy, forest management under changing atmospheric/climatic conditions). This project is highly interdisciplinary, lying at the intersection of computer science, environmental science, and climate studies. By integrating deep learning models with ecological simulation data, we bridge technical advances in artificial intelligence with pressing societal challenges in fire ecology and disaster management. The proposed visualization pipeline also overlaps with human–computer interaction and science communication, broadening its impact across disciplines.

The evaluation methods and data collection methods are also significant contributions and important steps in the process.

This project relies on the integration of observational, experimental, and simulated data sources.

Key datasets will include:

- **Climate and Atmospheric Data:** Historical climate variables (temperature, humidity, wind speed, precipitation) will be obtained from NOAA, ERA5, and NASA’s Bioclim2 datasets. Oxygen-related atmospheric data will be derived from ERA5 reanalysis outputs (air composition, O flux proxies such as vertical wind and pressure gradients) and, where available, from specialized combustion experiments (Thomas et al., 2017).
- **Fire and Ecological Data:** MODIS Fire Radiative Power (FRP) and burned-area products (MCD64A1) will be used to capture fire occurrence and spread patterns. Fuel and vegetation structure will be extracted from GEDI LiDAR products (canopy height, cover fraction) and corroborated with regional land cover maps (ESA CCI or USGS NLCD).
- **Soil and Terrain Data:** Soil moisture and composition will be sourced from SoilGrids and SMAP satellite products, while topography and slope will be derived from SRTM or ALOS DEM datasets.

- **Data Harmonization:** All datasets will be reprojected to a common coordinate system (WGS84) and temporally aligned to the simulation period. Interpolation (e.g., kriging) will fill gaps where necessary, and normalization/scaling will prepare data for model input.
- **Feature Extraction:** Statistical and machine learning techniques (e.g., PCA, correlation clustering) will be applied to derive dominant ecological and atmospheric patterns, including oxygen-related variables (flux gradients, near-surface concentration). Extracted features will condition both the simulation and the deep learning–based video generation model.

Empirical evaluation will combine quantitative performance testing and qualitative expert validation, ensuring that both scientific accuracy and visual interpretability are assessed. The evaluation framework will address three dimensions:

- **Simulation Accuracy and Sensitivity:**
  - Compare predicted fire spread metrics (rate of spread, flame height, burned area, heat release) from the proposed oxygen-integrated simulations against baseline models that omit oxygen flux.
  - Use statistical measures such as RMSE, MAE, and  $R^2$  to quantify improvement.
  - Conduct sensitivity and feature-importance analysis (e.g., SHAP, Sobol indices) to verify that the model correctly identifies oxygen flux and related variables as influential parameters.
- **Video Generation Quality:**
  - Evaluate generated 3D fire simulations using standard video generation metrics (Fréchet Video Distance, Structural Similarity Index).
  - Solicit expert feedback from fire ecologists and atmospheric scientists to score realism, interpretability, and ecological plausibility of generated visuals on a 5-point Likert scale.
  - Solicit expert feedback from fire ecologists and atmospheric scientists to score realism, interpretability, and ecological plausibility of generated visuals on a 5-point Likert scale.
  - Compare model outputs with real fire imagery (MODIS/VIIRS) for spatiotemporal pattern fidelity.
- **Interpretability and Practical Utility:**
  - Assess how effectively the visualizations help users understand cause–effect relations (oxygen vs. temperature, moisture, or wind).
  - Conduct a small user study measuring accuracy and response time for interpretive tasks (e.g., identifying key fire spread drivers from visual output).
  - Collect qualitative feedback on clarity, usability, and potential decision-support applications.
- **Generalization Tests:**
  - Train the framework on subsets of environmental variables (with and without oxygen flux, or with varying soil and moisture conditions) to assess robustness and generalization across ecosystems.
  - Report consistency of variable rankings and simulation outcomes across scenarios.

### 3 Related Work

Current fire simulation and visualization models have advanced significantly in coupling climate/weather, terrain, fuel type, and moisture, enabling 3D immersive visualizations that help with understanding fire spread and extinguishing behavior (e.g., [2], [6]). However, many of these models assume standard atmospheric oxygen and do not include oxygen flux or altered oxygen concentration.

A range of modelling and visualization techniques have been developed to simulate forest fire spread under differing environmental conditions. Cellular automata (CA) models, improved Huygens’ principle, finite state machines, and statistical/AI approaches are frequently used. For example, [2] produce immersive 3D visualizations of fire spread and extinguishing (sprayed retardants, precipitation) using CA combined with dynamic weather, terrain, and fuel moisture. [6] approximate tree morphology and use finite state machines to model pyrolysis, heating, and burning, enabling real-time 3D visualization of fire over extended landscapes.

Experimentally, [3] investigate how oxygen flux into fuel beds affects burning dynamics: flame duration, heat release, combustion efficiency, etc. [4] simulate at the global scale how increased atmospheric oxygen leads to higher fire activity, yet forests may still survive under feedbacks.

Although there are AI / statistical fire spread models (e.g., [5]), and classical models like cellular automata (e.g., [1]), these works rarely incorporate oxygen as an explicit modulating variable; seldom provide dynamic video generation or deep learning-conditioned visualization; and often do not allow the user to explore how changing oxygen/flux in combination with ecological parameters or microclimate alters outcomes.

### 4 Research Plan

The schedule for this project is six weeks, including time for writing the final abstract and preparing the final presentation. Weekly milestones are as follows:

- **Week 1–2: Data extraction and preprocessing.** Collect historical climate datasets, oxygen flux measurements, and ecological records. Standardize and preprocess the data to align temporal and spatial resolution. Begin exploratory analysis to confirm variable distributions and correlations.
- **Week 3: Simulation model training and feature ranking.** Train baseline deep learning models on the processed datasets. Implement systematic feature importance ranking to identify which ecological and atmospheric variables most strongly affect fire dynamics (e.g., oxygen availability, soil moisture). Begin comparing against existing feature sets used in the literature.
- **Week 4: Visualization pipeline development.** Construct a pipeline that transforms model outputs into dynamic 2D animated video simulations. Optimize the system for interpretability and ensure the spatiotemporal fidelity of fire spread patterns. Address challenges in training the video generation model on extracted climate–oxygen features.
- **Week 5: Evaluation and testing.** Perform user-level evaluations of the visualization outputs for clarity, interpretability, and ecological realism. Compare model outputs with known fire events to validate predictive performance. Conduct error analysis and refine the pipeline accordingly.
- **Week 6: Refinement and final deliverables.** Improve the visualization quality and streamline the pipeline. Draft the final project abstract, prepare a comprehensive report, and design the final pre-

sentation. Demonstrate the tool’s use for proactive intervention planning by simulating ecological prevention strategies.

Doability: A 3D simulation/video may be ambitious within this timeframe, but a 2D animated video generated according to ecological parameters is feasible. This project will leverage deep learning libraries, existing simulation datasets, and visualization platforms to ensure progress.

## 5 Resources

This project will require both computational and software resources for successful implementation:

### Hardware

- **High-performance GPU compute nodes** (NVIDIA A100, V100, or equivalent) to train deep learning models for fire simulation and video generation. A single multi-GPU server with at least 40GB of GPU memory will be sufficient for the proposed scale.
- **Multi-core CPU cluster access** for preprocessing large climate and ecological datasets (temperature, oxygen, soil moisture, etc.).
- **Local workstation** (laptop/desktop) with moderate GPU (e.g., NVIDIA RTX 3060 or above) for development, debugging, and small-scale visualization testing.
- **Storage capacity** of at least 2–3 TB for storing raw historical datasets, intermediate simulation outputs, and generated video visualizations.

### Software

- **Programming languages:** Python (primary for data processing, deep learning, visualization), C++ (optional for simulation backends).
- **Deep learning frameworks:** PyTorch or TensorFlow for model training and video generation.
- **Visualization tools:** ParaView, Blender, or Unity for rendering dynamic fire simulations and producing 2D/3D video outputs.
- **Scientific libraries:** NumPy, Pandas, SciPy for numerical analysis; Matplotlib and Seaborn for exploratory visualization.
- **Geospatial/Climate data handling:** GDAL, xarray, and netCDF4 for handling large-scale climate and ecological datasets.
- **Version control and collaboration:** GitHub/GitLab for code management and documentation.

### Facilities

The project will leverage Brown University’s computing resources, including GPU clusters available through the Center for Computation and Visualization (CCV). In addition, open-source datasets (e.g., NASA climate archives, SoilGrids, and forest fire incident reports) will be accessed as inputs to the model.

## References

- [1] A. Hernández Encinas et al. Modeling forest fire spread using cellular automata: a comparative study. *Simulation Modelling Practice and Theory*, 2015.
- [2] X. Meng et al. Immersive 3d visualization of forest fire spread and extinguishing simulation using cellular automata. *Journal of Visualization*, 2023.
- [3] J. C. Thomas, R. M. Hadden, and A. Simeoni. Experimental investigation of the impact of oxygen flux on the burning dynamics of forest fuel beds. *Fire Safety Journal*, 2017.
- [4] R. Vitali, C. M. Belcher, J. O. Kaplan, and J. Watson. Increased fire activity under high atmospheric oxygen concentrations is compatible with the presence of forests. *Philosophical Transactions of the Royal Society B*, 2016.
- [5] Z. Wu, B. Wang, M. Li, Y. Tian, Y. Quan, and J. Liu. Simulation of forest fire spread based on artificial intelligence. *Ecological Indicators*, 136:108653, 2022.
- [6] Y. You et al. Finite state machine based modeling of tree pyrolysis and fire spread for real-time visualization. *Fire Safety Journal*, 2022.

# Aditi Kannan

(401)-230-2439 | [aditi\\_kannan@brown.edu](mailto:aditi_kannan@brown.edu) | [linkedin.com/in/aditi-kannan-73289a252](https://www.linkedin.com/in/aditi-kannan-73289a252) | [github.com/aditileclerc](https://github.com/aditileclerc)

## EDUCATION

**Brown University**, Providence, RI

*Master of Science in Computer Science*

*Expected May 2027*

Relevant Coursework: Deep Learning, Interdisciplinary Scientific Visualization

**Shiv Nadar University**, Chennai, India

*Bachelor of Technology in Computer Science Engineering (IoT)*

*July 2025*

Relevant Coursework: Machine Learning, Embedded AI, Artificial Intelligence, Internet of Things, Distributed Computing

Honors: First Class with Distinction

## SKILLS

Computer Skills: C, C++, Java, Python, Arduino, NS2, CARLA, TensorFlow, Keras, PyTorch, HTML, CSS, JavaScript, React, Node, Streamlit, MySQL, Cypher, Git, Llama

Research Skills: Microsoft Excel, Tableau

## ACADEMIC PROJECTS

**Loan Status Prediction Model** [[Github](#)]

September 2022 - December 2022

- Developed a machine learning-based loan status prediction model tailored to the needs of small financial institutions and non-bank financial companies.
- The model leverages the Support Vector Machine (SVM) and Linear Regression algorithms to analyze borrower data and predict the likelihood of loan approval or default.

**GPU Accelerated Brain Simulations with Procedural Connectivity**

September 2024 - February 2025

- The project accelerates the Spiking Neural Network (SNN) simulation of Biological Neural Networks, on a single GPU system, using parallel computing methods and simulate the trials using GeNN simulator.

**Stock Price Prediction Using Gen AI and Opinion Mining**

March 2024 - November 2024

- The model predicts the market value of stocks using a generative AI model, such as a transformer model to perform opinion mining and judge the value of the stock. A financial RNN, FinBERT, which is a Transformer based model and VADER, was used for the opinion mining.

## PROFESSIONAL EXPERIENCE

**Zoho Corporation, Chennai, India**

December 2023 - January 2024

*Internship*

- As a part of the Key Manager Plus team, I worked with SSL and SSH key certificates and developed a system of storing and verifying security certificates using Java and JavaScript

**PurpleSlate Ltd, Chennai, India**

May 2024 - June 2024

*Internship*

- Worked mainly with Generative AI, Large Language Models, Prompt Engineering and Transformer model based chatbots, with the ParrotGPT team. Built a custom chatbot [[Github](#)]

## LEADERSHIP POSITIONS

**English Literary Club, Shiv Nadar University, Chennai**

July 2022 - May 2025

*Club President*

- Organized several events and club activities, implemented and managed the club structure.

**SSN-SNUC MUN, Shiv Nadar University, Chennai**

August 2024 - September 2024

*Under Secretary General of Technology*

- Worked along with my team of developers to redesign the conference website and worked on a lunch verification application.



# Olivia Heng

[hengoliviadja@gmail.com](mailto:hengoliviadja@gmail.com) | (609) 225-2109 | [linkedin.com/in/olivia-heng/](https://www.linkedin.com/in/olivia-heng/)

## EDUCATION

- Brown University, Providence, Rhode Island** | Computer Engineering, Sc.B. & Computer Science, Sc.M. September 2022 – **Expected Graduation May 2026**
- *Relevant Coursework:* Data Structures and Algorithms, Deep Learning, Computer Vision, Computer Systems, Digital Electronics Systems Design, Design of Computing Systems, Digital Integrated Circuits, Pattern Recognition and Machine Learning, Linear Algebra, Multivariable Calculus, Differential Equations
  - *Current Coursework:* Embedded Microprocessor Design (Capstone), Interdisciplinary Scientific Visualization, Electricity and Magnetism

## ENGINEERING EXPERIENCE

### School of Engineering, Brown University

- *Teaching Assistant ENGN0030: Introduction to Engineering* September 2024 – Present
  - Host office hours and teach EMG Claw project for **40+ students** in Wearables module focusing on **Arduino** sensors in addition to course grading
- *Lead Teaching Assistant ENGN0032: Introduction to Engineering: Design* February 2023 – January 2024
  - *Lead Studio TA:* Instruct **25+ students** in a weekly studio focusing on engineering concepts, such as vectors, coupling, moments, and MATLAB
  - *The Chair Prototype Project Lead:* Coordinate weekly development meetings, ensure clear communication with Design Workshop and 50+ staff
  - *Communications Lead:* **Redesigned lecture structure** and content based on previous iterations of the course, coordinate TopHat and EdStem tools, analyze feedback data to propose changes to course, develop the course website and update content on a weekly basis for the **340+ student course**
- *Teaching Assistant ENGN0040: Dynamics and Vibrations* January 2024 – May 2024
  - Host weekly **office hours** aiding students with problems sets and design projects focused on course topics, such as kinematics of rigid bodies/systems

## COMPUTER SCIENCE EXPERIENCE

### Microsoft, Azure Data Analytics

- *Software Engineer Intern* May 2025 – August 2025
  - Implemented **serialization pipelines** for **Substrait** query plans using alternative frameworks (FlatBuffers, Cap'n Proto, FBE) across **C#** and **C++**
  - Designed and **benchmarked** TPC-H workloads to assess performance **trade-offs** (schema translation complexity, memory footprint, message size) in real-world **data warehousing** scenarios, contributing findings to internal tooling and future architectural decisions
  - Partnered with engineers across Azure Data Analytics to validate **cross-language compatibility**, troubleshoot runtime integration issues, and refine schema evolution strategies, contributing to a more resilient and **scalable** serialization layer for internal and external consumers
- *Software Engineer Intern* May 2024 – August 2024
  - Integrated an **open-source specification** for complex query plans into SQL Server Management Studio, which is currently in deployment (SSMS 21)
  - Developed a **unique graphical visualizer** with the open-source integration using **C#, Visual Studio**, and **.NET libraries**, enhancing the efficiency of query development processes by providing intuitive access to critical performance data
  - Collaborated with the Azure Data team to enhance and leverage their robust cloud computing capabilities while learning best development techniques

### Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility

- *Education Collaboration Software Engineer Intern* January 2024 – May 2024
  - Project: *Exploring the Integration of Tensorflow Models into NVIDIA FLARE*
    - Integrated an open-source Tensorflow model into NVIDIA Federated Learning Application Runtime Environment for supercomputer **Frontier**
    - Utilized **Python**, **reStructuredText**, and **Tensorflow** for integration and contributed to **NVIDIA FLARE** and OLCF documentation
- *Technical and Professional Software Engineer Intern* August 2023 – December 2023
  - Project: *Utilizing Asynchronous Dynamic Content Rendering with Hotwire's Turbo and Applying Code Efficiency to ORNL Existing Applications*
    - Implemented **Hotwire's Turbo** within the website application: Facility Efficiency Tracking Application
    - Establish the **software lifecycle** for the provided project for operational software systems in a production environment
- *Pathways to Computing Software Engineer Intern* May 2023 – July 2023
  - Project: *Improving Metric Granularity to Enable Research on Optimal Work Situations through Software Development*
    - Built a **full stack** website application, the Facility Efficiency Tracking Application (FETA), to track the ORNL Software Services Development Team's efficiency using **Ruby on Rails 7**, **JavaScript**, **Git**, and **CI/CD**
    - Resulted in greater issue granularity, a **29% increase** in Search Engine Optimization, and **20% increase** in Accessibility compliance

### Hack@Brown, Brown University

- *Director* September 2022 – March 2025
  - Direct event logistics for Brown University's oldest and largest hackathon, **hosting 500+ students** annually for an in-person 24-hour hackathon
  - Lead and coordinate with **41 members** across Sponsorships, Experience, Dev, Workshops, and Design Teams to prepare for hackathon, **raising \$26k+**
- *Experience Team Lead:* Lead Experience Team to design the ideal hacker experience by planning all hackathon logistics, such as publicizing, sourcing panelists and judges, etc; Coordinated with team leads and directors throughout the year to effectively lead the executive team on the day-of the hackathon
- *Experience Team Member:* Connect with industry alumni and professionals and coordinate swag and registration

### Code With Klossy

- A 501(c)(3) charitable organization that creates learning experiences and opportunities for women and gender expansive youth through passion igniting programs
- *Machine Learning and Artificial Intelligence Instructor Assistant* February 2023 – August 2023
  - Led Culture of Tech sessions "Creating Opportunity" and "Designing Tech Responsibility" to inform scholars of the unique challenges that women, gender non-conforming people, and people of color face in tech along with forwardly thinking of solutions — working **individually with ~30 scholars**
  - Support scholars with **debugging**, **troubleshooting**, and **adapting content** to meet scholar needs throughout the program, specifically using **Co-ML**

## SKILLS AND INTERESTS

**Technical:** Java, Python, C#, JavaScript, HTML, CSS, Verilog, C, C++, MATLAB, SQL, Swift, Ruby, Ruby on Rails, ReasonML, Racket; Git, Bash, Linux, Tensorflow, NumPy, Docker, Google Colab, .NET libraries, Quartus, Sequel Pro, Fusion360, OnShape

**Certifications:** Azure Data Fundamentals, Azure AI Fundamentals (Microsoft); Cybersecurity (Google); High Performance Computing (ORNL, Purdue University)

**Industrial:** Laser cutting, woodworking, metalworking, 3D printing, soldering

**Other Skills:** American Sign Language (basic), certified in CPR

## PROFESSIONAL ACTIVITIES

**Publication:** DOI: 10.1145/3626203.3670571 - "Improving Metric Granularity to Enable Research on Optimal Work Situations through Software Development"

**Conferences:** PEARC24: Human Powered Computing (**Student Poster**), Supercomputing '25, '24, and '23 (**Lead Student Volunteer**, HPC Immersion Program), PEARC23: Computing for the Common Good (**Student Volunteer**, Student Program), Richard Tapia Celebration of Diversity in Computing Conference '23

**Extracurriculars:** Girls Who Code College Loops (**Treasurer**), Society of Women Engineers, Mosaic+, Society of Asian Scientists and Engineers, Women in CS



## Regarding Proposal for CSCI 2370 Class



**Kannan, Aditi** <aditi\_kannan@brown.edu>

to Divya ▾

10:25 PM (0 minutes ago)



Hello Divya,

This is Aditi Kannan from Dr. David Laidlaw's CSCI 2370 class! I've come up with a proposal idea for the Visual Analysis of an Ensemble Forest Fire Simulation project. I've attached the proposal below, and it would be great if I could get your thoughts and feedback on the idea!

As a part of the submission process, we're asked to confirm the collaborator's participation. If you could kindly reply to this email with your response, I'd greatly appreciate it! Thank you so much for giving us this opportunity.

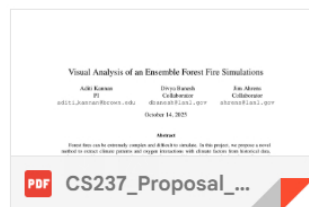
Looking forward to hearing from you,

Thank you again,

Best,

Aditi Kannan

**One attachment** • Scanned by Gmail ⓘ



# Towards 3D Visualization of Power Transmission Grid Systems

Aidan LaBella

PI

`alabella@brown.edu`

Yonas Amha

Co-PI

`yonas_amha@brown.edu`

Kenny Gruchalla

Collaborator

`Kenny.Gruchalla@nrel.gov`

Rahul Kadavil

Collaborator

`rahul.kadavil@nypa.gov`

October 13, 2025

## Abstract

Power transmission grids are complex systems that require extensive monitoring. Visualizing the data from these complex systems has become a challenge in recent years. Previous works ultimately fail to completely capture all the dynamics of these multivariate systems. We propose a novel approach that leverages 3D visualization methods with the potential for immersive virtual reality (VR).

# 1 Aims

The primary aim of this project is to implement a 3-dimensional power transmission grid visualization utility. Specifically, we aim to provide an immersive virtual reality experience for all types of users, whether they have a background in electrical engineering or not. The **scientific contribution** is that subject matter experts (SMEs) will be able to more quickly gather insights into the dynamics of power grid transmission systems, specifically when there are anomalies or other grid disruptions. Our **visual contribution** will be to provide an innovative and intuitive visualization utility that is effective at displaying the complexity of the multivariate data that power transmission grids produce.

We also aim for our software to be robust and abstract enough to operate on many different sources of transmission grid data. As described in section 4.2, not all of the data that needs to be visualized is available in the public domain. As a result, we propose building software that can work at any geographic location, so long as the data is complete.

# 2 Significance

The intricate workings of power transmission grid systems are difficult to explain, even for SMEs. The current state of the art is limited as the data cannot be explained in an intuitive manner to those that are not SMEs. Although those who work with this type of data may have developed an innate intuition as to what features are important during analysis, those who are not experts (such as shareholders, executives, project managers, etc) will likely not have such an intuition. Additionally, the current 2D representations contain lots of visual clutter, as demonstrated in figure 3, which can make analysis tricky and features left indiscernible. This is where 3D data representations, such as immersive virtual reality, have the potential to provide visual information in ways the previous state of the art has not been able to.

# 3 Related Work

The most recent work concerning the effectiveness of current power transmission data visualization methods has explored the usefulness and correctness of existing power grid data, such as with 2D contour maps [2], which were found to largely misrepresent the data. The main weakness of these 2D contour map representations is their inability to provide a streamlined representation of multivariate data, resulting in indiscernible data streams, particularly for voltage violations (i.e. the voltage exceeds or goes below its standard operating range). The human subjects tasked with finding these violations in the data were ultimately only to correctly identify (on average) five out of fifteen violations. We expect 3D representations to be a plausible solution to this problem as 3D representations, in general, are able to display data at a higher fidelity.

In 2011, Zhu et. al [8] took a data-driven approach to visualization of power grid data. This includes the paradigm of “query-driven” visualization, where the authors focus their visualization efforts on the data that users frequently query. This only addresses the issue of how large amounts of data are filtered, not necessarily the mechanics how it is presented. Our approach will differ in that we propose a volume-agnostic approach in the sense that users should be able to visualize all of the data, not just the most frequently queried data.

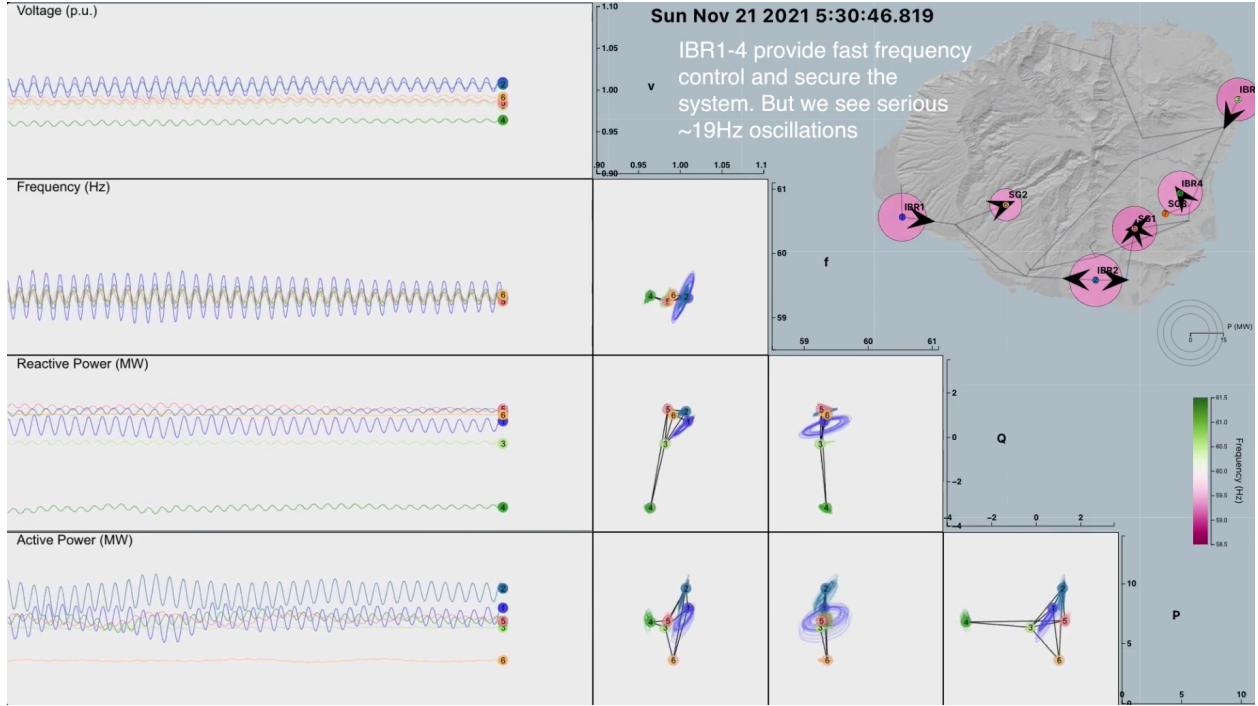


Figure 1: Recent 2D Visualization Method [2]

In the year 2000, Overbye et al [6] provided a novel 3D visualization method of transmission grid data. This approach was extremely limited in the sense that the authors only provided 3D projects on to existing 2D maps. Our approach will differ in that we will implement a 3D visualization method with existing 3D representations of transmission grid locations.

Other notable work includes physics-based visualizations [7], the use of broadband cable TV data [3] and web-based applications [4].

## 4 Research Plan

### 4.1 Western Electricity Coordinating Council (WECC)

The source of our data for this project will be from the western electrical coordinating council (WECC). We have already been provided approximately 200MB of WECC data in CSV format.

### 4.2 New York Power Authority (NYPA)

One target user for our novel visualization method will be the New York Power Authority, which monitors the transmission grids for the entire State of New York. We propose initially implementing our prototype with the WECC data and then adding a level of abstraction that can support data from other sources, such as the NYPA. Because the NYPA data is proprietary, we will require close collaboration with our contacts at the NYPA to ensure our data specifications are congruent.

| Week | Task                                                                                                               | Deliverables and Outcomes                                             |
|------|--------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| 1-2  | Implement the existing SOTA 2D method [5], get to know the data, collaborators and develop initial infrastructure. | A replication of what is shown in 3, with working codebase.           |
| 3-4  | Develop a 3D model of the transmission grid system and begin VR implementation.                                    | A deliverable 3D product that can be shipped to users for evaluation. |
| 5-6  | Conduct evaluation studies with users for 2D and 3D models.                                                        | Qualitative and empirical results from user studies.                  |

Table 1: Detailed weekly work plan

### 4.3 Evaluation

In coordination with our collaborators at the New York Power Authority (NYPA) and The National Renewable Energy Laboratory (NREL), we will task human evaluators with comparing our new 3D visualization method with the previous state of the art. We expect users to find our approach to be more intuitive and informative when compared to the previous SOTA. As many parts of this visualization task are subjective and user specific, we propose two evaluation studies: a qualitative and empirical analysis.

**Qualitative Analysis** The first will be a qualitative study for users at the NREL and NYPA where each participant evaluates their perceived usefulness with a likert scale. We will conduct two studies, one on the 2D evaluation method shown in 3 and the other on our 3D implementation. We will ask users to evaluate the ease of use, clarity, interactive-ness and perceived value.

**Empirical Analysis** The second study will consist of testing users on their ability to identify features of the data. This process will consist of 1.) creating a ground truth subset from the WECC dataset for various features (i.e. voltage, power, current, etc.). We will test the users ability to depict the correct values of these features for certain time steps for both the 2D and 3D models. In addition to testing how much of the data the users are able to accurately depict, we will also measure the time it takes for the users to make such depictions.

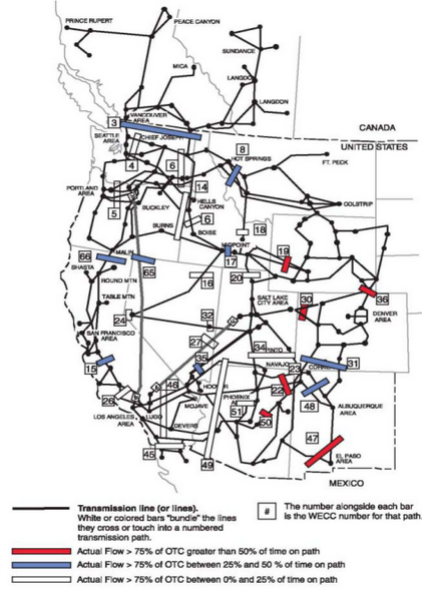


Figure 2: Map of the WECC [1]

## References

- [1] U. DOE. National electric transmission congestion study, 2006.
- [2] K. Gruchalla, S. Molnar, and G. Johnson. Reevaluating contour visualizations for power systems data. *IEEE Transactions on Smart Grid*, 14(6):4877–4887, 2023.
- [3] G. Johnson, K. Gruchalla, M. Ingram, N. Guba, R. Cruickshank, and S. Caruso. Vis-saga: Visual analytics for situational awareness of grid anomalies. In *2023 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pages 17–21. IEEE, 2023.
- [4] G. Johnson, S. Molnar, N. Brunhart-Lupo, and K. Gruchalla. Architecture for web-based visualization of large-scale energy domains. In *2024 IEEE Workshop on Energy Data Visualization (EnergyVis)*, pages 46–51. IEEE, 2024.
- [5] I. Lyons-Galante, M. Karimzadeh, S. Molnar, G. Johnson, and K. Gruchalla. Alternatives to contour visualizations for power systems data. In *2023 Workshop on Energy Data Visualization (EnergyVis)*, pages 11–15. IEEE, 2023.
- [6] T. J. Overbye and J. D. Weber. Visualization of power system data. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, pages 7–pp. IEEE, 2000.
- [7] P. C. Wong, K. Schneider, P. Mackey, H. Foote, G. Chin Jr, R. Guttromson, and J. Thomas. A novel visualization technique for electric power grid analytics. *IEEE Transactions on Visualization and Computer Graphics*, 15(3):410–423, 2009.
- [8] J. Zhu, E. Zhuang, C. Ivanov, and Z. Yao. A data-driven approach to interactive visualization of power systems. *IEEE Transactions on Power Systems*, 26(4):2539–2546, 2011.

# Aidan P. LaBella

Email : alabella@brown.edu

**Citizenship:** United States of America

<https://github.com/aidanlabella/>

<https://cs.brown.edu/people/grad/alabella/>

Mobile (Preferred) : +1 (518) 779-9285

## INTERESTS

---

**Data Science and Engineering • Machine Learning • Computational Science • Time Series Analysis**

## EDUCATION

---

- **Ph.D., Computer Science, Brown University** Providence, RI  
2023 - Present *In Progress*
  - **Advisor:** Stephen Bach
- **Sc.M., Computer Science, Brown University** Providence, RI  
2023 - 2025 *Conferred May 2025*
  - **Examination Committee:** Stephen Bach, Karianne Bergen, Jonathan Pober
- **B.Sc. with Honors, Computer Science, Rochester Institute of Technology (RIT)** Rochester, NY  
2018 - 2023 *Conferred May 2023*
  - **Advisor:** Travis Desell
  - **Minor and Immersion:** Criminal Justice

## TECHNICAL SKILLS

---

- **Languages:** Python, Java, C, C++, C#, Bash, ECMAScript, GoLang, R, MATLAB, Common Lisp, Icon, Prolog.
- **Tools:** Git, Maven, Vim, Visual Studio, jQuery, React, Bootstrap, Docker, Atlassian JIRA, Jenkins.
- **DBMS:** MySQL, MariaDB, MongoDB, PostgreSQL, SQLite.
- **Cloud:** AWS, Kubernetes.

## EXPERIENCE

---

- **Undergraduate Research Assistant** RIT, Rochester, NY  
*Distributed Data Science Systems Lab* *May 2020 - May 2023*
  - Worked with faculty and graduate students in research projects focused on Aviation Safety and Machine Learning.
  - Co-authored journal and conference articles covering the research conducted.
  - Trained new students coming into the lab that would be working on our various projects.
- **Software Engineer Co-Op** EagleView Technologies, Rochester, NY  
*Full member of an Agile Release Train (ART) and participant in the Agile SDLC* *Jan. 2022 - Apr. 2022*
  - Full-stack software developer as part of an Agile team responsible for delivering EagleView's imagery to customers via web-based applications.
  - Performed incremental front-end changes in JavaScript for EagleView's main software platform, ConnectLink.
  - Utilized GoLang on backend services to ensure server reliability.
- **Peer Mentor** RIT, Rochester, NY  
*Tutored students in the introductory Computer Science sequence at RIT* *Jan. 2020 - Apr. 2020*
  - Advised and mentored students with programming assignments, exam preparation, problem solving and technical issues.
  - Worked in both an in-person and online format to provide assistance to students.
  - Provided assistance for Java, C, C++ And Python languages.



## PROJECTS

---

### CURRENT

- **“SWIFT-SAT: RFI Detection Across Six Orders of Magnitude in Intensity: A Unifying Framework with Weakly Supervised Machine Learning”**: *Funded by the National Science Foundation (Award #2228989)*
  - Develop machine learning applications for the detection and removal of radio frequency interference (RFI) from 21cm cosmology data.
  - Introduce weak supervision to harvest RFI detection rules that produce meaningful training datasets.
  - Interdisciplinary work with members of Brown’s Computer Science and Physics departments.

### PAST

- **The National General Aviation Flight Information Database (NGAFID)**: *Funded by the Federal Aviation Administration*
  - Implemented calculations for Loss of Control and Stall Indices. Both of these indices are utilized by pilots to identify potentially unsafe flight conditions.
  - Performed numerous UI/UX enhancements, bugfixes and added features at the request of the project stakeholders.
  - Added features that enabled users to simulate previous flights in the X-Plane Flight Simulator.
- **Evolutionary eXploration of Augmenting Memory Models (EXAMM)**:
  - Implemented a backpropagation-free neural architecture evaluation method known as Random Error Sampling-based Neuroevolution (RESN).
  - Trained recurrent neural networks in predicting various time series columns from the NGAFID project data.

## PUBLICATIONS

---

- [1]: Aidan LaBella, Joshua Karns, Farhad Akhbardeh, Andrew Walton, Zechariah Morgan, Brandon Wild, Mark Dusenbury and Travis Desell. **Optimized Flight Safety Event Detection in the National General Aviation Flight Information Database** *In The 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22)*. Online. April 25-29, 2022.
- [2]: Hong Yang, Aidan LaBella and Travis Desell. **Predictive Maintenance for General Aviation Using Convolutional Transformers** *The Thirty-Fourth Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-22)*. Vancouver, British Columbia, Canada. February 24 - 26, 2022.
- [3]: Aidan LaBella, Charles Duong, Pak Iong Long, Nathan DePiero, Aditya Iyer, Elise Carman, Randall Balestrieri and Travis Desell. **GATS: A Time-Series Dataset for Addressing General Aviation Flight Safety** *1st ICML Workshop on Foundation Models for Structured Data*

## PUBLIC TALKS

---

- Aidan LaBella. **Optimized Flight Safety Event Detection in the National General Aviation Flight Information Database**. *The 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22)*. Online. April 25-29, 2022.
- Aidan LaBella. **Optimized Flight Safety Event Detection in the NGAFID**. *The 2022 RIT Undergraduate Research Symposium*. Rochester, NY. July 28, 2022.
- Aidan LaBella. **Data Mining in the National General Aviation Flight Information Database**. *DataFest 2023*. Rochester, NY. March 24, 2023.

## SERVICE

---

- **Computer Science Department Representative** Brown University  
*Graduate Student Council* *Sept 2023 - Present*
  - Participate in monthly meetings to discuss and vote on issues pertaining to graduate students at Brown.

MENTORING

---

- **SWIFT-SAT: RFI Detection Across Six Orders of Magnitude in Intensity** Brown University  
*Scientific Machine Learning* Sept 2023 - Present
  - **Kevin Scroggins, Jr.:** Sc.M., Computer Science
  - **Christopher Tripp:** Sc.M., Computer Science
  - **Ronit Kapoor:** Sc.M., Computer Science
  - **Sarah Liu:** Sc.B., Computer Science
  - **Charles Duong:** Sc.B., Computer Science
- **GATS: Time Series Data Analysis for General Aviation with Self-Supervised Learning** Brown University  
*Machine Learning for General Aviation Safety* Sept 2024 - Present
  - **Elise Carman:** Sc.M., Computer Science
  - **Justin Long:** Sc.M., Computer Science
  - **Nathan DePiero:** Sc.M., Computer Science
  - **Charles Duong:** Sc.B., Computer Science

**From:** Gruchalla, Kenny <[Kenny.Gruchalla@nrel.gov](mailto:Kenny.Gruchalla@nrel.gov)>

**Date:** Thursday, September 18, 2025 at 11:13 AM

**To:** Aidan LaBella <[aidan\\_labella@brown.edu](mailto:aidan_labella@brown.edu)>, Kadavil, Rahul <[rahul.kadavil@nypa.gov](mailto:rahul.kadavil@nypa.gov)>

**Cc:** [dhl@cs.brown.edu](mailto:dhl@cs.brown.edu) <[dhl@cs.brown.edu](mailto:dhl@cs.brown.edu)>, [bumjin\\_joo@brown.edu](mailto:bumjin_joo@brown.edu) <[bumjin\\_joo@brown.edu](mailto:bumjin_joo@brown.edu)>, [timothy\\_fong@brown.edu](mailto:timothy_fong@brown.edu) <[timothy\\_fong@brown.edu](mailto:timothy_fong@brown.edu)>, Korey Sam <[korey\\_sam@brown.edu](mailto:korey_sam@brown.edu)>, Amha, Yonas <[yonas\\_amha@brown.edu](mailto:yonas_amha@brown.edu)>

**Subject:** Re: NYPA Project -- Brown University

Hi Aidan,

I'd be happy to chat on Monday anytime after 1pm (Eastern).

Rahul,

As reminder this is regarding the VR class at Brown University that are considering exploring new ways to explore at the WECC data immersively.

-Kenny

Kenny Gruchalla, Ph.D.

Principal Scientist & Distinguished Member of Research Staff

Computational Science Center

National Renewable Energy Laboratory

<https://gruchalla.github.io>

Link to the manuscript pdf:  alabella.pdf

## Response Letter

Dear Editor and Reviewers,

Thank you for your attention to our proposal. We have addressed your concerns below:

[R 1.1]

<https://drive.google.com/file/d/19pxgbQ1OSi3BaELV0dL2PkN2ak2rkrT6/view?usp=sharing>

*Feasibility concerns: achieving robust 3D interactivity, real-time data integration, and meaningful user evaluation within six weeks may not be realistic.*

We understand that this is an optimistic timeframe, but want to stress that the main goal is to implement an intuitive 3D representation of the data; not necessarily to have a fully-implemented VR environment.

[R 2.1]

[https://drive.google.com/file/d/1TSAj\\_t4WdA0r8Pb4d6jZPzOaywrKTizv/view?usp=sharing](https://drive.google.com/file/d/1TSAj_t4WdA0r8Pb4d6jZPzOaywrKTizv/view?usp=sharing)

*Placed strong emphasis on "intuitiveness" but lacks a clear definition over the term.*

We understand that the intuitiveness may be vague, but we want to stress that ultimately we are looking for a "one size fits all" approach that any user of any experience level can use. That is what we ultimately mean by intuitiveness.

[R 2.2]

*Lack a detailed description about what new visualization approaches to take in addition to the existing ones.*

We understand that saying "visualize in 3D space" may be vague, but ultimately we are looking for a semi-realistic representation of the transmission grids. This includes illustrations of the infrastructure, with overlays representing the various measurements that are being taken at any given time.

# Visualizing LiDAR Preprocessing for Intra-Site Visibility at a Chachapoya Settlement

Charly Castillo  
PI

`charly_castillo@brown.edu`

Sir Dylan Allotey  
Co-PI

`sir_dylan_allotey@brown.edu`

Parker VanValkenburgh  
Collaborator  
`parker_vanvalkenburgh@brown.edu`

October 14, 2025

## Abstract

We present *VisArch*, a three-part visualization for exploring how LiDAR preprocessing affects archaeological visibility. The approach improves upon standard visibility modeling, which often relies on raw surfaces that obscure distinctions between architecture and vegetation. In a comparative case study, *VisArch* more clearly revealed architectural features and structure-to-structure relationships at a Chachapoya site.

# 1 Reviews & Responses

Dear Reviewers,

We appreciate the constructive feedback on the proposal, and we have prepared a revision accordingly. Please see below for an itemized list of your feedback and a brief description of how each is addressed in the revised manuscript.

## [R1] Gordan Milovac

### Overall: 3

**[R1.1]** Interdisciplinary: 2; A good interdisciplinary project connecting computer science (machine learning, data visualization) with archaeology to solve a domain-specific problem.

**[R1.2]** Scientific: 3; The project's goal to make the effects of LiDAR preprocessing choices transparent and understandable is a valuable scientific contribution to archaeology. It addresses a real-world problem where technical decisions can significantly alter scholarly interpretations of historical sites.

**[R1.3]** Visualization: 2; The core of the project is to build an interactive tool that allows users to see how different data cleaning methods change the resulting visibility networks. This is a direct and significant visualization contribution.

**[R1.4]** Significant: 3; The work is significant within its niche. For archaeologists working with LiDAR data in vegetated areas, a tool like this could be very useful for improving the rigor and transparency of their spatial analyses.

**[R1.5]** Novel: 3; While using machine learning to classify LiDAR data isn't new, the specific application: creating an interactive visualization tool to compare the interpretive consequences of these methods, appears to be a novel approach.

**[R1.6]** Goals clearly stated: 1; Excellent. The two primary aims are laid out as clear, numbered points, making the project's objectives easy to understand.

**[R1.7]** Likelihood of Success: 4; The technical plan to use XGBoost and develop a Shiny app is sound and seems achievable. The lack of a detailed week-by-week work plan makes it harder to assess, but the overall scope feels appropriate for a 6-week project.

**Response:** *The work plan has been further elaborated, with additional details such as concrete plans for user testing.*

### [R1.8] Strengths:

- Clear, Focused Problem: The proposal identifies a specific, practical problem in archaeological research and proposes a direct, targeted solution.
- Sound Technical Approach: The plan to use established machine learning (XGBoost) and visualization (R Shiny) tools is sensible and realistic.
- Well-Defined Goals: The project's two main objectives are stated with clarity and precision.

### [R1.9] Weaknesses:

- Missing Visuals: The proposal repeatedly refers to figures and visualizations that are placeholders, which makes it difficult to fully grasp the intended output.

**Response:** *The final proposal contains two figures, including a mock-up image that depicts the anticipated output of the tool.*

## [R2] Haonan Zhang

### Overall: 2

[R2.1] Interdisciplinary: 1; Excellent integration of archaeology, machine learning, and visualization. The collaboration with an archaeologist strengthens relevance.

[R2.2] Scientific: 2; Methodologically sound, using appropriate ML techniques (XGBoost, Random Forest) and quantitative validation. Could expand on error analysis or statistical robustness.

*Response: The revised proposal strengthens robustness through the addition of k-fold cross-validation.*

[R2.3] Visualization: 2; Strong visualization rationale—linking preprocessing to interpretive consequences is conceptually novel. Would benefit from more detail on visual encoding and interaction design.

*Response: The final proposal will include details about color visual encoding, side-by-side layout and organization, and interaction design.*

[R2.4] Significant: 2; Addresses a real research gap in archaeological transparency and could meaningfully influence how LiDAR data are interpreted.

[R2.5] Novel: 2; Innovative in connecting preprocessing decisions to interpretive visualization rather than performance metrics.

[R2.6] Goals clearly stated: 1; Clear objectives, milestones, and deliverables. The plan is realistic for the timeframe.

[R2.7] Likelihood of Success: 3; Technically feasible, but integration and Shiny app design may be time-intensive; risk of limited user feedback.

*Response: After discussing with our collaborator, the Shiny app has been replaced with a dashboard due to factors besides time. Additionally, the final version of the proposal more clearly states the inclusion of both non-expert and expert test audiences. It also outlines efforts to recruit expert archaeologists across different areas of the subject.*

### [R2.8] Strengths:

- Clear, research-driven motivation grounded in archaeological relevance.
- Combines ML classification and visualization transparency—a unique interdisciplinary niche.

### [R2.9] Weaknesses:

- Limited user-testing scope may restrict interpretability evaluation.
- May need careful scoping to ensure full app integration within time.

*Response: See response to R2.7.*

### [R3] Juliana Escuti

#### Overall: Not provided

[R3.1] Interdisciplinary: 1; This proposal combines visualization and archaeology.

[R3.2] Scientific: 2; This proposal clearly lays out how current preprocessing is limited in its application to the Chachapoyas region, and how solving this problem could improve interpretation of this data.

[R3.3] Visualization: 2; This proposal improves on the preprocessing step in this pipeline with the explicit goal of differentiating between architecture and vegetation.

[R3.4] Significant: 2; This proposal states how it will solve a current problem in this field using preprocessing. Finding ways to accurately reconstruct the history of the Chachapoya will offer new insight into indigenous communication systems and politics.

[R3.5] Novel: 2; This is a novel way of approaching the problem of classification in preprocessing for LiDAR point clouds in archaeology.

[R3.6] Goals clearly stated: 1; This project clearly lists two goals: evaluating preprocessing methods and building an interactive tool.

[R3.7] Likelihood of Success: 3; This is a challenging problem which has not been solved yet, but it seems to be within the skillset of this team.

#### [R3.8] Strengths:

- This proposal clearly defines the problem it aims to solve, and how solving the problem of classifying architecture could lead to significant new knowledge in this field.

#### [R3.9] Weaknesses:

- How will different preprocessing methods visually affect the final result? Will this be visible both to experts and non-experts? How will the feedback be quantified?

**Response:** *Firstly, the aim of the project will be further emphasized to ensure its significance is clearly conveyed. Secondly, although the tool will be tested with both expert and non-expert audiences, the intended primary audience is academics. This distinction will be made clearer in the final version. Lastly, feedback is planned to be largely qualitative, but new edits also include simple 1-5 rating metrics on factors such as clarity, usability, and interpretability.*



## **[R4] Korey Sam**

### **Overall: 4.7/5**

**[R4.1]** Interdisciplinary: 5/5; It bridges archaeology, geospatial data science, and visualization—linking LiDAR pre-processing (computer science) with interpretive archaeology (humanities) and UX/interface design.

**[R4.2]** Scientific: 4.2/5; The problem is well defined, and the approach comparing XGBoost vs. Random Forest vs. raw surfaces is methodologically rich.

**[R4.3]** Visualization: 4.7/5; The Shiny app enables direct interpretive comparisons of preprocessing choices—a brilliant example of transparency in modeling.

**[R4.4]** Significant: 5/5; Addresses an important methodological blind spot in archaeological LiDAR use, how preprocessing choices shape interpretation. The potential to uncover hidden site networks and challenge colonial biases in Andean archaeology adds social and cultural depth to its significance.

**[R4.5]** Novel: Not provided; Combining ML-based classification with interpretive visualization of visibility networks is original. Related work has focused on accuracy, this reframes the problem around interpretive consequences taking a conceptual leap.

**[R4.6]** Goals clearly stated: 5/5; The 6-week plan is realistic, milestones are measurable, and deliverables (Shiny app, user feedback, report) are clearly articulated. Stretch goals are properly scoped.

**[R4.7]** Likelihood of Success: 4.5/5; The core deliverables (classification and visualization) are well within reach, given existing R libraries and available data. The only mild risk lies in the limited timeframe for comprehensive user testing and polishing of the Shiny app.

### **[R4.8] Strengths:**

- There is a seamless integration of ML, LiDAR science, and archaeological theory. This proposal presents a focus on transparency and interpretive awareness, not just algorithmic performance. Finally, I felt that real cultural significance is tied to Chachapoya heritage and decolonizing research perspectives.

### **[R4.9] Weaknesses:**

- The user evaluation is small-scale and largely qualitative. A wider study could validate interpretive consistency. Further, there is limited detail on the visualization design, and the user interface may constrain replicability.

**Response:** See responses to R2.3 and R3.9.

## [R4] ChatGPT

### Overall: 2

[R4.1] Interdisciplinary: 1; Integration of computer-science visualization and archaeology is exemplary. Collaborator VanValkenburgh's expertise in Andean LiDAR interpretation ensures direct scientific grounding. Roles and collaboration plan are explicit.

[R4.2] Scientific: 2; Quantitatively links LiDAR filtering parameters to visibility metrics and site-placement hypotheses. This could revise long-standing assumptions about Chachapoya fortification and signaling networks. Clear research question, measurable outcomes.

[R4.3] Visualization: 2; Interactive side-by-side rendering of raw, filtered, and bare-earth DTMs for viewshed modeling is innovative. The design directly addresses interpretive uncertainty by enabling archaeologists to see how preprocessing changes visibility. Interface sketches and described controls (slider-based parameter exploration, overlay toggles) are concrete.

[R4.4] Significant: Not provided

[R4.5] Novel: 3; Prior work (e.g., Doneus 2008; Richards 2021) visualized visibility results but not preprocessing effects. VisArch adds an explanatory layer between data processing and interpretation, an area rarely addressed in archaeological visualization.

[R4.6] Goals clearly stated: 3; Timeline—prototype by Week 3, pilot testing by Week 6—is realistic. Dataset already acquired. User testing limited to domain collaborators, which is feasible.

[R4.7] Likelihood of Success: 2; Balanced technical and domain expertise; prior collaboration on LiDAR datasets ensures productivity.

### [R4.8] Strengths:

- Excellent problem framing with direct scientific payoff.
- Concrete visualization methods and evaluation design.
- Feasible timeline and authentic data.

### [R4.9] Weaknesses:

- Uncertainty visualization not explicitly addressed (e.g., error bands in DTMs).

**Response:** *The final version of the proposal includes the stretch goal of uncertainty visualization.*

## 2 Aims

We propose *VisArch*, a visualization tool for evaluating how Light Detection and Ranging (LiDAR) preprocessing choices affect intra-site visibility at Chachapoya hilltop settlements. In densely vegetated regions, raw LiDAR surfaces often conflate standing architecture with vertical vegetation, reducing confidence in visibility models. Our goal is to clarify how different preprocessing decisions shape local spatial conclusions and make these effects transparent through interactive visualization. To do so, we aim to:

1. Evaluate how different LiDAR preprocessing methods affect structure-to-structure visibility and site interpretation.
2. Build and evaluate a tool to visualize and compare the interpretive consequences of those preprocessing choices.



Figure 1: Preliminary mock-up of the visualization dashboard showing three LiDAR preprocessing outputs (Raw, Random Forest, and XGBoost) on a synthetic test site. Interactive controls are illustrative only.

To address the first goal, we will classify LiDAR point clouds using extreme gradient-boosted (XGBoosted) trees in R, trained on multiscale geometric features extracted with lidR to separate architecture from vegetation. We chose XGBoost because it is well-suited for complex datasets like LiDAR point clouds. For example, each point is described by multiple geometric features, such as slope and curvature, creating a high-dimensional space. Additionally, there are fewer architecture points than vegetation or ground points, making the dataset imbalanced. XGBoost has proven effective under these conditions.

Then, from the classified point cloud, we will generate cleaned surface models using terra. We will apply this process to a well-documented Chachapoya settlement and compare results to two baselines: (1) the unprocessed raw LiDAR elevation surface, and (2) the Random Forest pipeline published by Abate et al. [1]. Quantitative evaluation will include calculating metrics such as precision, recall, and F1 score on manually labeled subsets, alongside k-fold cross-validation to test model robustness. We will also visualize classification uncertainty to identify areas of lower confidence, with the exact threshold for uncertainty visualization determined after initial results.

For the second goal, we will develop a comparative dashboard (R flexdashboard) that integrates raw surfaces, Random Forest outputs, and XGBoost classifications into a unified visual environment. The dashboard will display surfaces alongside corresponding intra-site visibility maps, allowing archaeologists to evaluate how preprocessing decisions alter modeled visibility within the site. The visualization contribution lies in the design of an interactive dashboard that enables layer toggling, color-based visual encodings for classification confidence, and side-by-side comparative map alignment.

We will gather informal feedback from a group of 4-6 users, including domain experts (e.g., Dr. Parker VanValkenburgh) and non-experts (e.g., classmates). While the project will be tested with both expert and non-expert audiences, the intended primary users are archaeologists and researchers. Non-expert feedback will focus on usability and layout, while expert input will target analytical clarity and interpretive accuracy. To diversify perspectives, we will also seek input from archaeologists working in other subfields. Feedback will be primarily qualitative, supplemented by simple 1–5 rating metrics on clarity, usability, and interpretability.

As a stretch goal, we will incorporate uncertainty visualization that highlights areas of lower classification confidence, determined post-evaluation.

### 3 Significance

Archaeologists use visibility analysis to understand how ancient populations selected settlements and how structures may have been visually related within sites. In densely forested regions, where walls and structures are often obscured by vegetation, airborne LiDAR is one of the only tools that can reveal these features. However, visibility models often rely on raw LiDAR elevation surfaces, which do not reliably distinguish vegetation from built architecture. This limitation reduces confidence in local visibility results and may obscure important spatial relationships between structures.

This project introduces a visualization tool that makes the interpretive consequences of LiDAR preprocessing on intra-site visibility transparent. Existing approaches typically focus on classification accuracy or feature detection, but they rarely reveal how preprocessing choices alter later interpretations. By integrating machine learning (ML)-based classification into visibility modeling, researchers can compare spatial outcomes derived from raw and cleaned surfaces. The tool thus links technical preprocessing decisions directly to shifts in visibility networks and interpretive conclusions. In doing so, it contributes to archaeological research and to studies on transparency and interpretability in scientific modeling.

Improved visibility modeling may also help uncover local architectural relationships associated with the Chachapoya culture. Although the limited LiDAR data do not support regional intervisibility, this project establishes a framework for site-level visibility analysis that can inform future landscape-scale studies. Because Spanish colonization disrupted and dismissed Indigenous communication systems, much of what is known about their Chachapoya communication systems is limited. Additionally, insights could broaden attention beyond well-known destinations like Machu Picchu, supporting preservation and interpretation of less-studied landscapes.

### 4 Related Work

Airborne LiDAR has become an essential tool in archaeological research, particularly for detecting structures beneath vegetation. Štular et al. [4] emphasize that although LiDAR visualizations are often treated as straightforward products, they are the result of many subjective processing decisions. Their QGIS toolbox was designed to provide a transparent, archaeology-specific pipeline for point cloud classification and interpolation. By standardizing these steps, their work



Figure 2: Example of a reconstructed Chachapoya house showing original foundations and rebuilt structure.

highlights the efficiency of automated workflows and the interpretive risks of opaque preprocessing. Our project builds on this focus on transparency but goes further by providing an interactive comparative visualization that explicitly shows how preprocessing differences alter interpretive outcomes.

To reduce subjectivity, machine learning classifiers have been introduced into archaeological LiDAR analysis. Abate et al. [1] employed Random Forest models within CloudCompare to detect anthropogenic features at the Kastrí-Pandosia site in Epirus, Greece. They demonstrated that classifiers can outperform manual segmentation, but their evaluation was limited to reporting accuracy and did not explore consequences for archaeological interpretation. VisArch extends this approach by visualizing interpretive changes directly, emphasizing the downstream analytical effects rather than classification precision alone. Additionally, unlike Abate et al., our project tests a newer ensemble model (XGBoost) within the same comparative framework to evaluate whether it improves interpretability.

Serrano et al. [3] extended this line of work by applying XGBoost to large-scale digital elevation model (DEM) comparisons at the Valdecañas Reservoir in Cáceres, Spain. Their study showed strong predictive performance and provided insight into feature importance, suggesting the potential of ensemble methods for LiDAR analysis. Yet, as with earlier approaches, the focus remained on performance metrics rather than interpretive outcomes. Our work differs by assessing interpretive visibility networks and communicating these changes through visualization rather than statistical summaries.

Visualization techniques represent another critical dimension. Jaturapitpornchai et al. [2] compared multiple DEM-derived visualizations in semantic segmentation tasks and found that performance could vary by as much as 25% depending on the chosen visualization. Their findings demonstrate that visualization itself can bias detection and reinforce the view that LiDAR images must be treated as interpretive representations rather than neutral evidence. Similarly, Štular et al. [4] stress that LiDAR images are not “hard” data but require contextual interpretation. VisArch builds on these insights by treating visualization as an analytical tool in itself, using comparative layouts and coloring to expose uncertainty introduced by preprocessing choices.

Together, these studies establish advances in the use of LiDAR for archaeological classification and visualization. However, none explicitly evaluate how different classification approaches affect the outcomes of intra-site modeling. This project responds to this gap by integrating XGBoost-based classification with a comparative visualization framework comparing raw, Random Forest, and XGBoost outputs.

## 5 Research Plan

We propose a 6-week plan:

- **Week 1:** Prepare and clean LiDAR data for one well-documented Chachapoya site and label a small subset of points for training. *Deliverable: initial labeled dataset.*
- **Week 2:** Train XGBoost classifier and compare to Abate et al.’s Random Forest baseline. Generate cleaned surfaces and establish evaluation metrics using labeled subsets. *Deliverable: validated classification outputs and performance metrics.*
- **Week 3:** Generate DTMs and compute structure-to-structure visibility for each surface. *Deliverable: visibility networks for raw, Random Forest, and XGBoost surfaces.*
- **Week 4:** Develop prototype R flexdashboard. Integrate surfaces and visibility maps, including interactive features such as layer toggling, zoom, and visibility overlays. *Deliverable: working dashboard prototype with core visualization functions.*
- **Week 5:** Gather feedback from 4–6 users (experts and non-experts). Evaluate the dashboard’s usability and interpretive clarity using qualitative comments and 1–5 rating metrics. Make interface revisions based on results. *Deliverable: feedback summary and revised interface.*
- **Week 6:** Finalize dashboard, summarize classification and visualization results, and prepare final report and presentation. *Deliverable: final comparative dashboard and complete project report.*

The plan is ambitious yet achievable given the available data and the team’s expertise. Dr. VanValkenburgh will provide access to the necessary LiDAR datasets. The machine learning components will be implemented using existing libraries, such as `lidR`, `xgboost`, and `terra`. Dashboard development via `flexdashboard` replaces the previously planned Shiny app, aligning better with available data and project scope.

## References

- [1] N. Abate et al. An open-source machine learning-based methodological approach for processing high-resolution uas lidar data in archaeological contexts: A case study from epirus, greece. *Journal of Archaeological Method and Theory*, 32:1–34, 2025.
- [2] R. Jaturapitpornchai et al. Impact of lidar visualisations on semantic segmentation of archaeological objects. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 3361–3365, 2024.
- [3] A. Serrano et al. Classification of archaeological lidar data using xgboost: A case study from spain. *Remote Sensing*, 17:1306, 2024.
- [4] B. Štular et al. Airborne lidar point cloud processing for archaeology. pipeline and qgis toolbox. *Remote Sensing*, 13:1–37, 2021.

# CHARLY CASTILLO

Weehawken, NJ | (201) 682-5540 | [charly\\_castillo@brown.edu](mailto:charly_castillo@brown.edu) | [LinkedIn](#)

## EDUCATION

---

- Brown University**, Providence, RI Expected Graduation 05/2026
- *Sc.B. in Statistics and International and Public Affairs, Security*
  - **D.E. Shaw Discovery Fellow:** 1 of 49 sophomores nationwide selected for all-expenses paid program at DESCO HQ
  - **Undergraduate Teaching and Research Award:** Fall '23; 1 of ~100 Brown students awarded \$1,200 research stipend
  - **Relevant Coursework:** *Undergraduate:* Stochastic Calculus, Computational Probability, Modeling, Optimization, Theoretical Linear Algebra, Scientific Computing, Regression; *Graduate:* Scientific Visualization, Cybersecurity

## WORK EXPERIENCE

---

- Bank of America**, *Quantitative Analyst Intern*, New York, NY 06/2025–08/2025
- Evaluated 2024 policy replacing aging fees with desk-specific aged inventory limits, applying ITS analysis to measure shifts across BofA's 6 Global Credit flow desks (ASIA, CEEMEA, EMEA, LATAM, US HY, US IG).
  - Found 15-37% reductions on EMEA and CEEMEA desks; normalized aged inventory by book size across 12 US IG traders using coefficients of variation; proposed orthogonal random forest to isolate overlapping policy effects.

- JPMorgan Chase & Co.**, *Investment Banking Summer Analyst*, New York, NY 06/2024–07/2024
- Analyzed \$11B cross-border merger, assessing strategic fit by through revenue segmentation and market scope.

- Brown University Department of Biostatistics**, *Research Intern*, Providence, RI 09/2023–12/2023
- Benchmarked 46 distance and similarity metrics using philentropy to evaluate accuracy in matching multi-part names across datasets without unique identifiers; simulated 10 data corruption scenarios (substitutions, insertions, deletions, transpositions) on 1,000+ records using stringr and purrr to assess robustness under realistic data issues.

- New Jersey Cybersecurity and Communications Integration Cell**, *SECOPS Intern*, Newark, NJ 06/2023–07/2023
- Developed logistic regression pipeline with scikit-learn to classify 10,000+ algorithmically generated domains (AGDs), engineering 10 predictive linguistic features (e.g., consonant length, domain length, vowel/consonant ratio).
  - Achieved 64% precision in correctly identifying AGDs; improved model performance by 8% (F<sub>1</sub> score) through hyperparameter tuning with GridSearchCV; benchmarked against Random Forest and Gradient Boosting classifiers.

## LEADERSHIP EXPERIENCE

---

- Brown Office of Undergraduate Admissions**, *Student Head*, Providence, RI 02/2024–Present
- Lead campus tour program and conduct 3 weekly campus tours for audiences of up to 180; represent Brown as a keynote speaker at ADOCH, the official [virtual tour](#), and on Brown's official Instagram account (25k+ followers).

- Brown Statistics Department Undergraduate Group**, *President*, Providence, RI 10/2023–Present
- Organize academic events (e.g., alumni panels, course registration Q&As) attended by 200+ students annually; manage \$1,500 budget and negotiate program funding through UFunds to expand department programming.

## PROJECTS

---

- Total Variation Image Denoising via ADMM**, *Class Project* 04/2025–05/2025
- Applied ADMM with Gauss-Seidel iterations to Rudin-Osher-Fatemi total variation denoising; evaluated convergence (L<sub>2</sub> norms, iteration counts) and output quality across eight regularization parameters on grayscale images.

- Estimating U.S. Inflation Divergence from G7 Economies**, *Independent Project* 05/2024–09/2024
- Analyzed U.S. inflation trends using Difference-in-Differences, Bayesian Structural Time Series (CausalImpact), and Synthetic Control (Synth); processed IMF data in R using fixed effects regressions and counterfactual simulations.
  - Estimated inflationary effect of 30-140 basis points attributable to U.S. policies, including global spillover effects.

## SKILLS

---

- *Technical Skills:* Python (pandas, NumPy, scikit-learn, statsmodels, JAX, SciPy, Plotly), R, MATLAB, Julia, SQL
- *Tools:* IBM SPSS Statistics, Excel (PivotTables, VBA), Bloomberg, Git, LaTeX, Google Workspace
- *Languages:* English (native), Castilian Spanish (native), Galician (native), French (intermediate), Portuguese (basic)



# SIR DYLAN ALLOTEY

---

Bronx, New York, 10453 | [sirdylan.allotey1@gmail.com](mailto:sirdylan.allotey1@gmail.com) | (646) 773-9172 | [LinkedIn](#)

## EDUCATION

---

**Brown University**, Providence, RI

**Expected Graduation May 2026**

*Master of Science in Computer Science*

**Brown University**, Providence, RI

**Graduated May 2025**

*Bachelor of Science in Computer Science* GPA: 3.75

Honors: Sydney Frank Scholarship (2021), Zeta Boule Scholarship (2021)

## PROFESSIONAL EXPERIENCE & LEADERSHIP

---

**Modern Railway Systems**, Littleton, CO; Windsor, CA; Los Angeles, CA

**May 2023 – Present**

*Systems/Software Engineering Intern*

- **Fall 2025** - Maintained backend systems for applications managing live trains
  - Designed and optimized database update mechanisms
  - Implemented access controls and safety restrictions
- **Summer 2025** – Contributed to multibillion-dollar new rail infrastructure project for LA Metro
  - Performed duct bank cross-section calculations
  - Coordinated with contractors, stakeholders, and multidisciplinary teams to support engineering tasks
- **Summer 2024** - Used data collections to design and test hardware systems for use in the field
  - Developed software in Python to model aspects of vehicles and city plans such as approach velocity and warning time
  - Designed and wired several routers and switches to achieve redundancy and cyclic connections in the network
- **Summer 2023** - Worked with engineering teams to model several road conditions to create safe track speed restrictions for trains using kinematics
  - Designed router systems and CAD drawings to test and improve the current system's redundancy
  - Developed software in Python to represent data including time, costs, and resources a project

## PROJECTS

---

**Ray Tracer**

**December 2023**

*Developer*

- Designed ray tracer to build 3d scenes like Maya
- Rendered navigable scenes using primitive shapes (cone, cylinder, cube, sphere)
- Implemented Phone lighting reflection model
- Coded in C++ using the QT creator framework and the Open Graphics Library

**Computer Shell**

**December 2023**

*Developer*

- Coding in C with Ubuntu and Docker
- Featured memory allocation, extensive error-checking, background processes, and signal handling

## SKILLS

---

**Computer:** Proficient in Microsoft Applications (Excel, Word, & PowerPoint), Intermediate in Blender and Maya, Building Computers

**Languages:** Basic in Korean

**Coding Languages:** Korean Basic (Writing, reading, and speaking), Proficient in Java, Python, C, C++



**Castillo, Charly** <charly\_castillo@brown.edu>  
to Parker ▾

Fri, Oct 10, 1:57 AM (3 days ago) ☆ ↶ ⋮

Hi Parker,

Thank you, this is very helpful, especially the image. I also wanted to confirm that you're okay with collaborating in the capacity we discussed, mainly by offering feedback in the final stages. I need to note that confirmation in an upcoming assignment submission.

Thanks again for your help so far!

Best,  
Charly

...



**VanValkenburgh, Parker**  
to me ▾

Fri, Oct 10, 10:06 AM (3 days ago) ☆ ↶ ⋮

Absolutely!

...

# Revealing Patterns of Public Trust: A Comparative Visualization of Government Agencies

Gordan Milovac  
PI

`gordan.milovac@brown.edu`

Shravya Munugala  
Co-PI

`shravya.munugala@brown.edu`

Bertram Malle  
Collaborator  
`bfmalle@brown.edu`

October 9, 2025

## **Abstract**

We propose an interdisciplinary project to visualize patterns of public trust in U.S. government agencies. By combining survey-based trust dimensions with interactive visualization techniques, we aim to reveal comparative trust dynamics across agencies. A collaboration between computer science and social psychology, the project will evaluate novel visual encodings through user studies. Beyond this specific case, we seek to design a generalizable framework for visualizing any multidimensional dataset, thereby broadening the impact of our work across domains.

# 1 Response to Reviews

We appreciate the time and effort put into reviews and have included responses to reviewers' feedback on the attached proposal.

## 1.1 Haonan Zhang

### **Overall:** 2

A well-structured, interdisciplinary, and clearly motivated proposal with high potential for impact. Slightly ambitious for the 6-week scope but overall very strong.

### **Interdisciplinary:** 1

Excellent integration of computer science and social psychology. The collaboration effectively combines data visualization methods with conceptual modeling of public trust.

### **Scientific:** 2

Grounded in relevant literature with clear methodological reasoning. Could further specify statistical validation or data-processing details.

### **Visualization:** 2

Strong focus on multidimensional, interactive visualization using glyphs and radial encodings. Promising approach, though perceptual clarity and user-load management will be important to evaluate.

### **Significant:** 2

Addresses an important social-science challenge—making trust data interpretable and comparable—and offers generalizable visualization principles.

### **Novel:** 3

New in context and application; however, the chosen glyph and radar metaphors have known perceptual limitations that should be acknowledged and mitigated.

### **Goals clearly stated:** 1

Objectives, milestones, and evaluation strategy are clearly articulated and logically sequenced.

### **Likelihood of Success:** 2

Timeline and deliverables are realistic, and the team's technical background is strong. Main risk is limited time for user-study recruitment and analysis.

***Response:** We acknowledge this potential bottleneck. To mitigate this risk, we have updated the Research Plan and Aims sections to clarify that our collaborator, Professor Bertram Malle, will manage the recruitment and compensation of study participants. This streamlines the process, allowing us to stay on schedule and focus our efforts on the analysis and development phases.*

### **Strengths:**

- True interdisciplinary collaboration between visualization and psychology.
- Timely and socially meaningful topic (public trust) with potential generalization to other multidimensional datasets.
- Explicit inclusion of a user study to assess interpretability and usability.

### **Weaknesses:**

- The claim of a “generalizable framework” would benefit from clearer design.

***Response:** We agree that this point needed more detail. The proposal has been revised to clarify the design of the generalizable framework. We now specify in the Aims and Research Plan sections that the system will be built using a parametric approach, consisting of a modular data-processing*

*pipeline and adaptable visualization components (e.g., radial plots) that can be reconfigured for other multidimensional datasets with different structures.*

## **1.2 Juliana Escuti**

### **Overall: 2**

This proposal is both a novel visualization method and a novel application in this field.

### **Interdisciplinary: 1**

This proposal combines visualization with psychology and political science.

### **Scientific: 2**

This framework extends the applicability of the visualization system described. It will also clearly evaluate, using surveys, to what extent the new multi-dimensional visualization system is to use.

### **Visualization: 3**

Enabling the visualization of multivariate data on trust profiles is a novel application. Developing a framework for multivariable glyphs, radial plots, and comparative grids will reveal new insights into this data.

### **Significant: 3**

Visualizing and comparing trust between government agencies is very important. Additionally, this framework can extend to other applications outside of politics/psychology.

### **Novel: 4**

How will these new multivariate trust profiles improve upon previous versions? Is there a way to make the similarity between trust profiles visually evident?

***Response:** This is an excellent point. The revised Aims section now details a dual-visualization strategy to make similarities visually evident. We specify that we will use UMAP to provide a high-level overview that clusters agencies with similar trust profiles, making broad patterns immediately apparent. For detailed analysis, we will use interactive, overlaying radial graphs that allow for direct, multi-attribute comparisons between a small selection of agencies.*

### **Goals clearly stated: 1**

This proposal clearly states its goals and timeline.

### **Likelihood of Success: 2**

The visualization part of this proposal seems doable in the given timeframe.

### **Strengths:**

This proposal clearly states how a multivariate visualization method would be helpful in this field. It proposes three methods for visualization, and will evaluate to what extent they are effective and usable using surveys. Especially if it is interactive and online for anyone to use, this could be a valuable tool for visualizing trust in political agencies.

### **Weaknesses:**

It is hard to visualize how these new multivariate visualizations would look different from current visualizations. How will the glyphs/Chernoff faces be mapped? How will it better show relationships between the data for each agency?

***Response:** Thank you for this feedback. To improve clarity, we have revised the proposal to de-emphasize a broad exploration of glyphs like Chernoff faces and state a clear focus on our primary methods. The revised Aims and Related Work sections now explain that relationships will be shown through UMAP's spatial clustering and the direct comparison facilitated by the interactive radial graphs, which we believe are more effective for this specific dataset.*

### 1.3 Korey Sam

**Overall:** 4.8 / 5

This proposal combines social science rigor with computational visualization design in a way that is both ambitious and feasible.

**Interdisciplinary:** 5 / 5

The focus in Computer Science and social psychology, and political science provides the visualization methods and theoretical grounding in trust metrics. Collaboration with Bertram Malle, a leading expert in moral cognition and trust, ensures conceptual rigor and support for achievability.

**Scientific:** 4.6 / 5

The scientific motivation is clear and methodologically strong. The use of user studies and graphical perception tests grounds the evaluation in empirical methodology. A small limitation is that the proposal doesn't specify the sample size or statistical analysis plan for the user study.

**Visualization:** 5 / 5

It demonstrates extensive background in visualization theory referencing Chernoff faces, radial plots, glyph design, and perceptual bias research.

**Significant:**

Public trust in institutions is an issue of immense civic and scientific importance. By making trust data interpretable and explorable, this project offers real contributions to policy communication and media literacy in a time of significant mistrust of government institutions. There are also significant generalization opportunities to other domains (health, consumer behavior, psychology), amplifying its potential impact.

**Novel:** 4.5 / 5

The conceptual novelty lies in fusing multivariate visualization design with psychological constructs of trust.

**Goals clearly stated:** 5 / 5

The aims are explicit and coherent: build a visualization platform, prototype multiple encodings (glyphs, radial wheels, comparative grids), and evaluate through user studies.

**Likelihood of Success:** 4.3 / 5

The main constraints are time and the scope of user evaluation to ensure the product is viable. A small-scale but well-designed usability study should still yield strong insights.

**Strengths:**

- Deeply interdisciplinary, integrating visualization science, social psychology, and political communication.
- Ambitious yet concrete deliverables (interactive prototypes + empirical validation).
- Explicit plan for generalizability, increasing long-term research value.

**Weaknesses:**

- User study design could be elaborated (sample size, statistical power, tasks).

***Response:** We have expanded the description of the user study in the Aims and Research Plan sections. The proposal now specifies a sample size of approximately 100 paid participants. We have also clarified that the study will be coordinated by Professor Bertram Malle and will involve tasks where participants directly compare our visualizations against baseline methods to measure accuracy and intuitiveness.*

- Tight 6-week schedule may limit polish and robustness of final web platform.

***Response:** We recognize this constraint. By having Professor Malle manage the logistics of the user study (as noted in our response to 1.1), we can dedicate more focused development time to ensuring the final prototype is robust and functional for the purposes of our evaluation, even within the tight schedule.*

## 1.4 Kyle Wisialowski

**Overall:** 2-3

**Interdisciplinary:** 1

Collaboration between visualization, social psychology and political science.

**Scientific:** 3

Discovery of latent dimensions of trust, expose agency similarities / differences, generate new comparisons that were previously impossible with 2D models.

**Visualization:** 2

Design a generalizable framework for visualizing any multidimensional dataset.

**Significant:** 3

More accurate / interpretable visualizations of trust + methods of extrapolating these techniques to other domains.

**Novel:** 3

Previous work uses glyph-based designs (like Chernoff faces). This builds upon these strategies by using novel tools like radial plots and comparative grids with interactive elements like linked views and filtering.

**Goals clearly stated:** 4

The central goal of the project is to improve existing methods of conveying trust in various government agencies by exploring Chernoff faces, radial plots, and comparative grids.

**Likelihood of Success:** 3

**Strengths:**

This is an interdisciplinary project that combines visualization, social psychology, and political science. The PI has done a great job explaining the problem statement and current approaches to conveying trust visually. They have proposed several interesting ideas about how to improve current visualizations and have made it clear that their methods will be relevant to domains outside of political science. There are clear gaps in current work that this project seeks to resolve.

**Weaknesses:**

- Though the proposal does a great job proposing solutions to the problems at hand, I wish there was a bit more specificity in describing the goals. For example, “We will prototype and refine encodings such as Chernoff faces [3], radial trust wheels, and comparative agency grids, integrating them into the web platform.” I think statements like this could be solidified if the PI made some stronger choices about which techniques they plan to use and how they plan to implement them, rather than just throwing them out as possible avenues as exploration. If the PI was more specific about the visualization strategies, I would have rated the goals + likelihood of success a bit higher.

***Response:** We have revised the proposal to reflect a more decisive and focused research plan. The Aims and Related Work sections now clearly state that our primary implementation will concentrate on two specific techniques: UMAP for high-level pattern discovery and interactive radial graphs for detailed comparison, rather than a general exploration of all possible encodings.*

- Also, the significance has potential to be even stronger, too. Why should I even care about trust in government agencies? That is a question I would consider answering in the final proposal.

***Response:** We have strengthened the Significance section to more forcefully articulate the project’s real-world importance. The revised text now explains that tracking trust provides a powerful tool for predicting public behavior(e.g., economic confidence, policy compliance) and serves as an invaluable feedback mechanism on government performance through consistent monthly and yearly comparisons.*



Figure 1: Radial trust profiles for agency groupings; similarity is suggested but comparisons remain unclear.

## 2 Aims

The primary aim of this project is to design and implement an interactive visualization system that reveals comparative levels of public trust in different U.S. government agencies. Existing survey-based visualizations often struggle to represent complex, multidimensional trust metrics in a way that allows users to intuitively compare agencies or understand underlying patterns. Current displays are frequently incomplete, omitting some agencies or key dimensions, and in general lack intuitive design, making it difficult for researchers and the public to glean actionable insights (Figures 1, 2a and 2b). Our approach will explore a range of multidimensional visualization techniques, drawing inspiration from foundational work on glyph-based methods like Chernoff faces and comparative grids [3, 2]. While these methods will be considered, our primary focus will be on implementing interactive and comparative visualizations. Specifically, we will concentrate on using UMAP to reveal high-level structural patterns across all agencies [6, 8] and developing interactive, overlaying radial graphs to facilitate detailed, multi-attribute comparisons between a small selection of agencies [11].

By enabling these novel interactions, this project moves beyond static representations to provide a dynamic tool for exploratory data analysis. The key relevance of this work lies in its ability to make complex social data transparent and actionable. For researchers and policymakers, this system will facilitate the discovery of latent dimensions of trust and reveal subtle inter-agency similarities and differences that are currently obscured. For the public, it offers an intuitive way to engage with and understand the nuances of institutional trust, fostering greater data literacy. Furthermore, a central component of the project is the development of a system that generalizes to other multidimensional datasets. This will be achieved through a parametric approach, allowing the framework to be applied broadly in future data research across domains



that involve complex multivariate relationships. We will incorporate interactive elements, such as linked views, filtering, and dimensionality controls, to provide users with both overview and detail, improving interpretability for non-expert audiences [7, 12].

To evaluate the effectiveness and usability of our visualizations, we will conduct a rigorous user study involving approximately 100 paid participants. The recruitment and compensation for this study will be coordinated by our collaborator, Professor Bertram Malle, ensuring a timely and organized process. Participants will be asked to directly compare our new visualizations against baseline methods, and the study will measure not only how accurately they can interpret trust relationships but also their subjective assessments of intuitiveness and clarity, drawing from established methodologies in graphical perception [5]. The resulting insights will guide refinements to the system and provide empirical evidence of its value. Overall, the project aims to: visualize complex relationships within survey data from 30 government agencies, represent multidimensional trust metrics more clearly than existing techniques, generalize the framework to diverse datasets, and validate effectiveness through rigorous user testing. In doing so, we aim to produce a tool that is both scientifically useful and practically accessible for research, policy, and public engagement.

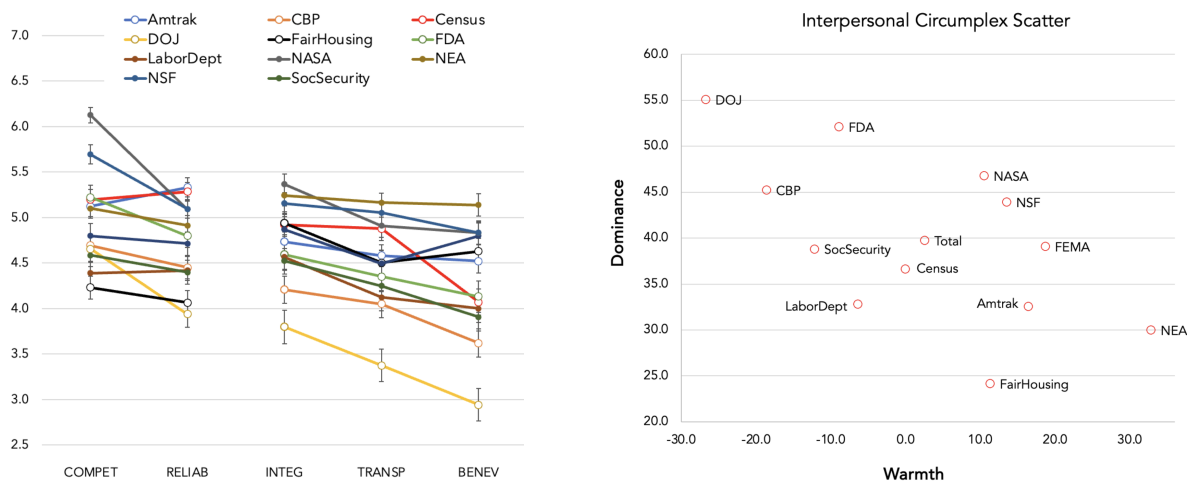


Figure 2: Visualizations of trust relationships (a) and interpersonal styles (b) across 16 agencies. Both displays become complicated and unintuitive even with fewer than the full set of agencies.

### 3 Significance

Trust in institutions is a central theme in political science, sociology, and psychology, influencing civic engagement, policy compliance, and social cohesion. The ability to track these trust metrics on a monthly or yearly basis provides invaluable insight into how public sentiment shifts in response to government decisions, offering a powerful tool for analysis and accountability. Despite its importance, survey data on public trust is frequently presented in simple bar charts, tables, or other static visualizations that fail to capture the multidimensional nature of trust judgments across different agencies. These traditional displays often omit agencies or key dimensions and provide little opportunity for users to explore the relationships among trust metrics in a meaningful way (Figures 1, 2a and 2b). This limitation can obscure latent patterns, prevent accurate comparisons, and reduce the overall interpretability of survey results.

Our project addresses this gap by introducing interactive visualizations that leverage UMAP to provide a high-level overview of agency clusters and interactive, overlaying radial graphs to reveal the subtleties of trust profiles across U.S. government agencies. By enabling side-by-side comparisons and interactive exploration, users can observe emergent patterns and interdependencies that are difficult to perceive in traditional static charts. This approach bridges disciplines, combining computational techniques for high-dimensional data visualization with conceptual frameworks from psychology and political science to ensure that representations of trust are both accurate and meaningful.

A further contribution of this work is the creation of a reusable visualization framework that generalizes beyond trust data. The system is designed to accommodate any multidimensional dataset, whether it concerns personality scales, product comparisons, health metrics, or other survey-driven research. By providing tools for interactive exploration and comparison, our project facilitates more nuanced interpretations, encourages responsible use of complex data, and supports broader efforts in data science to make multidimensional information comprehensible to both researchers and non-experts [5, 12]. Ultimately, the significance of this project lies not only in producing more accurate and interpretable visualizations of trust but also in establishing methods and principles that can enhance the study and communication of complex relationships across a wide range of domains.

## 4 Related Work

The challenge of representing multidimensional data has a rich history in the visualization community. Foundational work, such as Chernoff’s use of facial features [3], demonstrated that carefully designed visual metaphors could reveal complex data structures. Broader surveys on glyph-based design have since provided principles for encoding multiple variables into single perceptual objects [2, 10]. While our project will review these foundational ideas, our implementation will focus on modern, interactive techniques that are better suited for dynamic data exploration.

Our primary approach is informed by recent empirical studies that evaluate the efficacy of specific visualization types. For detailed comparisons, we will focus on interactive radial charts. The literature highlights that while static radial visualizations can suffer from perceptual bias, their effectiveness is significantly improved through careful design and interactivity [11, 1, 9]. We will directly incorporate the findings from these comparative papers, using their conclusions as a driving force for our design to mitigate known issues and enhance user comprehension. This evidence-based approach ensures our visualizations are built on proven principles of graphical perception [5].

For providing a high-level overview of the agency landscape, we will leverage UMAP, a powerful dimensionality reduction technique. Research in other complex domains, such as genomics, has shown UMAP’s exceptional ability to reveal cryptic population structures and heterogeneity from large datasets [6]. By applying this to trust data, we anticipate uncovering novel groupings and relationships between agencies. We are also mindful of the potential pitfalls of such techniques and will follow best practices to ensure its appropriate application for visual analytics [8].

While techniques like UMAP have been used to analyze complex data from transformer models [4] or in biology, their application to social-psychological survey data remains underexplored. This project

situates itself at this intersection, combining advanced visualization methods with the conceptual rigor of trust research, aided by consultation with experts like Prof. Bertram Malle. Furthermore, a key contribution is our focus on a generalizable framework. Whereas many systems are domain-specific, our design will be reusable for other multidimensional datasets, extending the applicability of these advanced visualization principles to new areas of social science research.

## 5 Research Plan

The proposed six-week project will proceed in three phases, each focused on a major milestone of development and evaluation.

### **Weeks 1–2: Foundation and Infrastructure.**

We will begin by reviewing relevant visualization literature, with a focus on multidimensional encoding and graphical perception. During this phase, we will also plan the system architecture, which will include a modular data-processing pipeline and adaptable visualization components, and build a website to serve as the platform for the interactive visualizations.

### **Weeks 3–4: Visualization Development and Generalization.**

The core focus of this phase will be designing and implementing interactive visualizations of trust across government agencies. We will prototype and refine encodings such as radial trust wheels and UMAP, and explore possibilities with Chernoff faces [3] and comparative agency grids, integrating them into the web platform. At the same time, we will extend the framework to support generalization to other multidimensional datasets (e.g., personality measures or product ratings), ensuring broad applicability beyond the initial case study. To achieve this, we will ensure that the data ingestion pipeline and visualization components are built parametrically, allowing them to adapt to new datasets with different structures and dimensions.

### **Weeks 5–6: Evaluation and Dissemination.**

In the final phase, we will coordinate with Prof. Bertram Malle to design and conduct a qualitative user study, where 100 participants will directly compare our visualizations against baseline methods. The evaluation will focus on intuitiveness, interpretability, and usefulness. To mitigate the risk associated with participant recruitment, our collaborator, Prof. Bertram Malle, will manage the recruitment and compensation of study participants. Based on these results, we will prepare a final paper, develop a polished presentation, and demonstrate the website to showcase the research outcomes.

We will use standard facilities: Python for data preprocessing, D3.js/Plotly for interactive visualization, and survey distribution platforms for user evaluation. The website will serve as both a research tool and dissemination medium, enabling others to explore and build on our approach.

## References

- [1] H. I. M. M. Abeynayake et al. Efficacy of information extraction from bar, line, circular (radar), and other chart types. *Open Access*, 2023.
- [2] R. Borgo, J. Kehler, D. H. S. Chung, E. Maguire, R. S. Laramée, H. Hauser, and M. Chen. Glyph-based visualization: Foundations, design guidelines, and applications. *Computer Graphics Forum*, 2013.
- [3] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973.
- [4] S. Chung and L. F. Abbott. Neural population geometry: An approach for understanding biological and artificial neural networks. *Current Opinion in Neurobiology*, 70:137–144, 2021.
- [5] W. S. Cleveland and R. McGill. Graphical perception: theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.
- [6] A. Diaz-Papkovich, L. Anderson-Trocmé, C. Ben-Eghan, and S. Gravel. Umap reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLOS Genetics*, 15(11):e1008432, 2019.
- [7] E. E. Firat et al. P-lite: A study of parallel coordinate plot literacy. *Open Access*, 2022.
- [8] H. Jeon et al. Stop misusing t-sne and umap for visual analytics. *arXiv*, 2025.
- [9] S. Kantabutra. Perceptual bias in mobile radar chart visualization. *International Journal of Analytic Applications*, 2025.
- [10] T. Ropinski et al. Survey of glyph-based visualization techniques for spatial multivariate data. *Computer Graphics Forum*, 2011.
- [11] A. Svalina et al. Assessing the design of interactive radial data visualizations. *Open Access*, 2023.
- [12] E. J. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990.

# Gordan Milovac

+1 (401) 471-9866 | [gordan\\_milovac@brown.edu](mailto:gordan_milovac@brown.edu) | [linkedin.com/in/gmilovac](https://www.linkedin.com/in/gmilovac) | [gordanmilovac.com](https://gordanmilovac.com)

## Education

**Brown University**, Providence, RI

*Master of Science in Computer Science – AI/ML Track*

Sep 2025 - May 2026 (Expected)

*Bachelor of Science in Computer Science and Economics - Professional Track*

Sep 2021 - May 2025

- Sigma Xi Honors Society, 2023 LINK Award recipient, GAMMA, GPA: 3.7/4.0

## Experience

**Teaching Assistant**, [Brown University Department of Computer Science](#)

Jan 2025 - Present

- Assisted developing course content, Held labs and office hours, Graded and overseen 200+ students for Data Science

**Sales Engineer Intern**, [InterSystems Corporation](#)

May 2024 - Aug 2024

- Built an AI powered Synthetic Patient Data Generator, Gave over 5 Sales Presentations for the HealthShare team  
- Received special recognition from the department VP, Director and Manager, alongside a part-time extension offer

**Research Assistant**, [Brown University School of Engineering](#)

Jan 2024 - May 2024

- Devised techniques for aerial and ground robots to detect and map dangerous yet unexplored areas (FastSLAM)

**iOS App Developer**, [ScanEat](#)

Jun 2023 - Sep 2023

- Gained startup experience, Developed and tested iOS apps using Swift, Built UI features on a 100,000+ user app

**SWE and Multimedia Intern**, [DUMP Split](#)

Sep 2020 - Dec 2020

- Practiced full-stack web development (HTML/CSS, JS, React.js), Gained experience in photo and video editing  
- Helped organizing the annual DUMP Days conference, attracting 2000+ followers and max live appointments

## Projects

**AI-SDA Generator**, InterSystems Corporation

May 2024 - Aug 2024

- Website utilizes AI to generate and populate SDA medical records, reducing creation time from 40h to 5min  
- Integrated GPT-4o interaction commands using precise prompt engineering and parallel API calls (React.js)

**PeachInnovators**, Brown University

April 2024 - May 2024

- Gathered telemetry data for Brown Crew and modeled it to see which factors contribute to the boat speed (SQL)  
- Utilized the Pearson Correlation test to analyze the relationship of several key Peach system variables (Python)  
- Implemented a KERAS model utilizing a straightforward sequential architecture and 4-fold cross validation training

**AiRobic**, Brown University

April 2023 - May 2023

- Website generates a user-specific plan of rowing workouts, using model generation on the backend server (Java)  
- Algorithm utilizes Markov Model with a matrix between a set of hidden states, workouts stored in Firebase database

## Leadership

**D1 Athlete**, Brown Men's Varsity Crew

Sep 2021 - May 2025

- 25+ hours weekly commitment, Represented Brown at the Henley Royal Regatta and IRA National Championship  
- 2024 Ivy League Champion, Head of The Charles Regatta course record holder, 2024 Sri Chinmoy Marathon runner

**National Team Athlete**, Croatian Rowing Federation

Sep 2019 - Aug 2021

- 3x National champion, Indoor champion and Junior record holder, World and European Rowing Championship  
- Rowing Coach, Croatian Youth Olympic Team member, Awarded "Accomplished athlete" Trophy in 2020 and 2021

## Coursework & Skills

Intro to Programming (OOP), Program Design: Data Structures & Algorithms, Discrete Structures and Probability, Software Engineering, Computer Systems, Computer Graphics, Artificial Intelligence, Machine Learning (Capstone), Data Science, Cybersecurity and International Relations, Intermediate Calculus, Linear Algebra, Statistical Inference

**Languages:** Python, Java, C++, C, Swift, SQL, JavaScript, TypeScript, HTML, CSS, InterSystems ObjectScript

**Frameworks:** React, Next.js, Angular, Jest, NumPy, Pandas, PyTorch, Spark, JUnit, Firebase, Qt, OpenGL

**Tools:** Git, GitHub, Perforce, IRIS, VSCode, JetBrains IDE's, XCode, Adobe CC, PowerPoint, Excel, LaTeX, Linux

# SHRAVYA MUNUGALA

(401) 215-3306 — shravya\_munugala@brown.edu — linkedin.com/in/shravya29M — github.com/Shravya29M

## EDUCATION

### Brown University

Expected May 2027

Master of Science in Computer Science

Providence, RI

- Relevant Coursework: Deep Learning, Interdisciplinary Scientific Visualization

### Vellore Institute of Technology, Chennai

May 2025

Bachelor of Technology in Computer Science, GPA: 3.97/4

Chennai, India

- Relevant Coursework: Data Structures and Algorithms, Operating Systems, AI, DBMS, Software Engineering

## EXPERIENCE

### Cloud Computing Intern

April 2025 - July 2025

Pentabrain Solutions Pvt Ltd

Chennai, India

- Assisted in building **serverless applications** with AWS Lambda and API Gateway, optimizing DynamoDB pipelines to **reduce data access latency** by 20%.
- Implemented **CI/CD pipelines** with GitHub Actions and CloudFormation for **automated deployments**.
- **Collaborated with cross-functional teams** through sprint planning and standups, leveraging Jira to support event-driven workflows and **improve collaboration** in a fast-paced environment.
- Maintained documentation in GitHub, **enhancing knowledge transfer** and accelerating team onboarding.

### Research Intern

May 2024 - Dec 2024

Center for Healthcare Advancement, Innovation and Research

Chennai, India

- Developed a **custom dual deep learning framework** to classify schizophrenia from 3D sMRI data. Optimized a CNN with 5-fold cross-validation, boosting accuracy by 13% across datasets.
- Adapted the model to diverse medical imaging datasets, maintaining **high clinical reliability and scalability**.

## PROJECTS

### Elyssa – AI Mental Health Assistant | Python, Gradio, LLM, Huggingface, Kubernetes

Feb 2025 – July 2025

- Built **Elyssa**, a novel AI mental-health assistant fusing facial-expression analysis, sentiment-shift detection, and confidence-weighted response generation.
- Developed a **scalable, load-balanced system**, also improving response quality by 10.46% (ROUGE F1) and ensuring reliable, context-aware interactions.

### AutoNewsSummarizer | PyTorch, BART, BERT, BeautifulSoup, Scikit-learn, FastAPI

Dec 2023 – March 2024

- **Engineered two novel NLP models:** a sentiment- and topic-aware summarizer for emotionally rich content, and a hierarchical long-form summarizer with iterative refinement.
- **Deployed a real-time URL-to-summary pipeline**, generating context-aware summaries from live web content.
- Outperformed ROUGE and BLEU, **producing coherent**, recall-optimized summaries beyond standard models.

## TECHNICAL SKILLS

**Languages:** Python, Java, C, C++, SQL, R, MATLAB

**Frameworks:** TensorFlow, PyTorch, Keras

**Developer Tools:** Docker, Google Cloud Platform, AWS (EC2, S3), Git, Github, Linux

**Libraries & Specializations:** Scikit-learn, Pandas, NumPy, OpenCV, Deep Learning, NLP, Data Structures & Algorithms

**Certifications:** AWS Cloud Practitioner (2025), Google Cloud Digital Leader (2025)

## PUBLICATIONS

- Shravya Munugala, et al. "Alzheimer's Disease Classification Using CUDA-Accelerated DenseNet201 and VGG16", IEEE ICICNIS, 2024.
- H.P, Shravya Munugala, et al. "GenoCryptNet: The Future of Secure Communication Through DNA", IEEE ICSTSDG, 2024.

## LEADERSHIP

### Head of HR and Management at PlaceXP Club

Feb 2024 – Feb 2025

- **Led 7+ events** and alumni panels, engaging 100+ students on college and internship opportunities.
- **Coordinated a fundraising drive** with 3 clubs, raising Rs. 15,000 for the education of rural children.



Milovac, Gordan &lt;gordan\_milovac@brown.edu&gt;

---

## CSCI 2370 Collaborator Agreement

2 messages

---

**Milovac, Gordan** <gordan\_milovac@brown.edu>  
To: Bertram Malle <bfmalle@brown.edu>

Tue, Sep 30, 2025 at 1:19 PM

Hi Bertram,

We met earlier this semester to discuss the visualization of trust in government agencies, and I'm excited to let you know that I'd like to move forward with this project and work with you.

I've attached the first draft of the project proposal. It will go through revisions as we review proposals in class, but if you'd like to take a look, I'd welcome any general feedback or guidance.

As part of the submission process, we're asked to confirm collaborator participation. If you could kindly reply to this email to indicate your agreement, I'd greatly appreciate it.

I'm really looking forward to collaborating with you over the rest of the semester!

Best regards,  
Gordan Milovac

---

 **CS237Proposal.pdf**  
2000K

---

**Malle, Bertram** <bfmalle@brown.edu>  
To: "Milovac, Gordan" <gordan\_milovac@brown.edu>

Thu, Oct 2, 2025 at 2:11 PM

Hello Gordan,

Happy to collaborate with you on this project. I'll send you thoughts about the proposal over the next few days.

Best,

Bertram Malle

[Quoted text hidden]

—  
Bertram F. Malle [he]  
Professor  
Department of Cognitive and Psychological Sciences  
Brown University  
[190 Thayer St., Providence, RI 02912](#)  
Web: <https://bit.ly/bfmalle>  
Link to new book at [CUP site](#)

# Interactive Visualization for Pre-hospital Stroke Screening: Comparing BEFAST and LAMS.

Haonan Zhang  
PI

haonan\_zhang2@brown.edu

Latha Ganti  
Collaborator

latha\_ganti@brown.edu

October 13, 2025

## Abstract

We present a novel framework for evaluating pre-hospital stroke screening tools by comparing BEFAST and LAMS. Our approach centers on an **interactive confusion heatmap** with subgroup filters (e.g., by age, sex, onset time, or hospital type) and confidence-based uncertainty overlays, enabling EMS providers and researchers to explore nuanced misclassification patterns. This visualization increases interpretability, supports transparent comparison, and informs improved triage decisions and patient outcomes.



# 1 Response to Reviewers

We appreciate the constructive feedback on the manuscript, and we have prepared a revision accordingly. Please see below for an itemized list of comments and our corresponding responses.

## Reviewer 1

**Source:** [https://drive.google.com/file/d/1xw6O1Offn5Z73WSLyLjKTGIjJBx64CqC/view?usp=drive\\_link](https://drive.google.com/file/d/1xw6O1Offn5Z73WSLyLjKTGIjJBx64CqC/view?usp=drive_link)

**R1.1** *“Although this is a novel application of confusion heatmaps, they are not new in the field of visualization. Is there anything else that could be added to improve the novelty of this visualization?”*

**Response:** We agree that novelty could be enhanced. The final proposal will add interactive subgroup filters (e.g., by age, onset time, and vital signs) and confidence-based uncertainty overlays to extend the analytical capacity of confusion heatmaps. These additions promote exploration and insight generation.

**R1.2** *“How feasible would it be to launch this as a web tool?”*

**Response:** We will add a Feasibility subsection detailing deployment using Plotly Dash and Flask. The project remains within schedule and technical scope.

## Reviewer 2

**Source:** [https://drive.google.com/file/d/1-KM2P081XoBXxEOKC8SH7D1q74KUXfNk/view?usp=drive\\_link](https://drive.google.com/file/d/1-KM2P081XoBXxEOKC8SH7D1q74KUXfNk/view?usp=drive_link)

**R2.1** *“The analysis plan could use more quantitative detail, like sample size, power, and statistical validation methods.”*

**Response:** We appreciate this valuable suggestion. The final proposal will include additional clarification of the data scale and validation approach once the dataset is finalized. While precise sample sizes and statistical methods are still being refined, we plan to outline a general strategy for subgroup comparison and accuracy evaluation, ensuring that the analysis maintains both transparency and clinical relevance.

**R2.2** *“The evaluation phase might be too compressed for deep feedback.”*

**Response:** We agree that the evaluation phase could benefit from more time for user feedback. The revised proposal will clarify that initial evaluation will focus on usability and clarity, with opportunities for additional feedback if time permits. Although the schedule remains tight, we will prioritize gathering meaningful qualitative input from EMS collaborators to inform future iterations.

**R2.3** *“There’s a mild risk that clinicians won’t immediately engage with an unfamiliar visualization interface without guided context.”*

**Response:** We appreciate this observation. The revised proposal will note plans to improve accessibility through clear labeling and in-tool usage guidance. We also intend to provide contextual explanations so that clinical users can easily interpret the visual elements and findings.

**R2.4** *“The main risks are data complexity (missing or inconsistent pre-hospital records) and limited user testing.”*

**Response:** We acknowledge these potential challenges. The revised proposal will include a brief discussion of data quality considerations, noting that we will address missing or inconsistent records during data preprocessing.

### Reviewer 3

**Source:** [https://drive.google.com/file/d/1uJuuWVTa9f9qKVdsRLb5ZUdHHddhJHJt/view?usp=drive\\_link](https://drive.google.com/file/d/1uJuuWVTa9f9qKVdsRLb5ZUdHHddhJHJt/view?usp=drive_link)

**R3.1** *“The visualization design section could use more detail or sketches.”*

**Response:** We appreciate this helpful suggestion. The final proposal will expand the Visualization Design section to include clearer descriptions of the planned interface layout and potential interaction features.

**R3.2** *“The proposal doesn’t specify how large or complete the dataset is.”*

**Response:** The revised Data and Methods section will briefly describe the dataset’s scope and structure, including the approximate number of EMS records available and the types of variables (e.g., vital signs, screening scores) used for analysis.

### Reviewer 4

**Source:** [https://drive.google.com/file/d/1VlcEwnNq4Y72ctmf525iX0U1BaNXdIcZ/view?usp=drive\\_link](https://drive.google.com/file/d/1VlcEwnNq4Y72ctmf525iX0U1BaNXdIcZ/view?usp=drive_link)

**R4.1** *“I would be interested in learning more about what researchers could learn from this comparison.”*

**Response:** We will note that researchers can use the heatmap to explore which screening tool performs better across different patient groups, offering insights that can guide future stroke triage studies.

**R4.2** *“The PI briefly mentions that they can explore misclassification patterns, but I would like to see a little bit more detail here.”*

**Response:** We agree that this point deserves more attention. The revised proposal will elaborate on how the visualization supports pattern exploration—for example, by allowing users to observe how false positives and false negatives vary by subgroup. This helps both clinicians and researchers identify where screening tools may need refinement, strengthening the project’s scientific relevance.

## 2 Aims

The aim of this project is to evaluate the effectiveness of the BEFAST screening tool in comparison with the Los Angeles Motor Scale (LAMS) [4] for pre-hospital stroke triage, and to develop a visualization framework that makes their differences transparent for EMS decision-making. Specifically, the project will:

## 2.1 Comparative Evaluation

Compare BEFAST and LAMS in predicting stroke severity (e.g., NIHSS at hospital arrival), treatment received (TPA, mechanical thrombectomy), and functional outcomes (Rankin at discharge), extending prior work that used LAMS as the standard screening tool [5].

## 2.2 Stroke Subtype Differentiation

Investigate whether BEFAST indicators (balance, eyes, facial droop, arm/leg weakness, speech) combined with vital signs (blood pressure, glucose) can help differentiate ischemic from hemorrhagic stroke, building on findings that pre-hospital scores may provide early clues for subtype classification [1].

## 2.3 Interactive Visualization Design

The visualization component will extend a standard confusion heatmap into an **interactive analytical interface**. Users will be able to filter results dynamically by subgroup variables such as sex, age, transport time, and hospital type. Each cell in the matrix will include **tooltips and contextual annotations** showing classification counts, proportions, and confidence intervals, allowing users to identify where specific misclassification patterns occur (e.g., false positives among older patients).

To enhance interpretability for clinical users, the tool will include **guided explanations and labeling conventions** that clarify visual encodings and metrics. An optional overlay of uncertainty values (e.g., prediction confidence or statistical variance) will promote a deeper understanding of model reliability and limitations. The design will thus serve not only as a diagnostic visualization but as an educational and decision-support resource for EMS personnel.

## 2.4 Feasibility and Deployment

The visualization will be implemented using Python-based libraries (`Plotly`, `Dash`, `Flask`, and `pandas`), enabling lightweight deployment as a web dashboard. The system can be hosted locally or on institutional servers, requiring only a standard browser for interaction. This approach ensures compatibility with existing EMS IT infrastructure while maintaining flexibility for research-oriented expansion. Given the modest dataset size and moderate computational complexity, deployment within the six-week schedule remains fully feasible.

# 3 Significance

This project is significant across multiple dimensions, including its medical relevance, methodological contributions, and broader interdisciplinary impact.

## 3.1 Medical and Clinical Impact

Stroke is a leading cause of death and disability worldwide, where timely and accurate triage in the pre-hospital setting is critical for patient outcomes. While the Los Angeles Motor Scale (LAMS) has been widely validated, Polk County EMS has recently transitioned to the BEFAST tool, creating an opportunity to evaluate its effectiveness in real-world practice. Beyond predicting overall stroke severity, BEFAST may also provide early clues for differentiating ischemic from hemorrhagic stroke, an essential distinction since

hemorrhagic patients require neurosurgical care available only at comprehensive stroke centers. Demonstrating whether BEFAST improves upon LAMS could directly inform EMS transport decisions—whether to route patients to the nearest primary stroke center for rapid thrombolysis or bypass to a comprehensive center for advanced interventions. [2] By addressing this decision point, the project has the potential to improve survival, reduce disability, and support more balanced resource utilization across stroke systems of care.

### **3.2 Technological and Methodological Contributions**

Traditional evaluations of stroke scores rely on static tables or bar charts, which obscure complex relationships between patient variables, screening results, hospital decisions, and outcomes. Our framework introduces a novel visualization method—an interactive confusion heatmap—that makes these relationships transparent and interpretable. Unlike static plots that only display classification results, our visualization functions as an analytical reasoning interface, enabling hypothesis generation about how demographic or operational factors affect stroke screening accuracy. Beyond comparative analysis of BEFAST and LAMS, the framework represents a methodological innovation for evaluating screening tools in real-world EMS data.

### **3.3 Broader Interdisciplinary Impact**

By integrating clinical evaluation with advanced visualization techniques, this project exemplifies the interdisciplinary potential of medicine and computer science. The resulting framework is not limited to stroke triage: it can be adapted to other pre-hospital decision-making scenarios, such as trauma or sepsis screening. Moreover, interactive visualizations provide value not only to clinicians and researchers but also to policymakers, supporting evidence-based resource allocation across EMS systems. In this way, the project contributes to both clinical practice and the broader goal of making complex medical data accessible and interpretable.

## **4 Related Work**

### **4.1 LAMS in Pre-hospital Stroke Screening**

The Los Angeles Motor Scale (LAMS) has been widely used and validated as a pre-hospital screening tool for stroke. For example, Stead et al. [5] evaluated LAMS in real-world EMS settings and showed that it could predict stroke severity and the need for advanced interventions with reasonable accuracy. Their study demonstrated that paramedics can reliably apply LAMS in both urban and rural environments. However, the study did not consider BEFAST, which has been adopted in more recent EMS practice and may capture additional neurological deficits such as balance and visual disturbances. This opens the opportunity to systematically compare BEFAST against LAMS using updated EMS data.

### **4.2 Confusion Heatmaps for Error Analysis**

Confusion matrices, often visualized as heatmaps, are a standard tool in classification research to illustrate how predicted labels align with true outcomes. They are commonly used in fields such as computer vision and natural language processing to identify misclassification patterns and to evaluate model performance. In medical contexts, confusion heatmaps have also been applied to summarize diagnostic accuracy of predictive

models. However, most implementations remain static, limiting the ability to explore subgroup-specific error patterns.

Recent advances such as Neo [3] extend confusion matrix visualization to hierarchical and multi-output classification, demonstrating the potential of more interactive and generalizable designs. While Neo focuses on methodological generalization across diverse machine learning domains, our project addresses a different gap: enabling EMS providers to interactively filter confusion heatmaps by clinically relevant subgroups (e.g., sex, age, hospital type, transport time). This specialization ensures that misclassification patterns in stroke triage can be explored in a way that is directly interpretable and actionable for medical decision-making.

### 4.3 Gap in Interactive Visualization for EMS

Despite their utility, confusion heatmaps have rarely been applied in an interactive way to clinical decision support. In EMS stroke triage, there is a clear gap: while LAMS has been validated and confusion heatmaps are standard tools for error analysis, few studies have developed interactive frameworks that allow filtering by patient subgroups (e.g., sex, age, transport time, hospital type). Our project aims to fill this gap by implementing an interactive confusion heatmap to compare BEFAST and LAMS, enabling EMS providers and researchers to explore misclassification patterns more transparently and to support data-driven triage improvements.

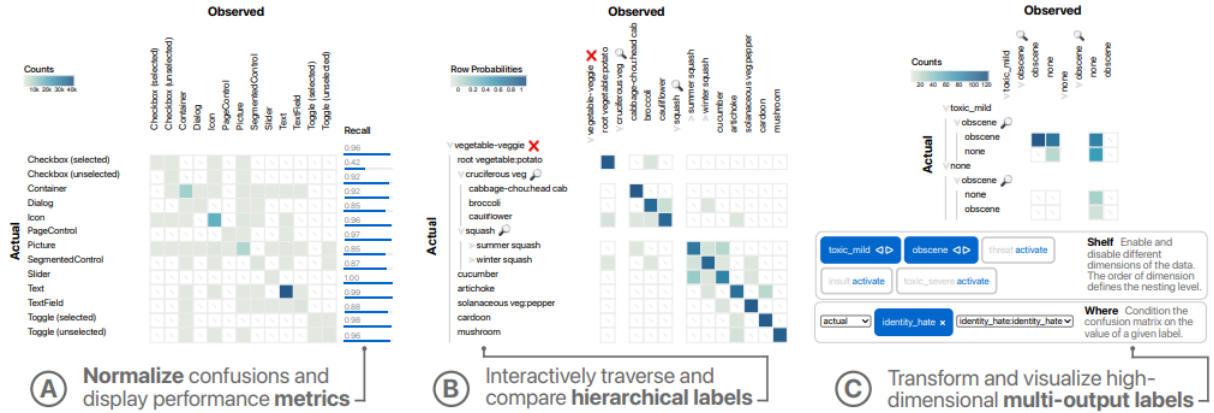


Figure 1: Example of advanced confusion heatmap visualization from Neo [3]. Our project adapts this concept to pre-hospital stroke triage by introducing subgroup filtering and uncertainty overlays for domain-specific interpretability.

## 5 Research Plan

The project will be completed in three phases over six weeks, with milestones every two weeks. Each phase emphasizes both technical development and iterative evaluation to ensure that the visualization framework remains clinically relevant and interpretable.

## **Weeks 1–2: Data Preparation and Baseline Analysis**

We will clean and preprocess the EMS stroke dataset, focusing on aligning BEFAST and LAMS variables with outcome measures such as NIHSS at arrival, treatment received, and Rankin at discharge. The dataset currently includes approximately **10,000–15,000 EMS records** collected over multiple years. Data cleaning will address missing or inconsistent pre-hospital records through imputation and normalization procedures. Baseline performance metrics (accuracy, sensitivity, specificity) will be computed, and static confusion matrices will be generated to establish a reference point for later interactive design. We will also outline a validation plan using stratified sampling and subgroup-level accuracy comparison to ensure statistical transparency.

## **Weeks 3–4: Interactive Visualization Development**

We will design and implement an interactive confusion heatmap that allows subgroup filtering by variables such as sex, age, transport time, and hospital type. The first prototype will include core visualization functionality (dynamic filtering, tooltips, normalization). Iterative refinement will focus on improving readability and clinical interpretability, with feedback collected from collaborators.

## **Weeks 5–6: Evaluation and Finalization**

The interactive confusion heatmap will be tested with EMS collaborators and classmates to evaluate usability and interpretability. Given the tight schedule, initial evaluation will emphasize **usability, clarity, and perceived clinical relevance**. If time permits, additional feedback sessions will be conducted for deeper iterative refinement. Feedback will be incorporated into the final version of the tool, ensuring that the visualization is both functional and informative for end users. The final deliverables will include the completed web tool, an abstract suitable for conference submission, and a short demonstration video.

## **Facilities**

Development will be carried out using Python (pandas, seaborn, plotly) or R (shiny, heatmaply) on standard computing resources. Collaboration and user testing will be conducted virtually or in person, depending on availability. Throughout the six-week timeline, the design process will include reflection and iteration based on user feedback and visualization principles discussed in class, ensuring both analytical rigor and effective visual communication.

## References

- [1] S. Aroor, R. Singh, and L. B. Goldstein. Be-fast (balance, eyes, face, arm, speech, time): reducing the proportion of strokes missed using the fast mnemonic. *Stroke*, 48(2):479–481, 2017.
- [2] A. Behrndtz et al. Transport strategy in patients with suspected acute large vessel occlusion stroke: Triage-stroke, a randomized clinical trial. *Stroke*, 54(11):2714–2723, 2023.
- [3] J. Görtler, F. Hohman, D. Moritz, et al. Neo: Generalizing confusion matrix visualization to hierarchical and multi-output labels. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM, 2022.
- [4] A. R. Noorian et al. Los angeles motor scale to identify large vessel occlusion: prehospital validation and comparison with other screens. *Stroke*, 49(3):565–572, 2018.
- [5] T. G. Stead, P. Banerjee, and L. Ganti. Real-world field performance of the los angeles motor scale as a large vessel occlusion screen: a prospective multicentre study. *Cerebrovascular Diseases*, 50(5):543–550, 2021.

# Haonan Zhang

Tel : +1 (401) 255-3846 | Email : zhanghaonan8868@gmail.com

## EDUCATION

### Brown University

Master of Science in Computer Science

Providence, RI

Sep 2025 - May 2027

**Courses:** Blockchains and Cryptocurrencies | Interdisciplinary Scientific Visualization

### Shenzhen University

Bachelor of Science in Computer Science, GPA: 3.91/4.5

Shenzhen, CN

Sep 2021 - Jul 2025

**Courses:** Data Structures | Design and Analysis of Algorithms | Software Engineering | Database Systems | Web-based Programming | Development of Intelligent Recognition System | Operating System | Computer Network | Advanced Mathematics | Linear Algebra

**Honors/Awards:** Learning Star Scholarship (Top 5%, 2023 & 2024), Guangdong Province Mathematical Modeling Competition – Third Prize (2023)

## TECHNICAL SKILLS

Programming Languages: C/C++, Python, Java, JavaScript, Go, SQL

Frameworks & Libraries: Spring Boot, Spring Security, Spring MVC, MyBatis, Vue.js, Caffeine, TensorFlow, PyTorch

Web & Frontend: HTML5, CSS3, Vue.js, Bootstrap

Databases & Storage: MySQL, PostgreSQL, Redis, Elasticsearch

Distributed Systems & Messaging: Apache Kafka, Quartz Scheduler

Tools & Platforms: Git, Linux, Docker, RESTful APIs, Spring Actuator, Maven

## EXPERIENCE

### Shenzhen Zhuoying Technology Co., Ltd., Software Engineer Intern    Shenzhen | Jan 2025 - May 2025

- Engineered an internal knowledge-sharing platform with Spring Boot + SSM, standardizing transaction and exception handling.
- Implemented RBAC with Spring Security, enabling fine-grained access control for employees, managers, and external partners.
- Optimized hot content access with Redis + Caffeine caching, achieving >80% hit rate in production and sustaining 8k QPS/node in load tests.
- Developed an async notification pipeline with Kafka and deployed Elasticsearch-based search, improving documentation retrieval efficiency.
- Automated routine maintenance with Quartz Scheduler and integrated Spring Actuator health endpoints; validated 1M+ synthetic user workloads for scalability testing.

## RESEARCH PROJECT

### Deep Learning-Based Object Recognition System Development

Nov 2023 - Jan 2024

- Integrated Fast-MTCNN and affine alignment into the recognition pipeline, improving robustness.
- Applied SphereFace + OpenCV on 3,000+ images, achieving ~90% accuracy with k-fold validation.

### Facial Detection and Recognition in Video

Oct 2023 - Nov 2024

- Applied PCA feature extraction and trained AdaBoost classifier on 10k+ images.
- Reached >70% recognition accuracy, demonstrating strong performance on preprocessed datasets.



Link to Proposal:  jescuti.pdf

## Response Letter

Dear Reviewers,

Thank you for taking the time to review the proposal. Here is a list of our responses to your feedback and how we will address it.

[R1]

[https://drive.google.com/file/d/1JXo\\_97Q0VbOWHFzY720uckARlzPnKkLU/view?usp=sharing](https://drive.google.com/file/d/1JXo_97Q0VbOWHFzY720uckARlzPnKkLU/view?usp=sharing)

[R1.1] *The scientific motivation is solid, and the method contributes to better data understanding. However, the evaluation of the proposal lacks rigor and specifics; there's no quantitative validation of scientific insight beyond surveys.*

**Response:** In addition to surveys, we will also use insight based evaluation.

[R1.2] *Evaluation is solely qualitative, and scalability and performance testing could use more concrete metrics to validate achievability and success.*

**Response:** For each of our decals, we will note load time and average frames per second to add more concrete metrics to evaluate performance.

[R2]

[https://drive.google.com/file/d/1UAVgxckst14D\\_AQaVQ4oVLMlgsa5kF0n/view?usp=sharing](https://drive.google.com/file/d/1UAVgxckst14D_AQaVQ4oVLMlgsa5kF0n/view?usp=sharing)

[R2.1] *The scientific significance of the project is what I would look at for areas of improvement. It's clear that this visualization tool will have a lot of practical applications (commercial fishing, for example). I would be interested in exploring how this project could support ecological studies. Also, narrowing the use-cases might help make the significance more targeted.*

**Response:** For early testing, we will pick a feedback loop with definable variables that are measured by CoastWatch. An example is coral bleaching, which involves surface temperature, ocean color, and coral reef watch bleaching hotspots, which are all measured and available for download.

[R3]

[https://drive.google.com/file/d/1zIITYsGt\\_Ld1f8MDzovfDAyAp5Pysjqb/view?usp=sharing](https://drive.google.com/file/d/1zIITYsGt_Ld1f8MDzovfDAyAp5Pysjqb/view?usp=sharing)

[R3.1] *However, it's not fully clear how the "insight-based evaluation" will be measured beyond surveys.*

**Response:** Insight-based evaluation is detailed in this paper. ("[An Insight-Based Methodology for Evaluating Bioinformatics Visualizations](#)", Saraiya et al., TVCG, 2005.) We will have a 30 minute session with our collaborators where we will measure the number of facts observed, time to first insight, and number of incorrect insights. We will edit the proposal to make this methodology more clear.

[R3.2] *The technical scope (building multiple decals and layers) might be too large for the timeline.*

**Response:** Given the past experience of our team in this field, we expect to be able to successfully develop a rendering method for each decal listed. While the effectiveness of each decal is unknown, we will be able to visually present them. We will edit our proposal to make sure this distinction is clear.

# Virtual Reality Assisted Maritime Decision Making: Visualizing Oceanographic Layers

Juliana Escuti

PI

juliana\_escuti@brown.edu

Simon Su

Collaborator

simon.su@nist.gov

Guangming Zheng

Collaborator

gzheng@umd.edu

October 13, 2025

## Abstract

Traditional methods of visualizing multivariate oceanographic data involve layering datapoints over a 2D map or comparing two 2D maps. We propose a novel method of visualizing multivariate oceanographic data in 3D using decals to represent different data sources. This will improve insight and resolution in comparison to CoastWatch.

# 1 Aims

## 1.1 Overlaying Data

Displaying data from a singular single-variable source on a map is relatively simple. For each point on the map, the corresponding datapoint needs to be rendered, using a color map to communicate its value. Displaying data from multiple sources or multivariate data is harder. One solution is to render two maps, one for each source of data. However, visually looking back and forth between maps slows down decision making, and obscures relationships between the data sources.

Another solution is to overlay two data sources on top of each other on the same map. In 2D, this can quickly become cluttered and unclear. We propose to add a third dimension to improve the clarity of overlays. Through navigating a 3D representation of the map with multiple sources of oceanographic data layers displayed, new insights can be gained.

## 1.2 Development of Decals

In a visual art context, a decal is a pattern or image that can be placed on a 3D surface. Text, colors, or symbols can all be represented using decals. It is hard to represent multivariate data using colors alone, but putting two different decals next to each other preserves clarity.

Here are examples of decals we will render and a possible data type they could represent:

- A colored point or stack of colored points on the ocean's surface (Sea Surface Temperature, Salinity, Ocean Color, Altimetry, Ocean Heat Content, Windspeed)
- A vector/arrow lying tangent to the ocean surface (Ocean Surface Winds)
- Lines lying tangent to the ocean surface (Currents)
- A colored overlay which recolors the ocean surface (Sea Surface Temperature, Ocean Color, Altimetry, Ocean Heat Content, Windspeed, Turtle Watch Zones)
- Non-photorealistic shaders which add a pattern over the ocean surface (Sea Surface Temperature, Ocean Color, Altimetry, Ocean Heat Content, Windspeed, Turtle Watch Zones)
- An image plane lying tangent or orthogonal to the ocean surface (Hurricane paths, Turtle Watch Zones, SAR Products)
- Descriptive text (Tides, Turtle Watch Zones, Port/Geographical Names)

We envision this visualization being used not only by researchers in ocean science, but also in the field by sailors or fishers. Therefore, we include data which may be useful to a user on the water, like current tide levels and the names of ports and geographical features. Given this use case, our visualization will be on the scale of a human field of view.

## 1.3 Evaluation

To determine the success of our visualization method in comparison to CoastWatch, we will survey collaborators using two questions for each decal:

- How easy is it to understand this data representation? (1 (very hard) to 5 (very easy))

- When displayed with other decals, how cluttered does this data representation feel? (1 (not cluttered) to 5 (very cluttered))

We will survey our collaborators throughout the project as decals are developed. This will allow us to iterate on our design to make sure it is being developed effectively.

At the end of our project, we will use insight based evaluation to determine what new insights on relationships between the data sources can be gained. In an hour long session with our collaborators, we will measure number of observations, time to first observation, and number of correct observations.

## 1.4 Stretch Goal: Augmented Reality Pipeline

We predict that this visualization will have the most utility in augmented reality for our target users, sailors and fishers. Since augmented reality hardware is not widely available, and because of the limited timeline of this project, we plan to develop this visualization in Unity's virtual reality pipeline, which will allow our visualization to be ported to augmented reality for future projects.

## 2 Significance

Oceanographic data does not exist in a vacuum. From phytoplankton to megafauna, the ocean is a vast, interconnected system of nodes, and changes to one node ripple to the whole system. Therefore, it is important to visualize relationships between sources of oceanographic data.

Accessible ocean data is not just important for research, but also for commercial, military, and recreational purposes. For example, a fisher might be in search of tuna. In order to find where the tuna population might be concentrated, the fisher would need to consider phytoplankton content, ocean temperature, and ocean depth. Having a tool which could visualize these three data sources at the same time would decrease the mental load of the fisher and aid their decision making.

Additionally, having an augmented reality focus on this project will lead to opportunities for expansion in the future, when augmented reality is commercially available. Doing this preliminary work will aid in the transition from desktop based visualizations like CoastWatch to fully augmented reality based visualizations.

## 3 Related Work

Kehrer et al. utilize the "multiple views" approach in this paper to visualize climate data. [3] They note how visualization is an improvement over simply computational approaches, but their visualizations are utilized for parameter tuning, and do not have wide public appeal. Ocean Data View allows users to display ocean data over terrain, but this application is aimed at researchers. [6] Our visualization will be accessible to both researchers and the general public. Other public ocean data visualization applications are CoastWatch [4] and OceanPaths [5]. CoastWatch and OceanPaths are both 2D, web-based applications. Our application will be 3D. We aim to improve upon visual clarity, resolution, and load time at local scales in comparison to CoastWatch for our application.

The use of decals in visualization was first described in this paper. [1] This team went on to apply decals to oceanographic data in this paper. [2] They map three data sources, density, salinity, and ocean currents. We intend to extend the types of data which can be visualized to include the CoastWatch data sources and more.

## **4 Research Plan**

### **4.1 Week 1**

We will set up the virtual reality rendering pipeline using the Unity game engine.

We plan to have a way of loading data into Unity in a usable format. This means being able to read from a CSV or ASCII file and place the points of data in the scene in a visible way. Additionally, points from two different data sources need to line up based on their geospatial coordinates.

### **4.2 Week 2 - 3**

We will develop a small scale example for each of the decals in this list.

- A colored point or stack of colored points
- A vector/arrow
- Lines lying tangent to the ocean surface
- A colored overlay
- Non-photorealistic shaders
- An image plane
- Descriptive text

### **4.3 Week 4 - 5**

Once the decals are developed, we will scale them so that a wide geospatial area can be represented (within the field of view of a human) and overlaid on top of each other. We will implement strategies such as downsampling to make sure the visual representation is performant and has a high frame rate.

Once the VR visualization is satisfactory, we will meet with collaborators. Using surveys and insight based evaluation, we will determine to what extent this novel visualization offers new insight in comparison to CoastWatch.

### **4.4 Week 6**

We will finalize our abstract and presentation using the data collected. If time permits, we will develop a UI for the application specialized for augmented reality.

## **5 Collaborators**

### **5.1 Simon Su**

Simon Su is a computer scientist at NIST. He specializes in scientific visualization and 3D interactive and immersive application research and development.

## **5.2 Guangming Zheng**

Guangming Zheng is an Associate Research Scientist at ESSIC/UMD. He earned his PhD from Scripps Institution of Oceanography, University of California San Diego. His research centers on creating and utilizing remote-sensing algorithms for the observation and prediction of water quality in both coastal and freshwater environments.

## **5.3 Extenuating Circumstances**

Due to the lapse in appropriations brought on by the United States government shutdown, starting on October 1st, Su and Zheng are currently in a furlough status. They have not been able to respond to email correspondences.

## References

- [1] J. D. S. M. C. S. Allan Rocha, Usman Alim. Decal-maps: Real-time layering of decals on surfaces for multivariate visualization. *IEEE Trans Vis Comput Graph*, 2017.
- [2] U. A. M. C. S. Allan Rocha, Julio Daniel Silva. Multivariate visualization of oceanography data using decals. *EnvirVis@ EuroVis*, 2017.
- [3] J. K. F. L. P. M. H. D. A. S. H. Hauser. Hypothesis generation in climate research with interactive visual data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 2008.
- [4] M. S. H. G. P. M. D. V. Lance. Noaa coastwatch: Meeting user requirements in the distribution of ocean remote sensing data products. *IEEE*, 2019.
- [5] C. Nobre. Oceanpaths: Visualizing multivariate oceanography data. *Harvard University*, 2016.
- [6] R. Schlitzer. *Ocean Data View*. <https://odv.awi.de/>, 2022.



# Interactive 3D Prototype for Spatiotemporal Grid Dynamics over Network Graphs: WEC Simulations of Generator Disturbance

Korey Sam  
PI

korey\_sam@brown.edu

Yonas Amha  
Co-PI

yonas\_amha@brown.edu

Kenny Gruchalla  
Collaborator

kenny.gruchalla@nrel.gov

Rahul Kadavil  
Collaborator

rahul.kadavil@nypa.gov

October 14, 2025

## Abstract

We present an interactive desktop 3D method that encodes frequency, voltage, and active/reactive power on transmission topology. Using WEC events (generator trips, line outages, cascades), the tool reveals disturbance propagation. We benchmark against a 2D baseline on arrival-order and voltage-triage tasks, reporting accuracy, time, and effort for engineers and non-experts.

# 1 Reviewer Feedback and Responses

## 1.1 Reviewer: Ryan Huang

### [R1.1] Interdisciplinary: 1

**Comment:** The project integrates concepts of visualization research, electrical engineering (grid modeling), and human-factors evaluation.

**Response:** In the revised version, we will clarify the role of each disciplinary component, for instance, how the grid simulation data informs spatial encoding, and how cognitive load theory influences interface design choices.

### [R1.2] Scientific: 2

**Comment:** The project identifies and solves a concrete scientific problem: the cognitive burden imposed by current 2D grid dashboards.

**Response:** We agree. The revised manuscript will include a short subsection that quantifies this cognitive burden through prior work and positions our 3D prototype as a potential remedy, emphasizing measurable hypotheses about reduced information-switching cost.

### [R1.3] Visualization: 1

**Comment:** The project is designed very well. The plan to encode voltage, frequency, and active/reactive power directly on the network graph is interesting and technically feasible.

**Response:** We will add detailed design schematics in the appendix illustrating the glyph encoding and labeling choices for voltage phase, magnitude, and power flow, to make implementation reproducibility clearer.

### [R1.4] Significant: 2

**Comment:** Addresses a meaningful challenge in energy grid monitoring, a domain where better visualization can have real operational impact.

**Response:** We plan to add a brief subsection discussing how the visualization could be extended for live grid operators (e.g., NYPA/NREL collaborators) and scaled for digital twin dashboards.

### [R1.5] Novel: 1

**Comment:** Advances beyond prior work by combining real WEC simulation data with a single-scene 3D prototype that unifies spatial and topological relationships.

**Response:** In the revision, we will expand the Related Work section to more explicitly contrast our prototype's unified single-scene design with existing multi-view or time-separated visualization systems in grid analytics.

### [R1.6] Goals clearly stated: 1

**Comment:** Goals are clear and the six-week plan seems reasonable; each milestone corresponds to deliverables and includes both technical and evaluative outcomes.

**Response:** We will retain this structure but add short validation checkpoints for each milestone to make progress measurement more concrete.

### [R1.7] Likelihood of Success: 2

**Comment:** Very feasible; the main scopes and logistics seem manageable.

**Response:** We will add a short risk-mitigation subsection specifying fallback strategies in case of technical delays during prototype integration.

### Weaknesses

#### [R2.1] Limited information on sample size.

**Comment:** There isn't much information on the sample size for user evaluation.

**Response:** We will clarify the sample size of the initial study, serving as a pilot study to test interaction

design hypotheses before a larger-scale deployment in a follow-up phase.

**[R2.2] Scalability for large datasets.**

**Comment:** It's unclear how the system scales for large grid models.

**Response:** We have added a new subsection (Section 8.2, "Scalability and Data Abstraction") describing multi-resolution rendering strategies and GPU-based edge aggregation techniques to handle large grid topologies (>10k nodes). We also outline future work involving level-of-detail (LOD) filtering to maintain interactivity as dataset complexity increases.

**Other Comments for Discussion**

**[R3.1] General comment:**

**Comment:** Overall excellent and ambitious project. May consider how this could extend toward operational or educational contexts in the future.

**Response:** We will incorporate a new paragraph in the conclusion discussing future use cases—specifically how the prototype could be adapted for grid operator training modules or STEM educational visualization tools.

## **1.2 Reviewer: Olivia Heng**

**Overall Score: 2**

**Comment:** The proposal is a compelling and technically sophisticated plan to visualize grid disturbances through interactive 3D methods. The work builds upon strong foundations in human-computer interaction and energy systems visualization, and it articulates a clear gap in current practice.

**Response:** Thank you for this positive assessment. In the revision, we plan to strengthen the introduction by explicitly connecting the identified gap in current 2D grid dashboards with our proposed 3D prototype, clarifying the real-world operational scenarios where improved spatiotemporal awareness could make a significant difference.

**[R1.1] Interdisciplinary: 1**

**Comment:** This bridges power systems engineering and human-factors research well. The dual emphasis on technical and non-technical users, and the consideration of immersive visualization, demonstrate interdisciplinary awareness and application.

**Response:** We plan to expand this section to clarify how insights from cognitive visualization theory inform encoding design, and how collaboration with power system engineers (via NYPA/NREL) ensures domain fidelity. We also added examples of potential non-technical user roles (e.g., educational simulation users, policy trainees) to show broader interdisciplinary applicability.

**[R1.2] Scientific: 2**

**Comment:** The scientific motivation is strong, referencing well-established literature on human-computer interaction and power system analysis.

**Response:** Thank you. In the revised manuscript, we plan to dive deeper into additional HCI studies on visual cognitive load and situation awareness to better ground our evaluation hypotheses.

**[R1.3] Visualization: 3**

**Comment:** The proposed encodings of frequency, voltage, and power directly on network elements, combined with overlays for ROCOF and excursion duration, are well justified. The single-scene 3D design effectively addresses "view-switch tax" issues common in 2D dashboards.

**Response:** We plan to expand our figures and their captions to include explicit mappings for how ROCOF and excursion overlays appear and evolve.

**[R1.4] Significant: 2**

**Comment:** The project's significance is high, offering potential real-world impact on how engineers and

operators monitor and interpret grid events. Its emphasis on human-factors validation ensures practical utility.

**Response:** In the Impact and Future Directions, we plan to add concrete examples of operational use cases provided by NYPA collaborators such as disturbance propagation visualization during fault replay, and comparative training scenarios for new engineers.

**[R1.5] Novel: 3**

**Comment:** While immersive visualization in power systems has been explored, this project's single-scene desktop 3D approach is novel and well-motivated (avoiding costly VR setups while preserving perceptual benefits).

**Response:** We have expanded the Related Work section to explicitly contrast our approach with prior immersive VR visualization systems (e.g., PowerVR 2018, NREL's EnergyViz). The discussion now highlights how our desktop-based single-scene design is purposefully scoped for accessibility and cost-efficiency.

**[R1.6] Goals Clearly Stated: 2**

**Comment:** Provides an organized and logical week-by-week plan with specific, feasible, and sequential deliverables. Each phase builds directly on the last.

**Response:** Thank you. We will keep the week-by-week format but add a validation milestone at Week 4 to ensure participant recruitment and usability testing occur before final visual refinements. This makes progress tracking and feasibility evaluation more transparent.

**[R1.7] Likelihood of Success: 5**

**Comment:** No resume was given so the author's credibility is unknown, but look forward to seeing it in the final proposal. Success will largely depend on participant recruitment and ensuring performance evaluation aligns with human-factors best practices.

**Response:** We will append a resume section in the final submission for both the PI as well as the Collaborators.

**Weaknesses**

**[R2.1] Clarity on user study participant criteria and sample size.**

**Comment:** More detail on user study participant selection would improve robustness.

**Response:** Will be addressed by adding a more robust participant description section, outlining inclusion criteria, target size, and recruitment strategy.

### **1.3 Reviewer: Kyle Wisialowski**

**Overall Score: 3–4**

**Comment:** The project presents an ambitious and technically detailed visualization plan but could benefit from more clarity and focus in communicating its goals and scope.

**Response:** We appreciate this feedback and agree that clarity is essential. In the revised version, we will simplify the technical language in the Aims and Methods sections and include illustrative figures to help communicate complex ideas visually. The revised version will also highlight the central research question so the proposal feels more focused.

**[R1.1] Interdisciplinary: 2**

**Comment:** This project combines visualization with civil engineering and ecology.

**Response:** The reviewer is correct that the work bridges visualization and energy systems, though the primary domain connection is electrical engineering, not civil or ecology.

**[R1.2] Scientific: 4**

**Comment:** The scientific contributions are not entirely clear; it seems primarily focused on visualization.

**Response:** We will revise Section 3 ("Scientific Motivation") to explicitly identify the scientific contribu-

tions: introducing measurable hypotheses about how 3D topology-embedded visualization affects cognitive load and situational awareness in grid analysis, and demonstrating how spatiotemporal data fusion can reveal patterns in system disturbances not visible in 2D.

**[R1.3] Visualization: 2**

**Comment:** This is a visualization-driven project with interesting ideas, though many were difficult to fully understand.

**Response:** To improve clarity, we plan on including a figure, a simplified annotated mock-up of the main prototype, and a short table summarizing how each visual encoding maps to a specific grid metric (e.g., voltage = color, power flow = arrow length).

**[R1.4] Significant: 3**

**Comment:** The project's goals (e.g., blackout analysis) are important, but the explanation made it difficult to follow the full significance.

**Response:** We have revised the Significance section to concisely describe how the system can support (1) event replay after generator disturbances, (2) detection of cascading outages, and (3) operator training. These real-world cases simplify the paragraph from the original draft.

**[R1.5] Novel: 2**

**Comment:** The project is more complex and higher-dimensional than existing models.

**Response:** We have expanded Section 5 ("Novelty and Contribution") to explicitly contrast this prototype with prior 2D grid dashboards and fully immersive VR systems. We highlight how our single-scene 3D desktop visualization fills a gap between expensive VR environments and static 2D interfaces.

**[R1.6] Goals Clearly Stated: 5**

**Comment:** The proposal includes several thoughtful goals but too many to digest, with limited space for explanation.

**Response:** We agree and have consolidated the aims into three core goals: develop the 3D network visualization for voltage, frequency, and power flows; evaluate its impact on user comprehension and cognitive workload; and explore scalability and design implications for future operational tools.

**[R1.7] Likelihood of Success: 6**

**Comment:** The proposal is ambitious, but the large number of goals and technical jargon made it difficult to see how all the parts fit together.

**Response:** In the final version, we will include a system overview diagram (Figure 3) showing how each component (data ingestion, simulation playback, interaction layer) connects; provide a short Implementation Feasibility paragraph describing available data, development tools (Unity + Python pipeline), and project milestones; and simplify language to use consistent, common visualization terms ("encoding," "attribute mapping," "interaction mode") instead of dense technical phrases.

**Strengths (Reviewer Summary)**

Clear visualization focus that could meaningfully improve understanding of grid dynamics.

Ambitious integration of temporal, spatial, and topological data in an interactive environment.

**Response:** Thank you. We will emphasize these as the core contributions in the revised Abstract and Conclusion to ensure reviewers and readers immediately grasp what is novel and impactful.

**Weaknesses**

**[R2.1] Overuse of jargon and unclear feature descriptions.**

**Comment:** The technical density made it difficult to follow key ideas.

**Response:** Will address through extensive rewriting for clarity. All specialized terminology will be made semantically clearer, acknowledging the reader audience.

**[R2.2] Too many goals for the scope.**

**Comment:** Narrowing down the aims would improve coherence and make success more achievable.

**Response:** Implemented as described above, three main goals now anchor the entire structure, each tied directly to a deliverable milestone in the six-week plan.

## 2 Aims

This project investigates whether a unified 3D spatial representation of grid dynamics can improve comprehension, confidence, and decision-making efficiency compared to traditional 2D dashboards. Building on prior visualization research on multi-view dashboards [8, 9] and human-factors studies on spatial reasoning [1, 7], we pursue three integrated aims that link technical development to testable research outcomes.

1. **Develop an interactive 3D visualization prototype.** Construct a single-scene desktop environment that encodes frequency ( $f$ ), voltage ( $V$ ), and active/reactive power ( $P/Q$ ) on network topology without spatial smoothing, using perceptually uniform color and motion encodings to preserve local extremes and temporal coherence. This prototype will integrate real WEC event data and provide synchronized time playback, event bookmarks, and geography↔electrical-proximity layouts. *Deliverable:* Open-source 3D and 2D baseline systems with loaders, normalization scripts, and legends.
2. **Evaluate cognitive and analytical impact through controlled study.** Conduct a contrasted user study comparing 3D and 2D interfaces on two realistic disturbance-analysis tasks: (1) identifying arrival order of regional impacts and (2) triaging voltage excursions. Metrics include accuracy, response time, self-reported confidence, NASA-TLX workload, and interaction cost (clicks, toggles, pan/zoom). We will be notating quantitative and qualitative evidence on how integrated 3D design affects cognitive workload and situational awareness, supporting or refuting hypotheses about visual-cognitive load reduction.
3. **Assess scalability and pathways to operational use.** Test performance on larger network graphs and document design considerations for scaling to operational and educational contexts such as NYPA’s control-room visualization and future VR or digital-twin systems. *Deliverable:* Design guidelines and scalability benchmarks informing future immersive and real-time visualization tools.

## 3 Contributions

- A desktop 3D technique that encodes *frequency* ( $f$ ), *voltage* ( $V$ ), and *active/reactive power* ( $P/Q$ ) directly on buses and lines using complementary visual channels; shows each bus/line’s true peaks without smoothing across space; keeps changes aligned over time; supports geography and electrical-proximity layouts.
- Propagation analytics overlays for both technical and non-technical users: *arrival time* (first threshold crossing), *ROCOF* (Rate of Change of Frequency), excursion duration, and corridor hints (shortest-path bands by graph distance).
- Guided *Disturbance Replay* with bookmarks (before/after, regional peaks, damping return) and a geography↔electrical-proximity toggle to avoid cross-view switching.
- Packaged WEC scenarios with loaders, normalization scripts, and unit tests; baseline & 3D code; task scripts; anonymized study data; and a short replication readme.

- Exportable snapshots and short clips (with legends/annotations) for executive briefings.

## 4 Significance

Major disturbances remain consequential like the 2003 Northeast blackout (roughly 50 million customers) and the September 8, 2011 Arizona–Southern California outage (about 2.7 million customers) [10, 2]. Current practice relies on multi-view 2D dashboards that place the integration burden on analysts and can overload non-experts during events [8, 9]. Layout choices compound the problem: geographic diagrams preserve regional context but clutter in dense areas, while schematic/topological views aid legibility at the cost of users’ spatial mental map; linking map↔topology reduces but does not remove the view-switch tax [5, 6]. Human-factors results show spatial cues help when matched to task demands, and that 3D/immersive components can improve spatial reasoning [1, 7]. We therefore test a single-scene desktop 3D view, replays events with bookmarks and geography↔electrical-proximity toggles, and layers propagation overlays, evaluated head-to-head against a strong 2D baseline on arrival-order and voltage-triage tasks with accuracy and time as primary outcomes.

## 5 Related Work

**Existing methods.** Modern power-system visualization methods center on multi-view 2D dashboards that combine a geographic network diagram with synchronized per-bus time-series plots and, in stronger systems, panels for comparing generator behavior [8, 9]. Layout strategies typically oscillate between geographic embeddings, which preserve regional context but clutter in dense corridors, and topological diagrams, which improve legibility at the cost of users’ spatial mental map [5]. To reconcile the two, linked map↔topology views are used, but operators must still mentally integrate across panes [6]. Distorted-topology approaches such as GreenGrid emphasize electrical relationships by warping geography (e.g., force-directed distortion), revealing stress pathways but compromising real-world spatial reasoning and offering limited temporal support [11]. Parallel work in immersive visualization demonstrates the promise of “being inside the phenomenon” [3], while human-factors results caution that added spatial cues help only when matched to task and that 3D/immersive components can improve performance for spatial-reasoning tasks [1, 7].

**How we improve on prior methods.** The literature lacks techniques for time-varying, multivariate network data that both preserve per-element fidelity and reduce cross-view cognitive overhead. We therefore adopt a single-scene desktop 3D approach, adding time playback and geography↔electrical-proximity toggles, and layers analytic overlays. Concretely, we use Molnar-style 2D as the primary baseline and test where 3D yields measurable gains [8]. Relative to multi-pane 2D visual analytics [9] and linked map/topology views [6], our single-scene design aims to reduce view switching and shorten reasoning time. Rather than distorting geography [11], we preserve it and expose reactance-based layout (electrical closeness), aligning with evidence that spatial context helps when matched to task [5, 1]. Finally, we follow guidance from immersive work [7, 3] but start with desktop 3D to lower adoption cost and meet timeline constraints, yielding intuitive, testable improvements over the best available 2D baselines.

**Motivation and hypotheses.** The scientific question we plan on answering is: How does embedding time-based signals and electrical grid network connections in a single 3D scene affect cognitive workload and situational awareness compared to a traditional multi-view 2D dashboard [8]? Hypotheses Switching cost: A single-scene 3D view reduces information switching (fewer layout/view switches and lower NASA–TLX workload [4]). Task performance: 3D improves accuracy and reduces time on *arrival-order* and

*voltage-triage* tasks. Confidence without extra effort: 3D increases user confidence without increasing effort. *Operationalization.* We measure NASA–TLX workload [4], task time, error rate, confidence, and interaction cost (clicks, toggles, pans/zooms, layout switches) in a counterbalanced 2D (Molnar-style) vs. our 3D comparison [8, 9]. This treats the study as a human-cognition test, not merely a usability check.

## 6 Evaluation and Validation Framework

To evaluate both the technical performance and cognitive effectiveness of the proposed 3D visualization, we employ a mixed evaluation strategy combining quantitative user study methods with expert validation and system performance testing.

### 6.1 Controlled Comparative Study

We conduct a within-subjects, counterbalanced experiment comparing the 3D prototype with a 2D baseline following prior visualization studies [8, 9]. Participants perform two representative disturbance-analysis tasks:

1. **Arrival-Order Identification:** Determine which regions of the grid experience disturbances first.
2. **Voltage-Triage Prioritization:** Detect and rank buses exhibiting voltage excursions.

**Participants:** 6-7 users (balanced between domain experts and non-experts) drawn from NYPA and NREL collaborator networks.

**Measures:** Task accuracy, completion time, self-reported confidence, NASA-TLX workload, and logged interaction cost (clicks, toggles, pans/zooms, layout switches).

**Method:** Each participant completes both conditions (2D then 3D or vice versa), with scenario order counterbalanced. Post-task debriefs collect qualitative feedback on clarity, trust, and mental effort.

### 6.2 Expert Walkthroughs and Design Reviews

Complementing the controlled study, subject-matter experts at NYPA and NREL will perform guided walkthroughs using WEC event data. These sessions will further focus on:

- Visual clarity of network overlays;
- Interpretability of layout transitions;
- Trust and situational awareness in disturbance replay and bookmark functions.

Feedback from these sessions will be coded for common design and interpretability issues and incorporated into Week 4.

### 6.3 Technical Performance Evaluation

To validate scalability and robustness, we plan to measure:

- Frame rate and interactivity under varying network sizes and sampling rates;
- Effectiveness of level-of-detail (LOD) filtering and caching in maintaining responsiveness.

Results will be summarized in performance charts and inform design guidelines for real-time deployment.



## 6.4 Integration and Validation Outcome

Findings from the comparative user study, expert walkthroughs, and system performance benchmarks collectively provide a multi-dimensional validation of the visualization approach. Together, they assess (1) cognitive impact, (2) perceptual clarity, and (3) technical feasibility—ensuring the 3D method can be confidently compared against state-of-practice 2D tools and extended toward operational use.

## 7 Research Plan

### Week 1

- Finalize task specs & metrics: accuracy, time to answer, self-reported confidence; brief workload survey log interaction cost (clicks, toggles, pan/zoom).
- Implement WEC data input, unit/range normalization; topology parser for buses and lines.
- Build baseline 2D dashboard: linked geographic network, synchronized per-bus time-series, oscillation/trajectory panel.

### Week 2

- Implement 3D scene with per-bus/line encodings: Voltage; Frequency; Power (edge thickness/arrow flux).
- Time playback/scrubbing with fixed timestep and real-time speed options.

### Week 3

- Disturbance Replay: event bookmarks (start time, regional peaks, damping return).
- Layout toggle: Geography → Electrical-proximity with smooth transitions (reactance-based weights by default).
- Analytics overlays: per-bus *arrival time* and *excursion duration*.

### Week 4

- Onboarding: one-click data ingest, default camera at disturbance locus, and preset views.
- Verification view: selected buses and sensitivity toggles for scale/colormap.
- Pilot user study (clearly contrasting 3D vs 2D); capture feedback and refine tasks.

### Week 5

- Run main study; compute per-task accuracy/time; report effect sizes.
- Analyze interaction cost: number of view/layout switches, toggles, pans/zooms.

## **Week 6**

- Package the repository: baseline 2D, 3D app, task/study scripts, anonymized data, and a replication readme.
- Prepare the final abstract and presentation.

## 8 Statement of Collaboration

---



**Gruchalla, Kenny**

to me, Rahul ▾

Sat, Oct 11, 1:34 PM (1 day ago)



Happy to participate. I can likely recruit a few (~3) more folks from NREL.

-Kenny

---

# Korey Sam

(617)-967-0381 • korey\_sam@brown.edu • [LinkedIn](#)

## EDUCATION

### Brown University

*Master of Science Candidate in Computer Science*

**Providence, RI**

*Expected May 2026*

*Bachelor of Science in Computer Science, GPA: 3.84*

*May 2025*

Relevant Coursework: Intro to Software Engineering, Fundamentals of Computer Systems, Program Design with Data Structures and Algorithms, Introduction to Object-Oriented Programming, Logic for Systems

## TECHNICAL SKILLS

**Programming Languages:** Java, React/TypeScript, C, C++, C#, HTML, CSS, and Python

**Software:** Adobe Creative Suite, SQL Server, ASP.Net

**Certifications:** AWS Cloud Practitioner

## PROJECTS

### Maps | Java / React / TypeScript

- Developed a full-stack web application mapping historical redlining data from the U.S. Census using Mapbox for interactive and customizable map views.
- Implemented API endpoints to overlay redlining data with advanced filtering and search capabilities, enhancing user experience and data accessibility.
- Integrated user-specific map pins linked to Google Firestore, allowing users to securely add, store, and delete pins, persisting data across sessions.

### Eco Equity | Java / React / TypeScript

- Created a full-stack application with Java for backend API connections and React/TypeScript for a dynamic, responsive front-end user experience.
- Integrated Google's Moshi Adapter for efficient JSON parsing from API requests, converting data into usable objects.
- Enabled sustainably conscious investors to access ESG and financial data, allowing them to favorite, search, and filter companies and relevant financial articles through seamless Google Firestore and Microsoft Bing integration.

## WORK EXPERIENCE

### Accenture, Technology Architecture Analyst

Boston, MA | Summer 2024

- Supported a \$45M Higher Education Workday implementation to bridge international campus financial systems
- Led a successful integration of Grants and Customer Accounts, achieving 35% completion of client data migration
- Built a dynamic excel tracking system to optimize project management and milestone tracking for 53 deliverables
- Developed a customer confirmation deck to ensure project alignment with client needs and facilitate decision-making

### Novanta, Cambridge Technology, Software Engineering Intern

Bedford, MA | Summer 2023

- Developed a full-stack web application using C# and HTML, enhancing UI/UX and ensuring smooth data handling with an asynchronously updated SQL server
- Gained proficiency in frameworks like Avalonia UI, Microsoft Blazor, and ASP.Net for web development
- Presented web application to a group of 10 engineers and business executives to support a seamless deployment
- Ensured project alignment through bi-weekly meetings with engineering team and senior leadership

### Brown University, Community Coordinator

Providence, RI | Fall 2022 - Current

- Developed a vibrant residential community for 110 residents, assuming the role of a peer leader
- Addressed resident concerns and mediated conflicts in weekly community hours, enhancing satisfaction
- Execute 3 large-scale educational and wellbeing programs each semester, boosting community participation

## LEADERSHIP AND PROFESSIONAL DEVELOPMENT

### Beta Omega Chi, Chapter President

Providence RI | Spring 2024 – Current

- Directed a 10-member board, organizing bi-weekly meetings to ensure task completion and team collaboration
- Raised \$15,000 through fundraising initiatives, funding community outreach and chapter activities
- Implemented diversity, equity, and inclusion initiatives into recruitment, ensuring a culture of safety for all members
- Launched an alumni mentorship program, pairing members with experienced alums for professional development

### National Society of Black Engineers, Executive Board Member

Providence, RI | Fall 2022 - Current

- Boosted member engagement and attendance by 25% through sponsor-backed events and corporate raffles
- Organized community service projects with local nonprofits, increasing chapter visibility and impact
- Led career development workshops, increasing participation by 30% and securing sponsorships from top tech firms
- Led collaborations with Fortune 500 companies, creating unique recruitment opportunities for chapter members

## INTERESTS

Basketball, photography, weight training, leisure autobiographical and fictional reading, travel

# Yonas Amha

Providence, RI | Phone: (703) 772-1887 | E-mail: [yonas\\_amha@brown.edu](mailto:yonas_amha@brown.edu) |  
[linkedin.com/in/yonas-amha-a85a5525b](https://www.linkedin.com/in/yonas-amha-a85a5525b) | [github.com/yonasA2025](https://github.com/yonasA2025)

## EDUCATION

**Brown University, M.S. Computer Science**

Providence, RI | **Expected Graduation May 2026**

Relevant Courses: *Machine Learning, Reintegrating AI, Computer Security, Computer Architecture*

**Brown University, B.S. Computer Science, 3.6/4.0 GPA**

Providence, RI | **September 2021 - May 2025**

Relevant Courses: *Computer Systems, Program Design: Data Structure & Algorithm, Logic of Computer Systems, Cryptographic Systems, Software Security Exploitation, Computational Molecular Bio, Applied Cryptography*

## EXPERIENCE

**HP Inc., DevOps Engineer**

Houston, TX | June 2024 – September 2024

- Assigned to the Data Science Scrum Team for the launch of HP AI Studio. Implemented database management, telemetry, and API connection using MongoDB, Azure DevOps, Kubernetes, and GoLand

**Communities Together, Web App Development**

New York, NY | July 2023 – Present

- Built a website that prepares Amharic speakers to take the U.S. Civics Test in Amharic. The website is equipped with full text-to-speech capabilities. Website link: <https://usacivictest.web.app/>
- Website uses Azure Text-to-Speech

**Ringle, Tutor**

Fairfax, VA | June 2023 – September 2023

- Tutored 18 high school students in weekly English lessons, advising them on grammar, vocabulary and sentence structures
- Assigned a reading for each lesson and sent detailed feedback analyzing student's performance during the lesson along with exercises and worksheets to target areas of weakness

**Second Brain Foods, Product Marketing Intern**

Alexandria, VA | June 2022 – September 2022

- Assisted with packaging, presentation and pricing of the health snacks product-line for the startup company
- Conducted market analysis on six barley products within the company's main line focusing on performance and visual appeal

## PROJECTS

**Database | C**

[github.com/yonasA2025/database](https://github.com/yonasA2025/database)

- Created a fully functioning database in which the user can store, pop and search for given values
- Implemented thread safety with coarse and fine grain locking

**Amharic Civics Test | TypeScript | JavaScript | HTML | CSS**

[github.com/yonasA2025/amhariccivictest-v3](https://github.com/yonasA2025/amhariccivictest-v3)

- Built a website that prepares Amharic speakers to take the U.S. Civics Test in their native language

**Ratty Reviews | JAVA | TypeScript**

[github.com/yonasA2025/rattyReviews](https://github.com/yonasA2025/rattyReviews)

- Created a website that allows users to rate the meals that they have at the Ratty (popular Brown dining hall) on any given day
- The website uses web scraping to acquire the meals from [brown.dining.edu](https://brown.dining.edu) and uses caching to store user ratings
- Google authentication was implemented so the user could sign into their gmail to submit a rating

**Maze | C**

<https://github.com/yonasA2025/maze>

- Built a 2D maze generator that could generate up to 25x25 size mazes
- Generated solutions to the maze using Breadth First Search (BFS), Depth First Search (DFS), and Dijkstra Algorithms and then encoded the solution in hexadecimal

**Search | Python**

[github.com/yonasA2025/search](https://github.com/yonasA2025/search)

- Created a search engine that queries all wikipedia pages for any given search
- Search engine implements PageRank and handles WikiLinks

## LEADERSHIP & PROFESSIONAL DEVELOPMENT

**Brown Ethiopian Eritrean Student Association, Senior Executive Board Member**

Providence, RI | October 2022 – Present

- Design Instagram posts, merch, and coordinate with other clubs around promoting events and general board meetings
- Design posts have yielded an 70% increase in engagement for the Instagram page

**Merrone Teklu for Alexandria City Council, Election Campaign Volunteer**

Alexandria, VA | May 2021 – July 2021

- Increased public awareness and advertising for Meronne Teklu, an Alexandria City Council candidate by actively engaging in phone banking, election-day campaigning, and distributing campaign flyers
- Participated in bi-weekly phone banking in the lead up to the election and called 100 voters and encouraged them to get to the polls and provided closest voting locations

**Georgia Canvassing Project, Election Outreach Volunteer**

Atlanta, GA | October 2020 – November 2020

- Collaborated on a team of 50 volunteers to help canvas voters for Georgia's runoff election
- Separated a list of 5,000 names into various nationalities so that they could be canvassed in their native language.

## SKILLS

**Front-End Development:** React, TypeScript, JavaScript, Java, Python, C

**Languages:** English (fluent), Spanish (fluent), Amharic (speaking: fluent)

**Kenny Gruchalla, Ph.D.**  
**Curriculum Vitae**  
(June 6, 2025)

## Personal Information

---

**Postal Address:** (available upon request)  
Golden, CO 80401

**Office:** 303-275-3713

**Cell:** 720-394-9347

**E-mail:** kenny.gruchalla@nrel.gov

**Web:** <http://gruchalla.github.io>

## Education

---

**2009**      **Ph.D. Computer Science**, *University of Colorado at Boulder*, Boulder, CO  
*Thesis:* Progressive Visualization-Driven Multivariate Feature Definition and Analysis  
*Advisor:* Professor Elizabeth Bradley  
*GPA:* 3.9/4.0

**2003**      **M.S. Computer Science**, *University of Colorado at Boulder*, Boulder, CO  
*Thesis:* Immersive Well-Path Planning: Investigating the added value of immersive visualization  
*Advisor:* Professor Clayton Lewis  
*GPA:* 3.9/4.0

**1995**      **B.S. Computer Science**, *New Mexico Institute of Mining and Technology*, Socorro, NM  
*GPA:* 3.5/4.0

## Experience

---

### National Renewable Energy Laboratory (NREL), Golden, CO (June 2009 - present)

#### *Distinguished Member of Research Staff*

*Jul 2024 - present*      **Principal Scientist**

*Jun 2009 - Jul 2024*      **Senior Scientist**

My core responsibilities include the design and management of a state-of-the-art visualization laboratory, collaboration with NREL domain scientists to visualize complex, large, multivariate data, and frequent interaction with C-level executives to align lab capabilities with their needs and expectations. As a Principal Investigator, I successfully led a portfolio of over \$7 million in competitively funded research projects, managing diverse technical teams.

### University of Colorado at Boulder, Boulder, CO (April 2001 - Mar 2006, May 2011 - present)

*Mar 2025 - present*      **Assistant Professor Adjoint**

*May 2011 - Mar 2025*      **Assistant Professor Adjunct,**  
*Department of Computer Science*

I provide mentorship to doctoral-level students in the field of scientific data visualization, supporting the development of graduate students' research skills, including project design, data analysis, software development and effective visualization techniques.

*Apr 2001 - Mar 2006*      **Professional Research Assistant,**

*CADSWES (Center for Advanced Decision Support for Water and Environmental Systems)*

I worked in an interdisciplinary research center on the design and the development of a commercial graphically-based decision support software system, *RiverWare*, implementing object-oriented simulation, rule-based simulation, and linear optimization to model watershed physical processes, water ownership, and policy.

*Jan 2004 - Aug 2004*      **Professional Research Assistant,**

*Department of Molecular, Cellular, and Developmental Biology*

I collaborated on the design and development of a pilot study to investigate the added value of using immersive visualization as a molecular research tool.

*Jul 2002 - Aug 2003*

**Graduate Research Assistant,**  
*BP Center for Visualization*

I designed and developed an interactive 3D immersive application capable of integrating geological, geophysical, reservoir and well data with drilling and platform planning in an immersive virtual environment.

**Colorado State University**, Fort Collins, CO (October 2021 - present)

*Oct 2021 - present*

**Affiliate Faculty,**  
*Richardson Design Center*

As an Advisor for the Richardson Design Center (RDC), I support the development and execution of their graduate certificate program in human-centered design thinking. I bring industry experience and technical expertise to my advising role, providing a valuable resource for faculty and students as they navigate the rapidly evolving field of VR.

**Red Canyon Engineering**, Denver, CO (June 2000 - January 2009)

*Jun 2003 - Jun 2009*

**Research Scientist / Principal Software Engineer**

*Jun 2000 - Jun 2003*

**Software Engineering Consultant**

I spearheaded the creation of new business proposals and successfully led two NASA-funded projects - the Mars Flight Simulator and Lunar Base Simulator. In addition, I conducted comprehensive software architecture and algorithm reviews for the Mars Odyssey and Genesis spacecraft programs, ensuring technical accuracy and project success.

**National Center for Atmospheric Research (NCAR)**, Boulder, CO (May 2006 - October 2008)

*May 2006 - Oct 2008*

**Visitor Appointment**

I collaborated on the design and development of VAPOR, a state-of-the-art volume visualization suite designed to interactively explore large-scale time-varying multivariate computational fluid dynamics (CFD) data.

**Raytheon**, Aurora, CO (July 1995 - February 2001)

*Jul 2000 - Feb 2001*

**Senior Analyst / Medical Officer,**  
*United States Antarctic Program*  
*Raytheon Polar Services*

I managed the data acquisition and visualization laboratory aboard the National Science Foundation's Antarctic research vessel, the Nathaniel B. Palmer, providing scientific support to NSF grantees. I also sailed as Medical Officer (EMT) aboard the Palmer.

*Oct 1997 - Jun 2000*

**Technical Software Lead,**  
*Space and Science Systems, Raytheon Systems Corporation*

I designed and developed animated meteorological visualization tools for the Cape Canaveral and Vandenberg space lift ranges that included the development of both real-time and analysis visualization algorithms and image processing software for radar and satellite instrumentation. As the technical lead I served as technical mentor, providing technical guidance across projects and organizations.

*Jul 1995 - Oct 1997*

**Software Developer,**  
*Satellite Mission Management Organization, Hughes Space Systems* (purchased by Raytheon in 1997)

I designed and developed a distributed satellite mission planning and scheduling software system that included interactive 2D computer graphic models of satellite and ground station resource allocation and 3D modeling tools used for satellite payload constraint analysis. I also helped design, develop, and maintain object-oriented class libraries designed for reuse and rapid development of satellite space and ground applications.

**Brookhaven National Laboratory (BNL)**, Upton, NY (January 1994 - May 1994)

*Jan 1994 - May 1994*

**Science and Engineering Research Intern,**  
*Advanced Technology Division*

I designed and developed interactive 2D visualizations of subsurface radioactive waste plumes created by a physically-based model of the breach, leach, and transport of radioactive waste material.

| <b>Funding (selected)</b> (>\$7M total competitive awards as PI or Co-PI)                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>August 2024</i>                                                                       | EVision 2040: Visualizing a Path Toward Sustainable Mobility, <b>PI</b><br>JOET                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <i>July 2024</i>                                                                         | Immersive Visualization Architecture for the NYS Power Grid Digital Twin, <b>PI</b><br>WFO-New York Power Authority                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <i>December 2023</i>                                                                     | DEWA Energy Visualization Analysis, <b>co-PI</b><br>WFO-Dubai Energy and Water Authority                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <i>August 2023</i>                                                                       | Immersive Digital Twin Laboratory for Engineering Education, <b>PI</b><br>WFO-Fort Lewis College                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <i>October 2020</i>                                                                      | Collaborative Visualization and Analysis for the Control Room of the Future, <b>PI</b><br>DOE/NREL LDRD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <i>August 2019</i>                                                                       | Cluster Perception in Scatterplots, <b>PI</b><br>DOE/NREL LDRD/PI4                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <i>April 2016</i>                                                                        | Advanced Energy Systems Design Architecture, <b>PI</b><br>DOE/NREL LDRD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <i>September 2014</i>                                                                    | Immersive Computational Steering & Modeling, <b>PI</b><br>DOE/NREL LDRD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <i>September 2014</i>                                                                    | Visualization and Simulation of a Manufacturing Line, <b>PI</b><br>WFO-Abengoa                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <i>January 2013</i>                                                                      | Integrated Energy Management and Analysis for the ESIF's Computational Systems, <b>Co-PI</b><br>DOE/NREL LDRD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <i>September 2012</i>                                                                    | Novel Visualization and Analysis for Extreme-Scale Wind Turbine Array Simulations, <b>PI</b><br>DOE/NREL LDRD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <i>September 2010</i>                                                                    | Wind Turbine Array Fluid Dynamic and Aero-Elastic Simulations, <b>Co-investigator</b><br>DOE/NREL LDRD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <i>January 2009</i>                                                                      | Lunar Base Simulator, <b>PI</b><br>NASA Glenn Research Center                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <i>March 2005</i>                                                                        | MarsFlight – An Immersive and Interactive Mars Airplane Simulator, <b>Co-PI</b><br>NASA Glenn Research Center                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <i>March 2003</i>                                                                        | Immersive Technology for Engineering Education, <b>Co-investigator</b><br>University of Colorado Engineering Excellence Fund                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Publications</b> (1000+ citations, g-index:31, h-index:15, i10-index:25, i1-index:56) |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <i>Book Chapters: (2)</i>                                                                | <ul style="list-style-type: none"> <li>E. Gould, S. Guerin, C. Smith, S. Smith, B. Bush and K. Gruchalla. Learning and Tracking Ad Hoc Fiducial Markers in Spatial Augmented Reality In J.G. Michopoulos, C.J.J. Paredis, D.W. Rosen, J.M. Vance (Eds.), <i>Advances in Computers and Information in Engineering Research Vol. 2</i> ASME Press June 2021.</li> <li>K. Gruchalla and N. Brunhart-Lupo. The Utility of Virtual Reality for Science and Engineering In W.R. Sherman (Ed.), <i>VR Developer Gems</i>, CRC Press, May 2019.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <i>Refereed Papers: (40)</i>                                                             | <ul style="list-style-type: none"> <li>S. Molnar, G. Johnson, K. Gruchalla, K. Potter. Opportunities and Challenges in the Visualization of Energy Scenarios for Decision-Making <i>Energy Visualization 2024</i>, November 2024.</li> <li>N. Brunhart-Lupo, K. Gruchalla, L. Williams, S. Elias. Situated Visualization of Photovoltaic Module Performance for Workforce Development. <i>Energy Visualization 2024</i>, November 2024.</li> <li>G. Johnson, S. Molnar, N. Brunhart-Lupo, K. Gruchalla. Architecture for Web-Based Visualization of Large-Scale Energy Domains. <i>Energy Visualization 2024</i>, November 2024.</li> <li>G. Johnson, K. Gruchalla, M. Ingram, N. Guba, R. Cruickshank, and S. Caruso. Vis-SAGA: Visual Analytics for Situational Awareness of Grid Anomalies <i>VizSec2023</i>, October 2023.</li> <li>S. Molnar, K. Gruchalla, S. Dong, and J. Tan. Visualization of the Oscillatory Dynamics of an Island Power System. <i>Energy Visualization 2023</i>, October 2023.</li> <li>I. Lyons-Galante, M. Karimzadeh, S. Molnar, G. Johnson, and K. Gruchalla. Alternatives to Contour Visualizations for Power Systems Data. <i>Energy Visualization 2023</i>, October 2023.</li> <li>N. Brunhart-Lupo and K. Gruchalla. Immersive Particle Advection: Through the Scales of Renewable Energy. <i>ACM PEARC23</i>, July 2023.</li> <li>K. Gruchalla, S. Molnar, and G. Johnson. Reevaluating Contour Visualizations for Power Systems Data. <i>IEEE Transactions on Smart Grid</i>, March 2023.</li> </ul> |



- A. Podder, K. Gruchalla, N. Brunhart-Lupo, S. Pless, M. Sica and P. Lacchin. Immersive Industrialized Construction Environments for Energy Efficiency Construction Workforce. *Frontiers in Virtual Reality*, March 2022.
- S. Molnar, E. Bradley, K. Gruchalla. Oscillatory Spreading and Inertia in Power Grids. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, December 2021.
- H. Sitaraman, N. Brunhart-Lupo, M. Henry de Frahan, S. Yellapantula, B. Perry, J. Rood, R. Grout, M. Day, R. Binyahib, and K. Gruchalla. Visualizations of direct fuel injection effects in a supersonic cavity flameholder. *Physical Review Fluids*, November 2021.
- K. Gruchalla, S. Raghupathi, N. Brunhart-Lupo. Structure Perception in 3D Point Clouds *ACM SAP 2021*, September 2021.
- N. Brunhart-Lupo, S. Yellapantula, K. Gruchalla, R. Grout. Visualization of Jet Impingement and Ignition in a Piston-cylinder Chamber. *Energy Visualization 2021*, June 2021.
- M. Whitlock, D. Albers Szafir, K. Gruchalla. HydrogenAR: Interactive Data-Driven Presentation of Dispenser Reliability. *ISMAR 2020*, November 2020.
- N. Brunhart-Lupo, B. Bush, K. Gruchalla, K. Potter, S. Smith. Collaborative Exploration of Scientific Datasets using Immersive and Statistical Visualization. *Improving Scientific Software. NCAR/TB-564+PROC*, August 2020.
- A.L. Figueroa-Acevedo, C.-H. Tsai, K. Gruchalla, Z. Claes, S. Foley, J. Bakke, J. Okullo, A.J. Prabhakar. Visualizing the impacts of renewable energy growth in the U.S. Midcontinent. *IEEE Open Access Journal of Power and Energy*, January 2020.
- B. Bugbee, B.W. Bush, K. Gruchalla, K. Potter, N. Brunhart-Lupo, V. Krishnan. Enabling Immersive Engagement in Energy System Models with Deep Learning. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, June 2019.
- P.A. Fleming, J. Annoni, L.A. Martinez, S. Raach, K. Gruchalla, A. Scholbrock, M. Churchfield, and J. Roadman. Investigation into the shape of a wake of a yawed full-scale turbine *Journal of Physics: Conference Series* volume 1037, June 2018.
- S. Molnar and K. Gruchalla. Visualizing Electrical Power Systems as Flow Fields. In proceedings of *Workshop for Visualisation in Environmental Sciences (EnvirVis 2018)*, June 2018.
- P.A. Fleming, J. Annoni, M. Churchfield, L. Martinez, K. Gruchalla, M.J. Lawson, and P.J. Moriarty. A simulation study demonstrating the importance of large-scale trailing vortices in wake steering *Wind Energy Science*, volume 3, May 2018.
- P.A. Fleming, J. Annoni, M.J. Churchfield, L. Martinez, K.M. Gruchalla, M.J. Lawson, and P.J. Moriarty. From wake steering to flow control *Wind Energy Science Discussions*, November 2017.
- K. Gruchalla, N. Brunhart-Lupo, K. Potter, J. Clyne. Contextual Compression of Large-Scale Wind Turbine Array Simulations. In proceedings of *Workshop for Data Reduction for Big Scientific Data (DRBSD-2)*, November 2017.
- B. Bush, N. Brunhart-Lupo, B. Bugbee, V. Krishnan, K. Potter, and K. Gruchalla. Coupling Visualization, Simulation, and Deep Learning for Ensemble Steering of Complex Energy Models. In proceedings of *Data Systems for Interactive Analysis DSIA'17*, October 2017.
- B. Bugbee, C. Phillips, H. Egan, R. Elmore, K. Gruchalla, and A. Purkayastha. Prediction and characterization of application power use in a high-performance computing environment. *Statistical Analysis and Data Mining: The ASA Data Science Journal* February, 2017.
- G.A. Ferguson, V. Vorotnikov, N. Wunder, J. Clark, K. Gruchalla, T. Bartholomew, D.J. Robichaud, and G.T. Beckham. Ab Initio Surface Phase Diagrams for Coadsorption of Aromatics and Hydrogen on the Pt(111) Surface. *The Journal of Physical Chemistry C*, November 2016.
- K. Gruchalla, J. Novacheck, A. Bloom. Visualization of the Eastern Renewable Generation Integration Study. In proceedings of *SC16*, Salt Lake City, UT, November 2016.
- C. Zhang, S. Santhanagopalan, M.J. Stock, N. Brunhart-Lupo, K. Gruchalla. Interpretation of Simultaneous Mechanical-Electrical-Thermal Failure in a Lithium-Ion Battery Module. In proceedings of *SC16*, Salt Lake City, UT, November 2016.
- D. Macumber, K. Gruchalla, N. Brunhart-Lupo, M. Gleason, J. Abbot-Whitley, J. Robertson, B. Polly, K. Fleming, M. Schott. City Scale Modeling with OpenStudio. In *proceedings of ASHRAE and IBPSA-USA SimBuild 2016*.

- N. Brunhart-Lupo, B. Bush, K. Gruchalla, S. Smith. Simulation Exploration through Immersive Parallel Planes. In proceedings of *IEEE Workshop on Immersive Analytics*, March 2016.
- J. Hinkle, P.N. Ciesielski, K. Gruchalla, K.R. Munch, B.S. Donohoe. Biomass accessibility analysis using electron tomography. *Biotechnology for Biofuels* volume 8, November 2015.
- S. Li, K. Gruchalla, K. Potter, J. Clyne, and H. Childs. Evaluating the Efficacy of Wavelet Configurations on Turbulent-Flow Data. in *proceedings of IEEE Symposium on Large Data Analysis and Visualization*, Chicago, IL, 2015.
- M. Lunacek, A. Nag, D. Alber, K. Gruchalla, C.H. Chang, and P.A. Graf. Simulation characterization and optimization of metabolic models with the high-performance systems biology toolkit. *SIAM Journal on Scientific Computing*, volume 33, pages 3402-3424, 2011.
- K. Gruchalla, M. Rast, E. Bradley, and P. Mininni. Segmentation and visualization of multivariate features using feature-local distributions. In *Advances in Visual Computing*, volume 6938 of *Lecture Notes in Computer Science*, pages 619–628. Springer Berlin / Heidelberg, 2011.
- M.A. Sprague, P.J. Moriarty, M.J. Churchfield, K. Gruchalla, S. Lee, J.K. Lundquist, J. Michalakes, and A. Purkayastha. Computational modeling of wind-plant aerodynamics. In proceedings of *SciDAC 2011, Denver, CO*, 2011.
- M. Guy, P. Earle, C. Ostrum, K. Gruchalla, and S. Horvath. Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies. In *Advances in Intelligent Data Analysis IX*, volume 6065 of *Lecture Notes in Computer Science*, pages 42–53. Springer Berlin / Heidelberg, 2010.
- J. Clyne, K. Gruchalla, and M. Rast. VAPOR: Visual, Statistical, and Structural Analysis of Astrophysical Flows. In proceedings of *Numerical Modeling of Space Plasma Flows: Astronom-2009 (Astronomical Society of the Pacific Conference Series)*, volume 429, pages 323-329, 2010.
- K. Gruchalla, M. Rast, E. Bradley, J. Clyne, and P. Mininni. Visualization-driven structural and statistical analysis of turbulent flows. In *Advances in Intelligent Data Analysis VIII*, volume 5772 of *Lecture Notes in Computer Science*, pages 321–332. Springer Berlin / Heidelberg, 2009.
- K. Gruchalla, M. Dubin, J. Marbach, and E. Bradley. Immersive examination of the qualitative structure of biomolecules. In proceedings of *International Workshop on Qualitative Reasoning about Physical Systems*, pages 36–41, 2008.
- K. Gruchalla, J. Marbach, and M. Dubin. Porting legacy applications to immersive virtual environments - a case study. In proceedings of *International Conference on Computer Graphics Theory and Applications (GRAPP 2007)*, pages 179–184, 2007.
- K. Gruchalla. Immersive well-path editing: investigating the added value of immersion. In proceedings of *IEEE Virtual Reality, 2004.*, pages 157 – 164, March 2004.
- K. Gruchalla. Immersive Digital Twin Laboratory for Engineering Education: Cooperative Research and Development Final Report, *CRADA Number CRD-23-23838, NREL/TP-2C00-93161*, June 2025.
- S.L. Choi, R. Jain, P. Emami, K. Wadsack, F. Ding, H. Sun, K. Gruchalla, J. Hong, H. Zhang, X. Zhu, and B. Kroposki. eGridGPT: Trustworthy AI in the Control Room NREL Technical Report *NREL/TP-5D00-87740*, May 2024.
- I. Manogaran, A. Farthing, J. Maguire, K. Gruchalla. Savings in Action: Lessons Learned for a Vermont Community with Solar plus Storage NREL Technical Report *NREL/TP-7A40-84660*, January 2024.
- M. Ingram, A. Florita, M. Martin; K. Gruchalla, X. Fang, M. Cai, G. Johnson, N. Guba, A. Hasandka, M. Mooney, et al. Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics NREL Technical Report No. *NREL/TP-5D00-85187*, February 2023.
- K. Gruchalla, G. Johnson. Heterogeneous Visualization: A Call for Empirically Driven Technology Advancement and Adoption. *ASCR Workshop on Visualization for Scientific Discovery, Decision-Making, and Communication*, January 2022
- R. King, A. Glaws, K. Gruchalla, M. Henry de Frahan. Extracting Maximum Information from Exascale Simulations *ASCR Workshop on Reimaging Codesign*, March 2021.
- M. Lawson, J. Melvin, S. Ananthan, K. Gruchalla, J. Rood, M. Sprague. Blade-Resolved, Single-Turbine Simulations Under Atmospheric Flow NREL Technical Report No. *NREL/TP-5000-72760*, January 2019.

*Non-refereed Papers: (17)*

- B. Palmintier, J. Giraldez, K. Gruchalla, P. Gotseff, A. Nagarajan, T. Harris, B. Bugbee, M. Baggu, J. Gantz, E. Boardman. Feeder Voltage Regulation with High Penetration PV using Advanced Inverters and a Distribution Management System: A Duke Energy Case Study NREL Technical Report No. *NREL/TP-5D00-65551*, November 2016.
- A. Bloom, A. Townsend, D. Palchak, J. Novacheck, J. King, C. Barrows, E. Ibanez, M. O'Connell, G. Jordan, B. Roberts, C. Draxl, K. Gruchalla. Eastern Renewable Generation Integration Study. NREL Technical Report No. *NREL/TP-6A20-64472*, August 2016.  
**NREL 2017 Innovation & Technology Transfer Outstanding Public Information Award**
- R. Elmore, K. Gruchalla, C. Phillips, A. Purkayastha, N. Wunder. An Analysis of Application Power and Schedule Composition in a High-Performance Computing Environment. NREL Technical Report No. *NREL/TP-2C00-65392*, January 2016.
- R.W. Grout, K. Malhorta, P. Ciesielski, K. Gruchalla, B. Donohoe, M. Nimlos. Computational Assessment of the effect of realistic intraparticle geometry on biomass heating rates and pyrolysis yields. *8th US National Combustion Meeting*, May 2013.
- G. Pech, K. Gruchalla, and J. Marbach. The case for visualization. *Exploration & Production*, January 2009.
- G.A. Dorn, G.S. Pech, K. Gruchalla, J. Marbach The Value of Visualization in Exploration and Production: Anecdotal Evidence and Quantitative Data *70th EAGE Conference & Exhibition-Workshops and Fieldtrips* June 2008.
- K. Gruchalla and J. Marbach. Interactively exploring multiple characteristics of hurricane simulation data. *Advanced Imaging*, 22, 2005.
- G. Dorn, K. Gruchalla, J. Carlson, J. Marbach, T. Southren, and A. Jamieson. A visualization-driven paradigm for adaptive well-path planning. In *Offshore Technology Conference*, 2004.
- K. Gruchalla and E. Joynt. Late Cretaceous and Cenozoic Reconstructions of the Southwest Pacific, Data Report NBP00-07B. United States Antarctic Program, 2000.
- K. Bliss, K. Gruchalla, and K. Gavahan. OBS Refraction Profiling for Crustal Structure in Bransfield Strait, Data Report NBP00-07A. United States Antarctic Program, 2000.
- H. Sitaraman, M. Henry De Frahan, S. Yellapantula, B. Perry, J. Rood, R. Grout, M. Day, N. Brunhart-Lupo, R. Binyahib, K. Gruchalla. Direct fuel injection effects in a supersonic cavity flameholder *APS Division of Fluid Dynamics Gallery of Fluid Motion*, Chicago, IL, 2020.  
**Gallery Award**
- S. Witter Hicks, M. Churchfield, K. Gruchalla. Visualization of a Simulated LiDAR-Based Wind Turbine Wake Measurement Campaign, *SuperComputing 2016 (SC16)*, Salt Lake City, UT, November, 2016.
- K. Gruchalla, M.J. Churchfield, P.J. Moriarty, , S. Lee S. Li, J.K. Lundquist, J. Michalakes, A. Purkayastha, and M.A. Sprague. Computational modeling of turbine-wake effects. *SciDAC 2011*, Denver, CO, 2011.  
**Awarded SciDAC 2011 People's Choice Award**
- K. Gruchalla, O. Desjardins, P. Pepiot, and A. Purkayastha. Numerical simulation of a turbulent liquid jet. *SuperComputing 2010 (SC10)*, New Orleans, LA, 2010.
- K. Gruchalla, M.J. Churchfield, P.J. Moriarty, and L. Martinez. Eddy simulation of wind farm / atmospheric boundary layer interaction. *SuperComputing 2010 (SC10)*, New Orleans, LA, 2010.
- K. Gruchalla, P. Pepiot, and O. Desjardins. Particle dynamics in a fluidized bed reactor. *SciDAC 2010*, Chattanooga, TN, 2010.  
**Awarded SciDAC 2010 Outstanding Achievement in Scientific Visualization**
- K. Gruchalla and J. Marbach. Atmosv: Immersive visualization of the hurricane Isabel dataset. *IEEE Visualization 2004*, Austin, TX, 2004.  
**Awarded second place in the 2004 IEEE Visualization Contest**
- I. Lyons-Galante, M. Karimzadeh, and K. Gruchalla. Alternatives to Contour Visualization for Power Systems Data. <https://github.com/isaiahlg/voltage-tiling>, 2023.
- K. Gruchalla, S. Molnar, and G. Johnson. Visualization Encoding Experiments for Power Systems Analysis. National Renewable Energy Laboratory. <https://data.nrel.gov/submissions/200>, 2022.
- S. Raghupathi, N. Brunhart-Lupo, and K. Gruchalla. Caerbannog Point Clouds. National Renewable Energy Laboratory. 10.7799/1729892, 2020

*Video (selected):*

*Published Data (4):*

- C. Phillips, K. Gruchalla, R. Elmore, A. Purkayastha, and N. Wunder. HPC Energy Research. National Renewable Energy Laboratory. <https://cscdata.nrel.gov/#/datasets/d332818f-ef57-4189-ba1d-beea291886eb>, 2017.
- Preregistrations (2):**
- K. Gruchalla, G. Johnson, and S. Molnar. *Voltage Outlier Detection*. OSF Preregistration, <https://osf.io/tmv6s>, July 2022.
  - K. Gruchalla, G. Johnson, and S. Molnar. *Frequency Distribution Analysis*. OSF Preregistration, <https://osf.io/zn7ag>, July 2022.
- Motion Picture Credits (1):**
- K. Gruchalla. Computer Graphics Rendering, *Green China Rising*. National Geographic Film, 2012.
- Thesis (2)**
- K. Gruchalla. *Progressive Visualization-Driven Multivariate Feature Definition and Analysis*. PhD thesis, University of Colorado at Boulder, 2009.
  - K. Gruchalla. *Immersive well-path planning: The added value of immersive visualization*. Master's thesis, University of Colorado at Boulder, 2003.
- Posters (selected):**
- M. Sprague, S. Ananthan, K. Gruchalla, M. Lawson, J. Rood, K. Swirydowicz, S. Thomas, G. Vijayakumar, P. Crozier, C.R. Dohrmann, J.J. Hu, A.B. Williams, J. Turner, A. Prokopenko, R. Moser, and J. Melvin. ExaWind: Exascale Predictive Wind Plant Flow Physics Modeling. Poster at *2019 Exascale Computing Project Annual Meeting*, Houston, TX. January 2019.
  - B. Bugbee, C. Phillips, K. Gruchalla, R. Elmore, and A. Purkayastha. Exploring HPC Application Power Usage. Poster at *2016 Conference on Data Analysis (CODA)*, Santa Fe, NM, 2016.
  - J. Hinkle, P. Ciesielski, K. Gruchalla, B. Donohoe, and K. Munch. Segmentation of Lamellar Sheets of Cellulose using Electron Tomography. Poster at *2014 Annual BioEnergy Science Center (BESC) Retreat*, Chattanooga, TN, 2014.
  - H. Scharf, K. Gruchalla, R. Elmore, N. Brunhart-Lupo, A. Purkayastha. Optimal Prioritized Compression using Wavelets for Analysis and Visualization of Extreme-Scale Wind Turbine Simulations. Poster at *2014 Conference on Data Analysis (CODA)*, Santa Fe, NM, 2014.
  - R. Elmore, M. Sheppy, N. Wunder, K. Munch, K. Gruchalla, A. Purkayastha. A Case Study in Demand-Response Strategies for Managing Power in HPC Environments. Poster at *2014 Conference on Data Analysis (CODA)*, Santa Fe, NM, 2014.
  - M. Dubin, A. Pardi, and K. Gruchalla. Using immersive virtual reality for visualization of macromolecules. Poster at *2004 Butcher Symposium on Genomics and Biotechnology*, Boulder, CO, 2004.
  - K. Gruchalla and J. Marbach. Atmosv: Immersive visualization of the Hurricane Isabel dataset. Contest Entry at *IEEE Visualization 2004*, Austin, TX, 2004.  
**Awarded second place in the 2004 IEEE Visualization Contest**
- Press (selected):**
- 38th Annual Gallery of Fluid Motion award winners announced, *AAAS EurekAlert!*, November 2020.
  - The Man Behind the Goggles: Meet NREL's Visualization Guru, *NREL News*, June 3, 2019.
  - R. Gold, Superpower: One Man's Quest to Transform American Energy. Simon & Schuster, June 2019.
  - Decoding the Weather Machine. *PBS Nova*, Episode 7, Season 45, April 2018.
  - Scientists Break This Virtual Power Grid to Save the Real One. *Popular Mechanics*, July/August 2015.
  - Daley, D. Visualize the Future: Simulation helps NREL troubleshoot scenarios yet to happen. *Sound & Communications*, August 2014. (Cover article)
  - Scanlon, W. Scientists go eye to eye with research at ESIF. *NREL News Feature*, July 2013. ([http://www.nrel.gov/news/features/feature\\_detail.cfm?feature\\_id=2254](http://www.nrel.gov/news/features/feature_detail.cfm?feature_id=2254))
  - Pierce, E.R. Tour the National Renewable Energy Lab's Latest Research Center *energy.gov* June 2013. (<http://energy.gov/articles/slideshow-tour-national-renewable-energy-lab-s-latest-research-center>)
  - Tucker, E. Supercomputing Drives Innovation. *Continuum Magazine*, (2), 6-9. 2012.
  - Mosher, D. 10 Award-Winning Scientific Simulation Videos *Wired.com Wired Science* August 2011. (<http://www.wired.com/wiredscience/2011/08/science-simulation-videos/>)

## Presentations (selected)

---

### *Keynotes (2)*

- “Immersive Analytics: Knowledge Discovery from the Inside,” Keynote Address to the 2016 CSU Virtual Reality Symposium, Fort Collins, CO, October 2016.
- “Scientific Data Analysis and Knowledge Discovery through Advanced Visualization,” Keynote Address to the 2016 CAAV Conference, Denver, CO, June 2016.

### *Invited (selected)*

- “Visualization for Insight, Analysis, and Decision Support in Energy Systems,” NITRD-DOE-NSF DECIDE Workshop: Computational Decision Support for High-Consequence, Complex Problems, Falls Church, VA, March 2025.
- “Immersive Visualization for Scientific Data Analysis and Knowledge Discovery,” Surveying VR Data Visualization Software for Research Seminar, Brown University, Providence, RI (virtual). January 2025.
- “Visualization within the Department of Energy: NREL,” IEEE VIS Application Spotlight. IEEE VIS 2024, (virtual), October 2024
- “Immersive Visualization for Scientific Data Analysis,” Immersive Visualisation for Science, Research, Art and Digital Twins Applications BOF, SIGGRAPH 2024, Denver, CO, August 2024
- “Advanced Visualization for Scientific Data Analysis and Insight,” Computational Science Department Seminar, Princeton Plasma Physics Laboratory, Princeton, NJ (virtual), November 2023.
- “The Future of Visualization Technology,” Visualization in Practice, IEEE VIS, Melbourne, Victoria, Australia, October 2023.
- “Visualization of Energy Systems,” Monash University, Clayton, Victoria, Australia, October 2023.
- “From Data to Discovery at NREL,” Computer Science and Engineering Seminar. New Mexico Tech, Socorro, NM. September 2023.
- “Advanced Visualization for Scientific Data Analysis,” Summer Workshop on Energy Education for Teachers (SWEET). Colorado School of Mines, Golden, CO. June 2023.
- “Sustainable Immersive Visualization: The Tale of Two Visualization Labs,” Workshop on Immersive Visualization Laboratories, IEEE VR 2023, Shanghai, China (virtual). March 2023.
- “Immersive Visualization for Scientific Data Analysis,” VR Data Visualization Software for Research Seminar, Brown University, Providence, RI (virtual). January 2023.
- “Advanced Visualization Infrastructure for Scientific Data Analysis and Knowledge Discovery,” International Computing in the Atmospheric Sciences (iCAS2022). Stresa, Italy. September 2022.
- “Embodied Cognition and Perception: Immersion for Scientific Data Analysis,” The CAAV Speaker Series, Virtual, August 2022.
- “Visualization-Supported Energy Analysis for Resilient Cities,” Workshop on Urban Visualisation: Energy Resilient Cities, Melbourne, Australia (virtual due to COVID). June 2020.
- “NREL Visualization Overview,” Electronic Visualization Lab, University of Illinois, Chicago, IL. June 2019.
- “NREL Visualization Overview,” SCI Institute, University of Utah, Salt Lake City, UT. May 2019.
- “NREL Visualization Overview,” Emerging Analytics Center, University of Arkansas, Little Rock, AR. April 2019.
- “Computational Science,” JeffCon 2019, Golden, CO, January 2019.
- “Immersive Analytics: Knowledge Discovery from the Inside,” LANL Information Science and Technology Institute (ISTI) Seminar, Los Alamos National Laboratory, NM, July 2016.
- “Visualization and Analysis of Large-Scale Wind Turbine Array Simulations,” Boulder Fluids Dynamics Seminar, Boulder, CO, February 2015.
- “Enabling Renewable Energy Research through Scientific Visualization,” Data Visualization Summit, Boston, MA, September 2013.
- “Computational Modeling and Visualization of Turbine-Wake Effects,” Frontiers in Computational Physics, Boulder, CO, December 2012.

*Contributed (selected)*

- “Visualization-Driven Multivariate Feature Analysis using Feature-Local Distributions,” Colorado School of Mines, Joint Colloquia of AMS and EECS, Golden, CO, December 2011.
- “Enabling Renewable Energy Research through Scientific Visualization,” University of Colorado, Department of Computer Science Colloquium, Boulder, CO, March 2011.
- “Statistically Guided Multivariate Visualization and Analysis of Turbulence Structures,” National Renewable Energy Laboratory (NREL), Golden, CO, April 2009.
- “Interactive visualization and analysis of turbulence structures and their statistics,” Laboratory of Computational Dynamics Turb-Helio Seminar. Boulder, Colorado, February 2009.
- “Extending VAPOR’s hardware-accelerated volume rendering capabilities,” Computational and Information Systems Laboratory (CISL) Seminar, National Center for Atmospheric Research (NCAR). Boulder, Colorado, August 2007.
- “Multivariate volume visualization,” Laboratory of Computational Dynamics Turb-Helio Seminar. Boulder, Colorado, November 2005.
- “Situated Visualization of Photovoltaic Module Performance for Workforce Development,” EnergyVis 2024, IEEE VIS 2024, October 2023.
- “Vis-SAGA: Visual Analytics for Situational Awareness of Grid Anomalies,” IEEE Symposium on Visualization for Cyber Security (VizSec2023), IEEE VIS 2023, Melbourne, Victoria, Australia, October 2023.
- “Visualization of the Oscillatory Dynamics of an Island Power System,” EnergyVis 2023, IEEE VIS 2023, Melbourne, Victoria, Australia, October 2023.
- “Alternatives to Contour Visualizations for Power Systems Data,” EnergyVis 2023, IEEE VIS 2023, Melbourne, Victoria, Australia, October 2023.
- “Reevaluating Heatmaps for Power Systems Visualization,” 2023 DOE Computer Graphics Forum, Idaho National Laboratory, Idaho Falls, ID, April 2023.
- “Immersive Digital Twins and Decision Making,” 2023 DICE Digital Engineering Conference, Idaho National Laboratory, Idaho Falls, ID, April 2023.
- “Finding the Killer Rabbit of Caerbannog,” 2022 DOE Computer Graphics Forum, Eugene, OR (virtual), September 2022.
- “Structure Perception in 3D Point Clouds,” ACM Symposium on Applied Perception, Paris, France (virtual), September 2021.
- “Knowledge Discovery through Immersive Analytics,” High Performance Computing Symposium 2021, Boulder, CO, May 2021.
- “Visualization Paradigms in the Renewable Energy Space,” IEEE VIS 2019, Application Spotlight, Vancouver BC, Canada, October 2019
- “Visualizing Electric Power Systems as Flow Fields,” 2019 DOE Computer Graphics Forum, Monterey, CA, March 2019
- “Contextual Compression of Large-Scale Wind Turbine Array Simulations,” Workshop for Data Reduction for Big Scientific Data, SCI7, Denver, CO, November 2017
- “Coupling Visualization, Simulation, and Deep Learning for Ensemble Steering of Complex Energy Models,” DSIA: Data Systems for Interactive Analysis, Phoenix, AZ, October 2017
- “Visualization of a Simulated LiDAR-Based Wind Turbine Wake Measurement Campaign,” RMACC High Performance Computing Symposium 2017, Boulder, CO, August 2017
- “Interpretation of Simultaneous Mechanical-Electrical-Thermal Failure in a Lithium-Ion Battery Module,” SuperComputing 2016 (SC16), Salt Lake City, UT, November 2016
- “Visualization of the Eastern Renewable Generation Integration Study,” SuperComputing 2016 (SC16), Salt Lake City, UT, November 2016
- “Virtual Reality Panel,” 2016 CSU Virtual Reality Symposium, Fort Collins, CO, October 2016
- “Transmission Grid Visualization,” Grid Modernization Laboratory Consortium: Tools and Data for Production Cost Modeling Workshop, Golden, CO, October 2016.
- “ESIF Insight Center: Visualization Hardware Panel,” 2013 DOE Computer Graphics Forum, Portland, OR, April 2013
- “Quantifying and Meshing Features in Microscopy Data,” 2012 NREL High-Performance Computing Workshop, Golden, CO, August, 2012
- “Visual Analysis of Fluidized Bed Reactor Dynamics,” DOE Computer Graphics Forum, Albuquerque, NM, April 2012.

- “Segmentation and visualization of multivariate features using feature-local distributions,” *International Symposium on Visual Computing 2011*, Las Vegas, NV, September, 2011
- “The Dawn of Scientific Visualization at NREL,” *2010 DOE Computer Graphics Forum*, Park City, Utah, April 2010.
- “Visualization-Driven Structural and Statistical Analysis of Turbulent Flows,” *2009 Intelligent Data Analysis Conference*, Lyon, France, September 2009.
- “Immersive Examination of the Qualitative Structure of Biomolecules,” *2008 International Workshop on Qualitative Reasoning about Physical Systems*, Boulder, Colorado, June 2008.
- “Porting legacy applications to immersive virtual environments – a case study,” *The 2007 International Conference on Computer Graphics Theory and Applications*, Barcelona, Spain, March 2007.
- “Hardware-accelerated visualization of non-uniformly gridded volume data,” *2007 DOE Computer Graphics Forum*, Peaceful Valley, Colorado, May 2007.
- “Accounting network visualization,” 2005 Annual RiverWare User Group Meeting, Boulder, Colorado, March 2005.
- “Optimization and rules policy editor,” 2005 Annual RiverWare User Group Meeting, Boulder, Colorado, March 2005.
- “Workspace migration to Qt,” 2005 Annual RiverWare User Group Meeting, Boulder, Colorado, March 2005.
- “Immersive well-path editing: Investigating the added value of immersion,” *IEEE Virtual Reality 2004 Conference*, Chicago, Illinois, March 2004.
- “The COE Kansas City flood control method,” 2004 Annual RiverWare User Group Meeting, Boulder, Colorado, March 2004.
- “The added value of immersive visualization,” 2004 Drilling Visualization Research Consortium Meeting, Boulder, Colorado, January 2004.
- “Corps of Engineers flood control methods,” 2003 Annual RiverWare User Group Meeting, Boulder, Colorado, June 2003.
- “Plotting simulation data,” 2001 Annual RiverWare User Group Meeting, Boulder, Colorado, June 2001.
- “Visualization and user interface development for breach, leach, and transport models,” Brookhaven National Laboratory SERS Seminar. Upton, NY, May 1994

## Honors and Awards

---

### Research

- Distinguished Visitor, IEEE Computer Society, 2025 Class
- ACM/Eurographics High Performance Graphics (HPG) 2024 Chair’s Award
- NREL December 2024 Employee Team of the Month  
(with J.D. Laurence-Chasen, G. Johnson, and E. Grant)
- NREL Distinguished Member of Research Staff (2021-present)
- APS DFD 2020 Gallery of Fluid Motion Award
- NREL 2018 Innovation & Technology Transfer Outstanding New Partnership Award
- NREL 2017 Innovation & Technology Transfer Outstanding Public Information Award
- NREL May 2016 Employee Team of the Month  
(with N. Brunhart-Lupo and M. Stock)
- NREL FY14 Staff Award – Outstanding Achievement
- NREL 2013 President’s Award
- DOE ACSR SciDAC 2011 People’s Choice Award
- DOE ACSR SciDAC 2010 Outstanding Achievement in Scientific Visualization
- Advanced Imaging Magazine 2005 Imaging Solutions of the Year
- IEEE Visualization 2004 Visualization Contest Second Place

### Industry (selected)

- Raytheon Space and Science Systems CHIP Award
- Raytheon Systems Company Outstanding Achievement Award
- Hughes Space Systems Team Achievement Award

### Academic

- New Mexico Tech Regents’ Scholarship (4 years tuition)

## Students

|                          |                                                                                                                                                                                                                                                                                                                                                                                                                       |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Ph.D. Advising</i>    | Samantha Molnar, University of Colorado, Aug 2015 - Dec 2021,<br>Inertia, Network Structure, and Power System Oscillations<br>Outcome: Hired as NREL Staff Scientist                                                                                                                                                                                                                                                  |
| <i>Postdoc Advising</i>  | Nicholas Brunhart-Lupo, NREL, Nov 2014 - Oct 2016<br>Outcome: Hired as NREL Staff Scientist<br>Roba Binyahib, NREL, Apr 2020 - July 2021<br>Outcome: Hired as Intel Graphics Engineer                                                                                                                                                                                                                                 |
| <i>Ph.D. Committee</i>   | Sandra Bae, University of Colorado, March 2025,<br>Advancing Interactive Physicalizations<br>Matthew Whitlock, University of Colorado, November 2021,<br>Immersive Augmented Reality for Data-Driven Workflows<br>Nicholas Brunhart-Lupo, Colorado School of Mines, Oct 2014,<br>Morse decompositions of three-dimensional piecewise constant vector fields                                                           |
| <i>External Examiner</i> | Cian Kelly, University College of Dublin, June 2025                                                                                                                                                                                                                                                                                                                                                                   |
| <i>Interns</i>           | Sandra Bae, CU-Boulder (APUP), 2024-2025<br>Isaiah Lyons-Galante, CU-Boulder (APUP), 2022-2024<br>Sunand Raghupathi, Cornell (SULI), 2019<br>Matthew Whitlock, CU-Boulder (RPP), 2019<br>Samantha Molnar, CU-Boulder (APUP), 2016-2020<br>Sean Yang, University of Washington (RPP), 2017<br>Shane Witter Hicks, Principia College (SULI), 2016<br>Nicholas Brunhart-Lupo, Colorado School of Mines (APUP), 2012-2014 |

## Expertise

|                  |                                                                                                                                                                                                                                                               |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Languages</i> | C++, R, Python, Lisp, Fortran, IDL, JavaScript, OpenGL Shading Language, CUDA                                                                                                                                                                                 |
| <i>Libraries</i> | OpenGL, Qt, VTK, ITK, OpenInventor, NetCDF, MPI, d3, X-Windows                                                                                                                                                                                                |
| <i>Software</i>  | VAPOR, ParaView, Avizo, JMP, $\LaTeX$ , Unity, Godot                                                                                                                                                                                                          |
| <i>Areas</i>     | Scientific visualization, immersive visualization, human-computer interaction, high-performance scientific computing, object-oriented development, numerical algorithm development, GUI, real-time programming, interactive computer graphics, and simulation |

## Professional Affiliations

|                            |                                                           |
|----------------------------|-----------------------------------------------------------|
| <i>Senior Member</i>       | IEEE, IEEE Computer Society                               |
| <i>Member</i>              | IEEE Visualization and Graphics Technical Committee       |
| <i>Professional Member</i> | Association for Computing Machinery (ACM)                 |
| <i>Professional Member</i> | ACM SIGGRAPH, ACM SIGCHI                                  |
| <i>Member</i>              | SIAM                                                      |
| <i>Member</i>              | Sigma Xi                                                  |
| <i>Professional Member</i> | American Institute of Aeronautics and Astronautics (AIAA) |

## Service

|                           |                                                                                         |
|---------------------------|-----------------------------------------------------------------------------------------|
| <i>Board Member</i>       | Computer Science & Engineering Advisory Board, New Mexico Tech, 2023-present            |
| <i>Steering Committee</i> | DOECGF 2016 - present                                                                   |
| <i>Chair</i>              | EnergyVis 2024, St. Pete Beach, FL                                                      |
| <i>Local Chair</i>        | High Performance Graphics 2024 (HPG24), Denver, CO                                      |
| <i>Chair</i>              | EnergyVis 2023, Melbourne, Australia                                                    |
| <i>Co-Chair</i>           | Data Challenges in Sustainable Urban Energy, October 2023, Monash University, Australia |
| <i>Chair</i>              | Visualization Paradigms in Renewable Energy, <i>IEEE VIS 2019</i> , Vancouver, Canada   |
| <i>Co-chair</i>           | EnergyVis 2021, Torino, Italy                                                           |
| <i>Program Chair</i>      | DOECGF 2018, Savannah, GA                                                               |
| <i>Site Chair</i>         | DOECGF 2017, Golden, CO                                                                 |
| <i>Program Committee</i>  | SimAUD 2017 - 2021                                                                      |
| <i>Program Committee</i>  | CityVis 2020                                                                            |
| <i>Program Committee</i>  | Intelligent Data Analysis 2011-2016                                                     |
| <i>Mentor</i>             | IEEE VR Doctoral Consortium, 2021                                                       |



|                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Coach</i>          | Maple Grove Math Olympiads Team, 2018-2019, Golden, CO                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <i>Judge</i>          | CSU Virtual Reality Hackathon, 2016, Fort Collins, CO                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <i>Instructor</i>     | Maple Grove Minecraft Club, 2016-2017, Golden, CO                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <i>Judge</i>          | Korea Institute of Science and Technology Information Visualization Competition, 2014                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <i>Research Diver</i> | Pacific Whale Foundation, Summer 1998, Wailuku HI                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <i>Reviewing</i>      | IEEE TVCG, IEEE Visualization, Lдав, SIGCHI, IEEE VR, IEEE Access, PRESENCE, EuroVis, PacificVis, IDA, SimAUD, Journal of Information Science and Engineering R Journal, DOE ASCR                                                                                                                                                                                                                                                                                                                                       |
| <i>DOE Workshops</i>  | Autonomous Research for Real-World Science (ARROWS), 2025<br>NITRD/DOE/NSF DECIDE Workshop: Computational Decision Support for High-Consequence, Complex Problems, 2025<br>ASCR Workshop on Visualization for Scientific Discovery, Decision-Making, and Communication, 2022<br>Computational Challenges in Energy Systems Integration and Grid Modernization, 2015<br>DOE Exascale Workshop on Data Analysis, Management, and Visualization, 2011<br>NREL Workshop on Scientific Data Management and Informatics, 2009 |
| <i>Graduate</i>       | <i>Graduate Representative</i> , 2005-2006 Faculty Search Committee                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <i>Undergraduate</i>  | <i>Member</i> , Solar Car Team, 1993-1994<br><i>Senator</i> , Student Senate, 1991-1992<br><i>Member</i> , Student Judiciary Board, 1990-1991                                                                                                                                                                                                                                                                                                                                                                           |

## Open-Source Development

---

**FLC Solar Lab**, <https://github.com/FLC-Solar-Lab>

*Role:* Principal Investigator, Design, Development  
Hardware and software infrastructure for integrating digital twin simulations with physical objects in a laboratory setting.

**NOODLES**, <https://github.com/InsightCenterNoodles>

*Role:* Design  
NOODLES is a protocol specification for scenegraph synchronization, based off GLTF, using CDDL for the specification and CBOR as the message format.

**3D Power Flow Visualization**, <https://github.com/nbrunhar/Intellicamp3>

*Role:* Design / Developer  
An immersive software suite to visualize power flow simulations.

**CatCost (Catalyst Cost Estimation Tool)**, [catcost.chemcatbio.org](http://catcost.chemcatbio.org)

*Role:* Design  
Cost estimation tool designed to reduce the cost uncertainty associated with pre-commercial catalyst materials development.

**Surface Phase Explorer**, [spe.nrel.gov](http://spe.nrel.gov)

*Role:* WebGL Development  
Understanding the coadsorption of two species on a surface requires the exploration of phase diagrams. To facilitate exploration this web tool creates interactive and downloadable phase diagrams from ab initio data.

**kaleidoscope**, [github.com/NREL/kaleidoscope](https://github.com/NREL/kaleidoscope)

*Role:* Sole Author  
An R package developed to visualize PLEXOS scenarios.

**VAPOR**, [www.vapor.ucar.edu](http://www.vapor.ucar.edu)

*Role:* Developer  
*Contributions:* Volume rendering engines, transfer function interface, model parsing and rendering  
An interactive 3D visualization and quantitative analysis software suite tailored towards terascale computational fluid dynamics data.

**MarsFlight**, [education.grc.nasa.gov/MarsFlight](http://education.grc.nasa.gov/MarsFlight)

*Role:* Principal Engineer  
*Contributions:* Terrain rendering, subsystems interface, map interface, and front-end configuration & deployment interface  
An interactive flight simulator of a Mars airplane concept vehicle, which includes a complete model of the Martian terrain based on MOLA data and rover imagery. The flight simulator is based on the open source FlightGear ([www.flightgear.org](http://www.flightgear.org)) project.

**iPyMOL**, contact [dubin@colorado.edu](mailto:dubin@colorado.edu)

**Role:** Principal Engineer

**Contributions:** Immersive port

An immersive port of the PyMOL ([pymol.sourceforge.net](http://pymol.sourceforge.net)) molecular visualization system, adding interactive visualization support for head-tracked, stereoscopic immersive virtual environments.

## Other

---

*Certifications*

Wilderness EMT / EMT-B

Open Water SCUBA Diver

*Hobbies*

Traditional rock climbing, mogul skiing, additive manufacturing, wood working

*Development Portfolio*

(available upon request)

*Security Clearances*

(available upon request)

*References*

(available upon request)

## References

- [1] D. A. Bowman and R. P. McMahan. Virtual reality: How much immersion is enough? *IEEE Computer*, 40(7):36–43, 2007.
- [2] Federal Energy Regulatory Commission and North American Electric Reliability Corporation. Arizona–southern california outages on september 8, 2011: Causes and recommendations. Technical report, FERC & NERC, 2012. Accessed 2025-10-02.
- [3] A. S. Forsberg, D. H. Laidlaw, A. van Dam, R. M. Kirby, G. E. Karniadakis, and J. L. Elion. Immersive virtual reality for visualizing flow through an artery. In *Proceedings IEEE Visualization 2000*, pages 457–460, 2000.
- [4] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, Amsterdam, 1988.
- [5] M. K. Hoang, S. Goodwin, M. Wybrow, and Y. Yang. Exploring the benefits of geography on power network diagrams. In *Proceedings of the 2023 Workshop on Energy Data Visualization (EnergyVis)*, 2023.
- [6] A. Jäger, S. Mittelstädt, D. Oelke, S. Sander, A. Platz, G. Bouwman, and D. A. Keim. Lessons on combining topology and geography—visual analytics for electrical outage management. In *EuroVA 2016: EuroVis Workshop on Visual Analytics*, 2016.
- [7] B. Laha, K. Sensharma, J. D. Schiffbauer, and D. A. Bowman. Effects of immersion on visual analysis of volume data. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):597–606, 2012.
- [8] S. Molnar, K. Gruchalla, and S. D. J. Tan. Visualization of the oscillatory dynamics of an island power system. In *Proceedings of the 2023 Workshop on Energy Data Visualization (EnergyVis)*, 2023.
- [9] M. Steiger, T. May, J. Davey, and J. Kohlhammer. Smart grid monitoring through visual analysis. In *2013 4th IEEE/PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, pages 1–5, 2013.
- [10] U.S.–Canada Power System Outage Task Force. Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations. Technical report, U.S. Department of Energy and Natural Resources Canada, 2004. Accessed 2025-10-02.
- [11] P. C. Wong, K. Schneider, et al. A novel visualization technique for electric power grid analytics. *IEEE Transactions on Visualization and Computer Graphics*, 15(3):410–423, 2009.

Dear Olivia, Ryan, and Shravya,

Thank you for your feedback! The consistency of observations made it clear that my proposal would benefit from improving my visualization specificity and survey explanation. I also noticed a few comments about the novelty and feasibility of my project. I've grouped these comments into four categories and have provided my plan to reconcile your suggestions (in the blue boxes).

## 1) Tech Specs

*"Think about more details on the actual visualization component."*

*"Needs more details beyond concept, such as a possible example layout or mock-up prototype. Who are you aiming to prioritize for the audience (experts or broad)?"*

*"Visualization design lacks concrete examples or prototypes."*

*"Could maybe more explicitly define how the "interactive" component will be implemented (web-based? Python notebook? JavaScript app?)"*

*"The ideas are strong conceptually but would benefit from tighter focus and clearer implementation details"*

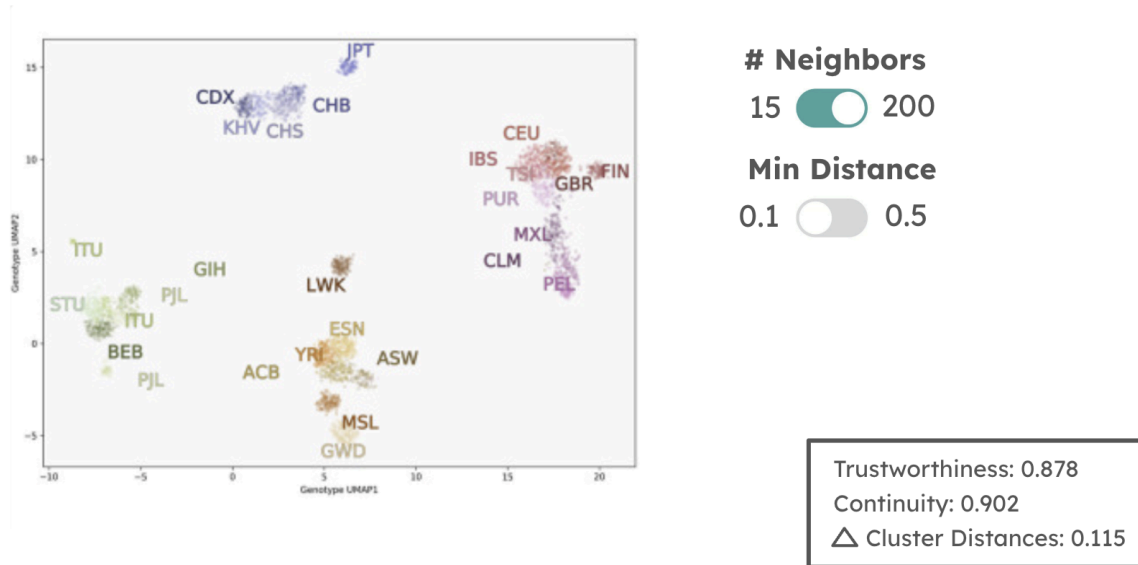
*"The idea of using toggles and overlays to visualize hyperparameter effects is promising, but the plan leans more conceptual than technical. Details on how interactivity will be achieved (e.g., libraries, UI structure, user interaction flow) would strengthen this section."*

To make my visualization plan more concrete, I propose using the following tools to generate an interactive UMAP model:

- I will design a static webpage using Typescript, which will host all of the information and interactions available to the user
- I will generate each UMAP permutation using the Python-learn library\*
- I will generate interactive parameter toggles, UMAP feature tooltips, and metric keys using React components

\*Since it takes awhile (several minutes) to render large UMAP models, it would be challenging to re-render each permutation in real time. I plan on rendering each permutation as a static image, which can then be presented to the user depending on which parameter settings they select.

Here is the mock-up I generated for my presentation:



## 2) Survey Explanation

“Success will depend on survey recruitment and execution.”

“Evaluation plan could include more detail on survey design, like the participant numbers and overall analysis.”

“This needs to be user-tested, but: slight risk of oversimplification when communicating complex UMAP behaviors to non-experts”

“Evaluation strategy (survey) is vague and the criteria for success and participant demographics are not clearly stated.”

To clarify my survey, I’ve created a more structured plan:

Survey 1 (Non-Domain-Specific): This survey will be provided to 20-30 students who are not domain-specific experts (they do not have a strong background in population genetics). Half of the participants will be presented with a static UMAP plot. The other half will be presented with an interactive UMAP plot generated from the same underlying high-dimensional data. Both groups will have 2 minutes to explore the features of the plots before they may begin answering questions (though they will have access to the plots throughout the entire duration).

Quiz: There will be 10-15 multiple choice questions that will test the students' understanding of key UMAP concepts as well as specific inferences related to the data on their plots. The questions will have objectively correct answers. Here is an example:

*What is the distance between the CEU cluster and the MXL cluster?*

- A) 10 U (UMAP units)
- B) 10 BP (basepairs)
- C) Distance is proportional to the distance between sampling locations
- D) Distance can only be expressed relative to other distances on the plot
- E) I am not sure

Qualitative analysis: Following the quiz, the students will be presented opportunities to express opinions about their experience. Here is an example:

*Were the clusters more, less, or equally related as you would have expected? Please explain.*

Survey 2 (Domain-Specific): This survey will be presented to 5 experts in the domain of population genetics (or similar fields). These questions will be more centered around the utility of the toggle functionality and less focused on inference accuracy. Here is an example:

*How do you typically decide which parameter values to use in your dimensionality reduction plots? How would this toggle functionality impact this exploration process if you were to integrate this into your research plan?*

### 3) Novelty

“It primarily repackages known ideas in an interactive form rather than introducing fundamentally new techniques.”

“It lacks depth on how the experiments will be structured or how the results will be validated quantitatively.”

“Similar ideas exist in explainable AI and dimensionality reduction research. The novelty lies more in its application context (population genetics) and user-centered design than in algorithmic innovation.”

I would like to defend the novelty of my project by breaking it down into 2 sections:

Biological novelty: I would like to extend current parameter optimization strategies by studying the impact of parameter adjustments on local and global distance preservation. These metrics will provide a comprehensive, qualitative, and novel understanding of how well a low-dimensional embedding preserves relationships from the high dimensional data. By

emphasizing these distance preservations, diversity will be prioritized over group separation in the final plot (as opposed to focusing on formulating distinct clusters). I will emphasize the difference between diversity preservation and categorical accuracy in my final proposal to ensure the novelty of my plan is clear to the reader.

Visualization novelty: Perhaps I could rephrase my goal as “educational clustering,” rather than “explainable clustering” (as not to be conflated with explainable AI). Methods for dimensionality reduction strategies are well-understood by researchers. Conveying this intuition quickly to a less experienced audience is where the novelty of my project exists. There is also novelty in how this project can serve experts. In my literature review, I found that “many studies mention future or ongoing efforts to improve parameter selection” [Bae et. al.]. There is a clear need for novel parameter visualization strategies — especially as a toolkit for biologists. I will emphasize this need in my final proposal and during my project.

#### **4) Feasibility**

“Completing both the computational experiments and user study in six weeks seems optimistic.”

To improve the feasibility of my timeline, I will consider developing the survey in tandem with the visualization tool. By designing these two products at the same time, I can ensure that they are tailored to one another while also working more efficiently. This way, as soon as I am complete with the visualization tool, I can release the survey (rather than dedicating an entire week to survey design).

# Explorative Clustering: An Interactive UMAP Model in the Domain of Population Genetics

Kyle Wisialowski  
PI

`kyle.wisialowski@brown.edu`

Skylar Walters  
Co-PI

`skylar.walters@brown.edu`

Dr. Alex Diaz-Papkovich  
Collaborator  
`alex.diaz-papkovich@brown.edu`

October 14, 2025

## Abstract

We present a novel, interactive method of visualizing UMAP plots that emphasizes the volatility of the plots with respect to their hyperparameters. This method overcomes a limitation of the traditional UMAP model, which does not capture this volatility, nor does it properly attempt to explain general UMAP features — particularly for lay users. We report findings from a survey which shows that an interactive model improves expert performance in determining optimal hyperparameter settings and non-expert performance in drawing accurate inferences.



# 1 Aims

**1) Determine how key UMAP parameters preserve local and global distances.** Our central goal is to explore and expose the underlying mechanisms that give rise to clusters on UMAP plots. These mechanisms are often briefly represented in UMAP captions or in the methods section of scientific papers. A comprehensive literature survey has shown that these explanations are not sufficient in conveying the abstract features of UMAP models or how these features can be used to draw inferences. [3] Our goal is to make these mechanisms more intuitive and interpretable by proposing a method for visually highlighting how key parameters influence how clusters form. The parameters we will investigate include the number of neighbors used to draw inferences, the minimum distance between clusters, the dimensionality of the embedding (2D versus 3D), and the distance metric used to generate the plot. The dependent variables of this experiment include a local distance metric (to measure how well the low-dimensional embedding maintains the distances present between individual samples in the high-dimensional data) and a global distance measurement (to measure how well the low-dimensional embedding maintains the distances present between distinct clusters in the high-dimensional data), as explained in Figure 1.

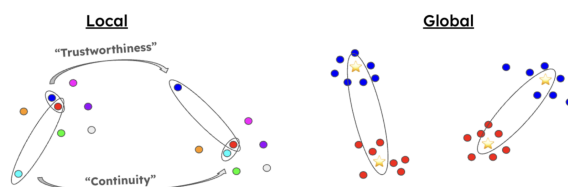


Figure 1: A mock-up for the proposed interactive UMAP mode. A “standard” UMAP model will be presented to the user with adjacent toggles to view the model with alternative parameter settings. There will be a key of local distance metrics (trustworthiness and continuity) and a global distance metric (cluster distances) corresponding to the current parameter setting. The tooltip in this image displays the intuition behind the coloring for the selected group of clusters.

**2) Render an interactive UMAP model that visualizes hyperparameters, as shown in Figure 2.** The model will include 16 example UMAP plots (only one plot will be viewable at a time) that iterate across 2 permutations of the following variables: `n_neighbors` (number of neighbors), `min_dist` (minimum distance between clusters), `n_dim` (number of dimensions), and `dist_metric` (distance metric). Each hyperparameter will have a toggle between two distinct values that will allow the user to understand how the parameter impacts the results, as well as an informational overlays providing an intuitive explanation of the plot features such as color gradients, axes, parameters, and distance preservation metrics.

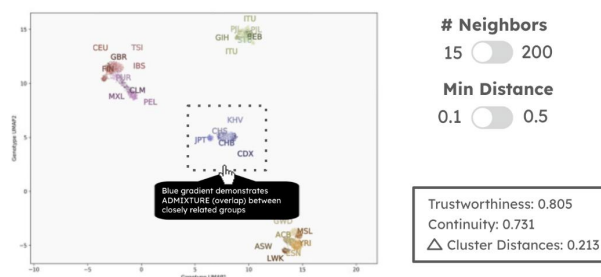


Figure 2: A mock-up for the proposed interactive UMAP mode. A “standard” UMAP model will be presented to the user with adjacent toggles to view the model with alternative parameter settings. There will be a key of local distance metrics (trustworthiness and continuity) and a global distance metric (cluster distances) corresponding to the current parameter setting. The tooltip in this image displays the intuition behind the coloring for the selected group of clusters.

**3) Compare results between novel UMAP renderings and traditional UMAP renderings with a survey of experts and non-experts.** This non-expert survey will be provided to 20-30 students who are not domain-specific experts (they do not have a strong background in population genetics). Half of the participants will be presented with a static UMAP plot. The other half will be presented with a novel interactive UMAP plot generated from the same underlying high-dimensional data. Both groups will have 2 minutes to explore the features of the plots before they may begin answering questions (both groups will have the same questions), as shown in Figure 3. The expert survey will be presented to 3-5 researchers in the domain of population genetics (or similar fields). These questions will be more centered around the utility of the toggle functionality and less focused on inference accuracy, with questions like: How do you typically decide which parameter values to use in your dimensionality reduction? How would this toggle functionality impact this exploration process if you were to integrate this into your research plan?

The distance between the CEU cluster and the MXL cluster?

A 10 U (UMAP units)

B 10 BP (basepairs)

C Distance is proportional to the distance between sampling locations

D Distance can only be expressed relative to other distances on the plot

E I'm not sure

Were the clusters more, less, or equally related as you would have expected? Please explain.

Figure 3: The left image shows an example of how a student’s comprehension of the UMAP rendering will be measured through objective quiz-like questions. The right image shows an example of how the survey will include a qualitative analysis.

## 2 Significance

**Biological significance.** The hyperparameters that we will investigate ( $n\_neighbors$ ,  $min\_dist$ ,  $n\_dim$ , and  $dist\_metric$ ) are common “knobs” in UMAP plot design that scientists tune to best fit the data they are trying to present. Previous studies have explored the effects of these parameters, but these studies only emphasize the overall clustering accuracy (i.e. if samples are grouped into the correct clusters) rather than how the parameters affect the preservation of local and global distances. [2] By extending the analysis of hyperparameters beyond the clustering accuracy, we can better understand how to select hyperparameter values that maintain the most accurate preservation of local and global distances. It will also be interesting to optimize these parameters on the distance metrics, then visualize the associated plot to determine how these mathematical optimizations correlate with visual representations.

**Visualization significance.** The biological inferences drawn from the above experiments, combined with survey results, will inform a method of determining how to visually display the influence of hyperparameter modifications on UMAP clustering results. The visualization significance can be divided into 2 categories: explorative (for experts) and educational (for non-experts). The explorative significance is that our proposed project will provide an opportunity for experts in the field to easily toggle between hyperparameter values to determine which set of values produce the most optimal plot for their research purposes. The educational significance is that, in addition to the hyperparameter exploration, our project will present informative overlays which will help develop and reinforce non-experts’ understanding of UMAP features. If integrated into scientific literature, this interactive visualization may increase the clarity of plots and the rate at which readers (expert and non-expert) can draw meaningful and accurate inferences.

**Sociopolitical significance.** The central reason why the non-expert use-case is included in this proposal is to decrease the likelihood of misinterpretation or manipulation of analyses pertaining to population genetics. There have been historical instances where dimensionality-reduction plots have been annotated with misleading labels to corroborate white-supremacist and eugenic agendas. By making more nuanced, interactive UMAP visualizations, drawing these malicious conclusions will be more challenging. The plots that have led to these misrepresentations are static and contain rigid boundaries. By providing an interactive display that accurately conveys the volatility and triviality of cluster distances, it will be much more challenging to extract these extremist and reductionistic conclusions, as shown in Figure 4.

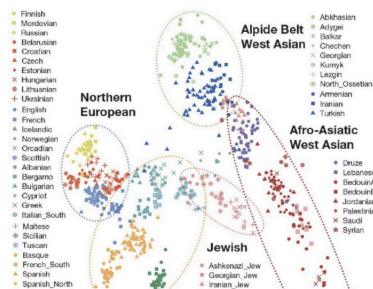


Figure 4: Example of a UMAP plot with rigid boundaries and definitive labels. This is a static image that does not present information in an appropriately nuanced manner. There is little intuition about how the clusters are grouped together, nor any gradient representing genetic overlap between clusters.

### 3 Related Work

The emphasis on visualizing the effects of UMAP hyperparameters and providing educational tooltips is motivated by a survey conducted by Bae et. al. After surveying over 100 papers related to human-visualization interaction in the context of clustering plots, they determined that explainability and confidence were among the most influential qualities associated with the user’s overall understanding and attitude towards the figures. [3] Since this is a survey paper, it is limited in that the authors did not include any practical applications of their results. Our proposed project aims to verify the results of this survey by focusing on an explainable visualization that also conveys the volatility of results — ultimately providing the end-user with a better overall understanding of the data as well as an accurate metric (local and global distance preservation) that can inform the confidence in which they can draw inferences.

Other methods of visualization were considered because there is sparse literature on interactive UMAP visualizations — specifically in the context of population genetics. Palma-Martinez et. al. propose an entirely different method of visualization in which genetic differences are displayed as SNP (single-nucleotide polymorphism) plots grouped by sampling region. [4] This method was considered because it emphasizes cross-group similarity (different regions may have similar SNP ratios) and in-group variation (a geographic population may have a diverse range of SNPs), which would be helpful in preserving sample distances between high dimensional data and low dimensional visualizations. While these features are useful, reducing the metric of genetic differentiation to SNP ratios makes it difficult to draw more complex conclusions. Additionally, the overlay of discrete pie-charts above geographic regions makes it difficult to visualize how different populations are related to each other; on these maps, distance is purely geographically-rooted and there is no gradient between populations.

One important consideration in this project is who the end-user will be. The most practical end-user is domain-specific scientists who will use UMAP visualizations to draw conclusions about population dynamics. Das et. al. explore this use-case in a study that prepares an intuitive toolkit for biologists to interactively study clustering data [5]. The benefit of this approach is that by building dynamic tools, the computer scientists can offload the biological exploration onto the biologists. We will incorporate some of this approach in our work by developing interactive hyperparameter modification. As mentioned, the focus for the experts is to provide a tool for them to determine which hyperparameters best fit their research interests. We will extend this work by considering the possibility of lay-users as well. This use case requires attention to educational features (like information overlays) and careful labeling strategies to avoid misuse.

One final study that provides motivation for this project is from Lawson et. al., who developed a tutorial on how to avoid over-interpreting Structure and AdMixture plots. [6] This study is instructional in teaching scientists how to critically examine dimensionality reduction plots. Their central concern is that highly similar low-dimensional plots can be rendered from highly deviant high-dimensional data. Rather than offloading this interpretive burden onto scientists, our proposal seeks to make these misconceptions easily understandable by the user by visualizing the volatility of UMAP plots with regard to the modification of hyperparameters.

## **4 Research Plan**

### **Weeks 1-2**

We will generate a summary of results from an experiment that permutes over an array of hyperparameter values. The summary will include how these hyperparameters influence local cluster distances (trustworthiness and continuity), and global cluster distances (centroid euclidean distances). To run this experiment, we will use the Python-learn library to generate UMAP plots and the biopython library to evaluate distance metrics.

### **Week 3**

Using Typescript and React, we will produce a basic UMAP display that dynamically updates depending on which parameter settings the user has selected. Since UMAP renderings take a while to load, all of the preprocessing will be complete before the interactive visualizations gets rendered; the toggles simply switch between a series of static images. We will also curate a set of questions for experts to receive preliminary feedback.

### **Week 4**

We will collect survey results from experts and modify visualizations to best capture the expert recommendations. We will also create a non-expert survey with quiz-like questions and qualitative analyses, and send out this survey to a group of 20-30 students.

### **Week 5**

We will synthesize the results from the non-expert survey and modify the visualizations to account for any feedback. We will then review any visualization updates with the expert population from the Week 3 survey to ensure project updates are compatible with both the expert and non-expert users.

### **Week 6**

We will summarize the results in a final paper.

## References

- [1] Liu, Z., et al. “Assessing and Improving Reliability of Neighbor Embedding.” PMC, 2025. PMC12125374.
- [2] Xiang, R., Wang, W., Yang, L., Wang, S., Xu, C., Chen, X. (2021). A Comparison for Dimensionality Reduction Methods of Single-Cell RNA-seq Data. *Frontiers in Genetics*, 12, 646936.
- [3] Bae, J., Helldin, T., Riveiro, M., Nowaczyk, S., Bouguelia, M. R., Falkman, G. (2020). Interactive clustering: A comprehensive review. *ACM Computing Surveys (CSUR)*, 53(1), 1–39.
- [4] Palma-Martínez, M. J., Posadas-García, Y. S., López-Ángeles, B. E., Quiroz-López, C., Lewis, A. C. F., Bird, K. A., Lasisi, T., Zaidi, A. A., Sohail, M. (2024). The multi-scale complexity of human genetic variation beyond continental groups. *bioRxiv*, arXiv:2024.12.11.627824.
- [5] Das, S., Saket, B., Kwon, B. C., Endert, A. (2020). Geono-cluster: Interactive visual cluster analysis for biologists. *IEEE Transactions on Visualization and Computer Graphics*.
- [6] Lawson, D. J., van Dorp, L., Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1), 3258.

# Kyle Wisialowski

☎ (860) 800-4466 | ✉ [kyle\\_wisialowski@brown.edu](mailto:kyle_wisialowski@brown.edu) | 🌐 <https://github.com/kylewisialowski> | 💼 <https://www.linkedin.com/in/kyle-wisialowski-0551121a9/>

## EDUCATION

- Brown University, Sc.M. Computer Science (AI/ML Track)** Providence, RI | **Expected Graduation May 2026**
- Current Coursework: Deep Learning, Human-AI Interaction, Interdisciplinary Scientific Visualization
- Brown University, Sc.B. Applied Math - Biology, 3.88/4.0 GPA** Providence, RI | Class of 2025
- Relevant Courses: Software Engineering, Foundations of Artificial Intelligence, Program Design: Data Structures and Algorithms, Statistical Inference I, Ordinary Differential Equations, Multivariable Calculus, Linear Algebra, Inference in Genomics
  - Sigma Xi International Scientific Research Honor Society
- Old Saybrook High School, 103.8/100 GPA** Old Saybrook, CT | Class of 2021
- Salutatorian, National Merit Commended Scholar, Bausch + Lomb Science Award, Varsity Cross Country, Swim, & Crew

## PROFESSIONAL EXPERIENCE

- Brown University Department of Computer Science, Teaching Assistant** Providence, RI | September 2025 – Present
- Host weekly office hours and grade assignments for students in CSCI 0410 (Foundations of AI)
  - Help students understand and debug implementations for common AI algorithms
- Center for Computation and Visualization, Software Development Intern** Providence, RI | June 2025 – August 2025
- Developed a parsing algorithm using ANTLR to extract parameter information from commonly used biological workflows
  - Integrated a self-hosted Ollama Large Language Model into the above parsing algorithm using Unsloth and Hugging Face to autonomously annotate workflow parameters
  - Trained a Llama 3 LLM to understand prompting with a niche biological workflow language (Workflow Development Language)
  - Built a Docker Container to host the ANTLR + LLM software, deployed through GitHub Actions
  - Restructured the above Docker workflow into a script compatible with Brown's High Performance Computing cluster
  - Automated cross-referencing between 2 genomic databases with dynamically-updated React dropdown components
- Center for Computation and Visualization, Research Assistant** Providence, RI | June 2024 – May 2025
- Student researcher under Dr. Katherine Siddle's SARS-CoV-2 genome study with Brown's Computational Biology Core
  - Hosted and edited Nextstrain phylogeny software on Brown's High Performance Computing cluster to produce phylogenetic analyses for a dataset of 9 million SARS-CoV-2 genomes
  - Merged a subsampling pipeline and a phylogenetic pipeline into a combined script with a Terminal User Interface — reducing script-running workload for Dr. Siddle and a doctoral student by 50%
  - Developed R-analyses for phylogenetic data (temporal distribution plots, frequency distribution plots, variance calculations)
  - Synthesized results in semesterly presentations for a group of researchers, doctoral students, and undergraduates
- Brown University Department of Biology, Teaching Assistant** Providence, RI | January 2024 – May 2024
- Graded weekly lab write-ups and provided feedback for students in BIOL 0285 (Inquiry in Biochemistry)
  - Assisted students during lab exercises, including gel electrophoresis, PCR reactions, protein induction, and kinetics assays
  - Helped students produce Michaelis-Menten and Lineweaver-Burke plots to interpret their results
- Pods Swimming, Head Instructor** Providence, RI | March 2023 – August 2024
- Taught weekly lessons in water safety and basic swimming techniques for kids aged 3-9, teens, and adults
  - Cultivated a fun learning environment where students mastered lifesaving skills like floating/kicking comfortably on their back
- College Matters, Summer Intern** Madison, CT | June 2022 – August 2022
- Used Excel, Unito, Trello, and Zapier to automate the client onboarding process for a college application consulting firm
  - Re-developed an existing protocol to efficiently store and organize consultant notes from college visits

## EXTRACURRICULAR/VOLUNTEER EXPERIENCE

- Club Swim Team, Co-Captain** Providence, RI | September 2021 – Present
- Help lead a team of 11 board members and 150+ swimmers
  - Manage team finances, travel logistics, practice schedules, and social events
  - Attended CCS National Championship in 2024/2025 and team record holder in 100 fly, 200 fly, 100 IM, and 200 IM
- Hermes Free Telehealth, Medical Outreach Team Member** Virtual | June 2022 – May 2024
- Assisted with outreach for a student-led telehealth nonprofit working to increase access to healthcare for underserved populations
  - Attended weekly meetings to brainstorm methods for recruiting volunteer health professions
- Applied Math Directed Reading Program, Participant** Providence, RI | September 2023 – December 2023
- Selected to participate in Brown's independent reading program lead by the Division of Applied Mathematics
  - Studied mathematical models of disease under the guidance of Dr. Juniper Cocomello
  - Worked with another student to develop a presentation about the theory, application, and modification of the SIR model

## SKILLS & INTERESTS

**Technical Skills:** Proficient in Python, Java, Linux/Unix environments, Javascript/Typescript, React, Matlab, R, data visualization

**Language:** High School Seal of Biliteracy in Spanish (elementary proficiency)

**Interests:** Rollercoasters, trying new food, pickleball, skiing, sunsets, anything involving a body of water

# Skylar Walters

69 Brown St, Box 3979 | Providence, RI 02912 | Phone: (973) 610-8724 | E-Mail: skylar\_walters@brown.edu

## EDUCATION

**Brown University, B.S. Computational Biology, Computer Science Track** Providence, RI | **Grad. May 2026**

- **Coursework:** Deep Learning, Computational Biology, Statistics and Inference in Genomics, Genetics, Data Structures & Algorithms, Linear Algebra, Statistical Inference I, Biotechnology in Medicine, Organic Chemistry

## EMPLOYMENT

**Humans in Data Integration, Visualization, and Exploration (HIDIVE) Lab, Intern** Boston, MA | June-Oct. 2025

- Developing a system to collect and analyze Gosling visualization specifications for genomics data, enabling task-driven visualization design and recommendation.
- Leveraging a formal taxonomy of genomics data, tasks, and visual encodings to support systematic analysis and potential machine learning applications.
- Associated with Harvard Department of Biomedical Informatics Summer Institute in Biomedical Informatics

**Surveillance of Infectious Disease Lab (SIDL), Research Assistant** Providence, RI | Aug. 2023-Present

- Develops computational techniques to detect defective viral genomes (DVGs) from large metagenomic datasets (1 billion+ data points) as a means of understanding mechanism by which DVGs trigger antiviral host responses
- Applies recombination mapping techniques to Ebola and SARS-CoV-2 datasets via shell scripts
- Granted \$2,500 Brown University Undergraduate and Teaching Research Award to continue research
- Present at weekly journal club meetings to gain proficiency in science communications and group collaboration

**Brown University, Deep Learning and Viral Metagenomics Researcher** Providence, RI | June-Aug. 2024

- Awarded NIH R25 grant to design a deep learning-based approach to identify novel and highly divergent viral genomes free of reference sequences, providing an alternative diagnostic tool for unidentified illnesses
- Adapts NLP techniques, variational autoencoders, and clustering algorithms to process 15,000+ viral and bacterial whole-genome sequences using Keras, h5py, TensorFlow, Scikit-Learn, and more
- Constructed latent-layer PCA models, weighting schema to reduce bias in data, and plots of loss and accuracy per epoch

**Brown University Department of Computer Science, Meta Teaching Assistant** Providence, RI | Oct. 2024-Present

- Collaborates with Director of Undergraduate CS Studies to organize the daily activities of the Computer Science Department
- Oversees hiring of over 300 TAs and Head TAs and organizes department-wide TA trainings
- Scripts algorithms to streamline hiring, evaluations, and room booking

**Brown University Department of Computer Science, TA & Head Teaching Assistant,** Providence, RI | Aug. 2023-Present

- Facilitates dynamic learning environment for 100 students in Computer Science: an Integrated Introduction
- Liaisons between students, Professor Philip Klein, and team of 11 Teaching Assistants to ensure course runs smoothly
- Leads development and launch of weekly homework assignments, while hosting office hours, labs, and grading sessions

## PEER-REVIEWED PUBLICATIONS

**Sargent Walters, S., Valderrama, A., Smits, T., Kouřil, D., Nguyen, H., L'Yi, S., Lange, D., Gehlenborg, N., "GQVis: A Dataset of Genomics Data Questions and Visualizations for Generative AI," In Proc, IEEE VIS, 2025**

## CAMPUS/COMMUNITY INVOLVEMENTS

**Appalachian String Ensemble, Banjo & Mandolin** Providence, RI | Sept. 2022-Present

- Studies traditional Appalachian folk music in small and large ensembles, including Brown's Old Time String Band
- Trained to learn and teach tunes by ear, continuing centuries-old tradition of Old Time Music

## AWARDS

|                                                                                                    |           |
|----------------------------------------------------------------------------------------------------|-----------|
| <b>Barry Goldwater Scholarship</b>                                                                 | Mar. 2025 |
| <b>Annual Biomedical Research Conference for Minoritized Students, Computational Biology Award</b> | Nov. 2024 |
| <b>Research Grant via NIH R25 Student Grant</b>                                                    | June 2024 |
| <b>Brown University Undergraduate Teaching and Research Award</b>                                  | Nov. 2023 |
| <b>United States Presidential Scholar</b>                                                          | May 2022  |
| <b>Coca-Cola Scholar</b>                                                                           | May 2022  |



# Dr. Alex Diaz-Papkovich

## WORK EXPERIENCE

Teaching assistant and workshop leader

2018 - 2023

McGill University, Montreal

I have assisted in teaching courses in population genetics, and foundational courses in my PhD program. I have also designed and taught workshops on statistics and statistical programming in R as well as dimensionality reduction and ODEs.

Instructor

2018

McGill University, Montreal

I was the instructor for EPIB613: Introduction to Statistical Software for McGill graduate students in epidemiology. The course covered how to implement statistical approaches using R.

Mathematical statistician

2010 - 2017

Statistics Canada, Ottawa

I was responsible for the development and implementation of statistical methods for several surveys and projects. My responsibilities included survey design, sample selection and allocation, weighting, estimation, statistical programming, record linkage, and writing and presenting technical reports to stakeholders as well as supervision duties. My projects included:

- The International Travel Survey
- The Canadian Survey of Economic Well-Being
- The Canadian Income Survey
- The Longitudinal Immigration Database

Technical consultant

2007 - 2010

The Co-operators, Guelph

I was responsible for developing internal applications for insurance agents and database maintenance for the company. Most of the work was done in Visual Basic, VBA, and SQL.

## ONLINE CONTENT

[Dimension reduction workshop](#) - An interactive introduction to dimension reduction of genetic and single-cell data for the McGill Initiative in Computation Medicine.

[Introduction to statistics in R](#) - An introductory workshop for students in biology to learn the theory and application of statistics in R. Covers hypothesis testing, ANOVA, and regression.

[Introduction to linear regression workshop](#) - An introductory workshop on theory and applications of linear regression with accompanying code.

[Dimension reduction of genotypes](#) - Repository of code used to study dimension reduction on genotype data and to accompany my PLOS Genetics publication.

[An interactive demonstration of dimension reduction on genotype data](#) - A jupyter notebook comparing PCA, t-SNE, and UMAP on 1000 Genomes Project data.

[Modelling population growth using ordinary differential equations](#) - An introduction to using ODEs to solve models of population dynamics.

[Death by Car](#) - A project designed to collect and analyze news stories of traffic violence against cyclists and pedestrians in Canada.

## PUBLICATIONS

- Topological stratification of continuous genetic variation in large biobanks  
*Alex Diaz-Papkovich, Shadi Zabad, Chief Ben-Eghan, Luke Anderson-Trocmé, Georgette Femerling, Vikram Nathan, Jenisha Patel, Simon Gravel*  
bioRxiv (2023)
- A review of UMAP in population genetics  
*Alex Diaz-Papkovich, Luke Anderson-Trocmé, Simon Gravel*  
Journal of Human Genetics (2020)
- UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts  
*Alex Diaz-Papkovich, Luke Anderson-Trocmé, Chief Ben-Eghan, Simon Gravel*  
PLOS Genetics (2019).
- On the Genes, Genealogies, and Geographies of Quebec  
*Luke Anderson-Trocmé, Dominic Nelson, Shadi Zabad, Alex Diaz-Papkovich, Ivan Kryukov, Nikolas Baya, Mathilde Touvier, Ben Jeffery, Christian Dina, Hélène Vézina, Jerome Kelleher, Simon Gravel*  
Science (2023).
- Conserved whole-brain spatiomolecular gradients shape adult brain functional organization  
*Jacob W Vogel, Aaron Alexander-Bloch, Konrad Wagstyl, Maxwell Bertolero, Ross Markello, Adam Pines, Valerie J Sydnor, Alex Diaz-Papkovich, Justine Hansen, Alan C Evans, Boris Bernhardt, Bratislav Misic, Theodore Satterthwaite, Jakob Seidlitz*  
bioRxiv (2022).
- Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits  
*Melissa L Spear, Alex Diaz-Papkovich, Elad Ziv, Joseph M Yracheta, Simon Gravel, Dara G Torgerson, Ryan D Hernandez*  
eLife (2020).
- Don't ignore genetic data from minority populations  
*Chief Ben-Eghan, Rosie Sun, Jose Sergio Hleap, Alex Diaz-Papkovich, Hans Markus Munter, Audrey V. Grant, Charles Dupras, Simon Gravel*  
Nature (2020)

Link to the manuscript reviewed pdf:

[https://drive.google.com/file/d/1C7T\\_PdHYMHvSSMVW\\_lx5lrcJwcwP6hoh/view?usp=drive\\_link](https://drive.google.com/file/d/1C7T_PdHYMHvSSMVW_lx5lrcJwcwP6hoh/view?usp=drive_link)

## Response Letter

Dear Editor and Reviewers,

We thank you for the thoughtful and encouraging feedback on our proposal, “Visualizing Wind-Driven Vorticity in Wildfire Simulations.” We appreciate that both reviewers recognized the project’s clarity, interdisciplinary relevance, and feasibility. Below, we respond to each comment and describe how the proposal has been strengthened accordingly.

[R1]: Ryan Huang

[\[https://drive.google.com/file/d/1yTOgzbLIU\\_3EEVj9TK0EM2jkedrseldX/view?usp=drive\\_link\]](https://drive.google.com/file/d/1yTOgzbLIU_3EEVj9TK0EM2jkedrseldX/view?usp=drive_link)

[R1.1] Could provide more detail on visual design choices.

Response:

We have expanded the Research Plan section to provide a more detailed rationale for the visualization design choices. Specifically, we describe how streamline seeding strategies, contour thresholds, and glyph density will be varied to highlight different vorticity structures and temporal dynamics. These additions clarify the intended design space exploration and align with Lam et al.’s Evaluating User Performance (UP) scenario, in which visual encodings are iteratively refined for clarity and interpretability rather than stylistic variation.

[R1.2] How “interpretability” will be assessed remains vague.

Response:

We have clarified that interpretability will be evaluated through expert assessment rather than case studies or general-user testing. Our collaborator, an expert in wildfire modeling at LANL, will review the visualization outputs in consultation with other domain specialists. Expert assessment will examine whether vortex visualizations successfully reveal airflow features relevant to known fire-spread mechanisms and whether the visualization aids domain understanding.

[R2]: Shravya Munugala

[\[https://drive.google.com/file/d/1GPI-eFMb3BlgAvtaSYO04c5zKjegaLzx/view?usp=drive\\_link\]](https://drive.google.com/file/d/1GPI-eFMb3BlgAvtaSYO04c5zKjegaLzx/view?usp=drive_link)

[R2.1] Limited discussion of evaluation methodology; “informal feedback” is vague.

Response:

We have replaced the “informal feedback” plan with a structured expert evaluation conducted by our collaborator and her network of wildfire researchers. This assessment will use targeted prompts to evaluate:

1. Interpretive accuracy (does the visualization correctly convey vorticity structures?), and
2. Analytic usefulness (does it reveal new or confirm known relationships between airflow and fire spread?).

[R2.2] Methods (streamlines, contours, glyphs) are standard—novelty lies mainly in application, not technique.

Response:

We have clarified in the Significance and Related Work sections that the innovation of this project is conceptual and domain-oriented, focusing on surfacing underexplored vorticity structures rather than proposing a new rendering algorithm.

[R2.3] Scope limited to exploratory visualizations rather than methodological innovation.

Response:

We have explicitly acknowledged in the Aims section that this project's contribution lies in developing an exploratory framework rather than algorithmic innovation. Our work contributes both to fire science, by exposing wind-driven vorticity patterns, and to visualization methodology, by showing how insight-based evaluation can be applied to ensemble simulation data.

[R2.4] Could benefit from having a better evaluation plan.

Response:

We have strengthened the evaluation plan to include two complementary phases:

- Direct qualitative assessment of interpretive clarity and scientific validity.
- Comparative review of vorticity-focused versus conventional fire-intensity visualizations.

This approach ensures that evaluation is contextually rigorous (via domain expertise) and feasible within the project timeline, aligning with Lam et al.'s guidance that expert review is an appropriate form of summative evaluation for specialized visualization tools.

[R3]: Sir Dylan Allotey

<https://drive.google.com/file/d/1Qx2Vy4AjYjvrf-7mVwinPLLZHYoaepgb/view?usp=sharing>

[R3.1] Could strengthen the quantitative validation beyond visual inspection.

Response:

We have addressed this by expanding the Evaluation to include quantitative comparison metrics in addition to expert assessment. Specifically, we will compute feature agreement between visually identified vortices and analytic vorticity magnitude fields, and measure spatiotemporal coverage of vortex structures across ensemble simulations. These metrics provide numerical evidence of interpretive accuracy and coverage, ensuring that the evaluation goes beyond qualitative visual inspection.

[R3.2] Insights could be better tied to quantitative improvements in interpretation or prediction.

Response:

We have clarified in the Significance section that quantitative insight will be assessed through measurable agreement between visualized features and model-computed fields, as well as through expert ratings of interpretive clarity on a scale. While predictive modeling lies beyond the project scope, this evaluation directly quantifies the visualization's contribution to model interpretability and validation.

[R3.3] Future work could explore real-time interactivity.

Response:

We agree that interactive exploration would meaningfully extend the framework. While not feasible within the six-week project scope, we have noted this as a direction for future development in the Research Plan. Specifically, we envision that real-time streamline manipulation could support dynamic analysis of evolving fire–wind interactions.

\*only 3 reviews were done on my proposal as of 10/13\*

## Summary of Revisions

- Expanded explanation of visual encoding choices and design rationale
- Reframed “informal feedback” as expert assessment and interpretive validation.
- Clarified the conceptual novelty of applying standard methods to an underexplored domain
- Strengthened evaluation plan with two-phase structure (qualitative insights + iterative refinement).

We thank all reviewers again for their constructive feedback. These revisions improve the methodological grounding of our proposal and situate it clearly within established frameworks for visualization evaluation.

Sincerely,  
Olivia Heng  
Brown University

# Visualizing Wind-Driven Vorticity in Wildfire Simulations

Olivia Heng

PI

`olivia_heng@brown.edu`

Divya Bansh

Collaborator

`dbanesh@lanl.gov`

Aditi Kannan

Co-PI

`aditi_kannan@brown.edu`

October 14, 2025

## Abstract

We propose a novel visualization framework WindVis for exploring wind-driven vorticity in wildfire simulations. While existing wildfire visualization emphasizes terrain, vegetation, and fire intensity, airflow structures remain underexplored despite their critical role in lateral fire spread. Using ParaView and Python/VTK, we will design methods to reveal vortex structures and streamline dynamics across ensemble simulations, producing visualizations that connect airflow features to fire behavior. By addressing this gap, this framework contributes to wildfire science and to visualization research on exploratory multifield analysis.

# 1 Aims

Our project aims to create exploratory visualizations that expose wind-driven vortices in wildfire simulations. Vorticity structures are implicated in rapid, lateral fire spread and “blow-up” events, yet current visualization pipelines rarely foreground these features. We will investigate how targeted visual encodings of wind-driven vorticity can improve interpretability in wildfire simulations. Specifically, we will:

1. Design and implement vortex-centered visual encodings using ParaView and Python/VTK, including streamlines, contours, and glyphs optimized for highlighting airflow structures.
2. Compare these vortex-focused visualizations to baseline fire-intensity renderings across ensemble simulation data.
3. Evaluate interpretive value through expert assessment, measuring clarity, feature detectability, and analytic usefulness in collaboration with wildfire scientists.
4. Analyze results to identify how visual encoding choices affect expert reasoning about wind-fire interactions.

## 1.1 Evaluation and Validation Plan

We will evaluate WindVis using both qualitative expert feedback and quantitative visual metrics:

- **Expert Assessment:** Wildfire modelers and wildfire scientists will compare baseline and vortex-focused visualizations, rating interpretability, clarity, and analytic usefulness on a scale.
- **Feature Agreement:** We will compare identified vortex regions with computed vorticity magnitude fields to measure alignment between visual and analytic representations.
- **Coverage Metrics:** We will quantify the spatial and temporal coverage of vortex structures captured in each visualization to assess how well different strategies expose relevant airflow patterns.

These evaluations will generate quantitative and qualitative evidence of interpretive improvement, linking visualization design decisions to domain insight.

# 2 Significance

Wildfires are increasingly destructive events that affect human communities, ecosystems, and climate dynamics. Accurate models and visualizations of wildfire spread are essential for understanding behavior and informing fire management strategies, such as prescribed burns and suppression tactics. While prior work has visualized fire intensity, terrain coupling, and vegetation effects, wind-driven vortices are rarely targeted as primary visualization objects.

This project introduces a visualization framework that makes vortex structures visible in ensemble simulations, thereby addressing a critical interpretive gap. Fire science has documented that vortices can trigger sudden lateral fire expansions and complicate suppression efforts. By making these features accessible to visual exploration, we connect small-scale airflow dynamics to broader fire spread outcomes. Beyond domain applications, this work contributes to visualization studies by demonstrating how choices in representation can either obscure or reveal consequential structures in multifield data.



Beyond its domain application, this project contributes to visualization research by investigating how specialized visual encodings influence interpretive reasoning in complex, multifield datasets. WindVis serves as a case study in applying expert-centered evaluation (Lam et al., 2012) to environmental simulations, emphasizing interpretability, trust, and analytic value over purely aesthetic or algorithmic improvements. Quantitative measures of vortex coverage and expert agreement will help formalize the link between visualization design and domain insight, which is an underexplored dimension in wildfire visualization research.

### 3 Related Work

Visualization of wildfire simulations has a strong foundation in scalar field analysis.

The 2022 IEEE Scientific Visualization Contest winners [1] produced ensemble wildfire visualizations that clarified terrain-fire interactions but did not highlight wind fields or vorticity. Our project builds directly on this work by narrowing attention to vortex structures.

In fire behavior modeling, FIRETEC [4, 3] remains a cornerstone of physics-based wildfire simulation. Subsequent studies [5, 2] use such models to investigate coupled fire-atmosphere dynamics, but the visualizations presented are often static and do not isolate vortices.

Wildfire science research [7, 6] has emphasized pyrogenic vorticity and fire whirls as destructive hazards, motivating our focus on vortex visualization. These studies underscore the real-world significance of vortices, but they rely primarily on analytic and mathematical descriptions rather than visual representation.

Finally, broader visualization work [8] demonstrates the role of high-resolution modeling in capturing turbulence in fire simulations, yet stops short of proposing interactive frameworks. Our project responds by creating exploratory visualization pipelines that make wind-driven structures explicit.

Previous IEEE SciVis and VisContest challenges (2019–2023) have featured airflow and turbulence visualization tasks that partially overlap with this project’s focus. While those studies demonstrated the aesthetic and computational challenges of representing vortical structures, they did not evaluate interpretive usefulness or domain relevance. WindVis extends this tradition by explicitly connecting airflow visualization to wildfire-specific scientific questions.

### 4 Research Plan

We propose a six-week timeline modeled on comparable visualization projects:

- **Weeks 1–2:** Load wildfire ensemble datasets into ParaView and Python. Extract vector fields for wind and scalar fields for fire intensity. Produce baseline scalar visualizations.  
*Milestone: working dataset pipeline and fire intensity visualization.*
- **Weeks 3–4:** Implement vorticity visualizations using streamline tracers, glyphs, and contours. Test seeding strategies and thresholds to emphasize vortex structures.  
*Milestone: preliminary vortex visualizations for selected timesteps.*
- **Weeks 5–6:** Develop comparative animations (wind-focused vs. intensity-focused). Gather informal feedback from collaborator Divya Banesh (LANL) and peers. Refine for interpretability and prepare deliverables.  
*Milestone: completed visualization narrative and presentation.*

## References

- [1] G. Borrelli et al. 2022 ieee scientific visualization contest winner: Multifield analysis of vorticity-driven lateral spread in wildfire ensembles. *IEEE Computer Graphics and Applications*, 44(1):40–49, 2024.
- [2] G. Forney. Visualization, a tool for understanding fire dynamics. [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=861441](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=861441), 2007. Accessed September 30, 2025.
- [3] R. Linn, J. Reisner, J. J. Colman, and J. Winterkamp. Studying wildfire behavior using firetec. *International Journal of Wildland Fire*, 11(4):233–246, 2002.
- [4] R. R. Linn. *A transport model for prediction of wildfire behavior*. PhD thesis, New Mexico State University, 1997.
- [5] Q. Meng, Y. Huai, J. You, and X. Nie. Visualization of 3d forest fire spread based on the coupling of multiple weather factors. *Computers & Graphics*, 110:58–68, 2023.
- [6] J. J. Sharples and J. E. Hilton. Modeling vorticity-driven wildfire behavior using near-field techniques. *Frontiers in Mechanical Engineering*, 5:69, 2020.
- [7] J. J. Sharples, A. E. Kiss, J. Raposo, D. X. Viegas, and C. C. Simpson. Pyrogenic vorticity from windward and lee slope res. In *International Congress on Modelling and Simulation (MODSIM)*, page 29197, Gold Coast, Australia, 2015.
- [8] Q. Wang, M. Ihme, R. R. Linn, Y.-F. Chen, V. Yang, F. Sha, C. Clements, J. S. McDanold, and J. Anderson. A high-resolution large-eddy simulation framework for wildland fire predictions using tensorflow. *International Journal of Wildland Fire*, 32:1711–1725, 2023.

# Olivia Heng

[hengoliviadja@gmail.com](mailto:hengoliviadja@gmail.com) | (609) 225-2109 | [linkedin.com/in/olivia-heng/](https://www.linkedin.com/in/olivia-heng/)

## EDUCATION

**Brown University, Providence, Rhode Island** | Computer Engineering, Sc.B. & Computer Science, Sc.M. September 2022 – **Expected Graduation May 2026**

- *Relevant Coursework:* Data Structures and Algorithms, Deep Learning, Computer Vision, Computer Systems, Digital Electronics Systems Design, Design of Computing Systems, Digital Integrated Circuits, Pattern Recognition and Machine Learning, Linear Algebra, Multivariable Calculus, Differential Equations
- *Current Coursework:* Embedded Microprocessor Design (Capstone), Interdisciplinary Scientific Visualization, Electricity and Magnetism

## ENGINEERING EXPERIENCE

**School of Engineering, Brown University**

- *Teaching Assistant ENGN0030: Introduction to Engineering* September 2024 – Present
  - Host office hours and teach EMG Claw project for **40+ students** in Wearables module focusing on **Arduino** sensors in addition to course grading
- *Lead Teaching Assistant ENGN0032: Introduction to Engineering: Design* February 2023 – January 2024
  - *Lead Studio TA:* Instruct **25+ students** in a weekly studio focusing on engineering concepts, such as vectors, coupling, moments, and MATLAB
  - *The Chair Prototype Project Lead:* Coordinate weekly development meetings, ensure clear communication with Design Workshop and 50+ staff
  - *Communications Lead:* **Redesigned lecture structure** and content based on previous iterations of the course, coordinate TopHat and EdStem tools, analyze feedback data to propose changes to course, develop the course website and update content on a weekly basis for the **340+ student course**
- *Teaching Assistant ENGN0040: Dynamics and Vibrations* January 2024 – May 2024
  - Host weekly **office hours** aiding students with problems sets and design projects focused on course topics, such as kinematics of rigid bodies/systems

## COMPUTER SCIENCE EXPERIENCE

**Microsoft, Azure Data Analytics**

- *Software Engineer Intern* May 2025 – August 2025
  - Implemented **serialization pipelines** for **Substrait** query plans using alternative frameworks (FlatBuffers, Cap'n Proto, FBE) across **C#** and **C++**
  - Designed and **benchmarked** TPC-H workloads to assess performance **trade-offs** (schema translation complexity, memory footprint, message size) in real-world **data warehousing** scenarios, contributing findings to internal tooling and future architectural decisions
  - Partnered with engineers across Azure Data Analytics to validate **cross-language compatibility**, troubleshoot runtime integration issues, and refine schema evolution strategies, contributing to a more resilient and **scalable** serialization layer for internal and external consumers
- *Software Engineer Intern* May 2024 – August 2024
  - Integrated an **open-source specification** for complex query plans into SQL Server Management Studio, which is currently in deployment (SSMS 21)
  - Developed a **unique graphical visualizer** with the open-source integration using **C#, Visual Studio**, and **.NET libraries**, enhancing the efficiency of query development processes by providing intuitive access to critical performance data
  - Collaborated with the Azure Data team to enhance and leverage their robust cloud computing capabilities while learning best development techniques

**Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility**

- *Education Collaboration Software Engineer Intern* January 2024 – May 2024
  - Project: *Exploring the Integration of Tensorflow Models into NVIDIA FLARE*
    - Integrated an open-source Tensorflow model into NVIDIA Federated Learning Application Runtime Environment for supercomputer **Frontier**
    - Utilized **Python**, **reStructuredText**, and **Tensorflow** for integration and contributed to **NVIDIA FLARE** and OLCF documentation
- *Technical and Professional Software Engineer Intern* August 2023 – December 2023
  - Project: *Utilizing Asynchronous Dynamic Content Rendering with Hotwire's Turbo and Applying Code Efficiency to ORNL Existing Applications*
    - Implemented **Hotwire's Turbo** within the website application: Facility Efficiency Tracking Application
    - Establish the **software lifecycle** for the provided project for operational software systems in a production environment
- *Pathways to Computing Software Engineer Intern* May 2023 – July 2023
  - Project: *Improving Metric Granularity to Enable Research on Optimal Work Situations through Software Development*
    - Built a **full stack** website application, the Facility Efficiency Tracking Application (FETA), to track the ORNL Software Services Development Team's efficiency using **Ruby on Rails 7**, **JavaScript**, **Git**, and **CI/CD**
    - Resulted in greater issue granularity, a **29% increase** in Search Engine Optimization, and **20% increase** in Accessibility compliance

**Hack@Brown, Brown University**

- *Director* September 2022 – March 2025
  - Direct event logistics for Brown University's oldest and largest hackathon, **hosting 500+ students** annually for an in-person 24-hour hackathon
  - Lead and coordinate with **41 members** across Sponsorships, Experience, Dev, Workshops, and Design Teams to prepare for hackathon, **raising \$26k+**
- *Experience Team Lead:* Lead Experience Team to design the ideal hacker experience by planning all hackathon logistics, such as publicizing, sourcing panelists and judges, etc; Coordinated with team leads and directors throughout the year to effectively lead the executive team on the day-of the hackathon
- *Experience Team Member:* Connect with industry alumni and professionals and coordinate swag and registration

**Code With Klossy**

- A 501(c)(3) charitable organization that creates learning experiences and opportunities for women and gender expansive youth through passion igniting programs
- *Machine Learning and Artificial Intelligence Instructor Assistant* February 2023 – August 2023
  - Led Culture of Tech sessions "Creating Opportunity" and "Designing Tech Responsibility" to inform scholars of the unique challenges that women, gender non-conforming people, and people of color face in tech along with forwardly thinking of solutions — working **individually with ~30 scholars**
  - Support scholars with **debugging**, **troubleshooting**, and **adapting content** to meet scholar needs throughout the program, specifically using **Co-ML**

## SKILLS AND INTERESTS

**Technical:** Java, Python, C#, JavaScript, HTML, CSS, Verilog, C, C++, MATLAB, SQL, Swift, Ruby, Ruby on Rails, ReasonML, Racket; Git, Bash, Linux, Tensorflow, NumPy, Docker, Google Colab, .NET libraries, Quartus, Sequel Pro, Fusion360, OnShape

**Certifications:** Azure Data Fundamentals, Azure AI Fundamentals (Microsoft); Cybersecurity (Google); High Performance Computing (ORNL, Purdue University)

**Industrial:** Laser cutting, woodworking, metalworking, 3D printing, soldering

**Other Skills:** American Sign Language (basic), certified in CPR

## PROFESSIONAL ACTIVITIES

**Publication:** DOI: 10.1145/3626203.3670571 - "Improving Metric Granularity to Enable Research on Optimal Work Situations through Software Development"

**Conferences:** PEARC24: Human Powered Computing (**Student Poster**), Supercomputing '25, '24, and '23 (**Lead Student Volunteer**, HPC Immersion Program), PEARC23: Computing for the Common Good (**Student Volunteer**, Student Program), Richard Tapia Celebration of Diversity in Computing Conference '23

**Extracurriculars:** Girls Who Code College Loops (**Treasurer**), Society of Women Engineers, Mosaic+, Society of Asian Scientists and Engineers, Women in CS

# Aditi Kannan

(401)-230-2439 | [aditi\\_kannan@brown.edu](mailto:aditi_kannan@brown.edu) | [linkedin.com/in/aditi-kannan-73289a252](https://www.linkedin.com/in/aditi-kannan-73289a252) | [github.com/aditileclerc](https://github.com/aditileclerc)

## EDUCATION

**Brown University**, Providence, RI

*Master of Science in Computer Science*

*Expected May 2027*

Relevant Coursework: Deep Learning, Interdisciplinary Scientific Visualization

**Shiv Nadar University**, Chennai, India

*Bachelor of Technology in Computer Science Engineering*

*July 2025*

Relevant Coursework: Machine Learning, Embedded AI, Artificial Intelligence, Distributed Computing

Honors: First Class with Distinction

## SKILLS

Computer Skills: C, C++, Java, Python, Arduino, NS2, CARLA, TensorFlow, Keras, PyTorch, HTML, CSS, JavaScript, React, Node, Streamlit, MySQL, Cypher, Git, Llama

Research Skills: Microsoft Excel, Tableau

## ACADEMIC PROJECTS

**Loan Status Prediction Model** [[Github](#)]

September 2022 - December 2022

- Developed a machine learning-based loan status prediction model tailored to the needs of small financial institutions and non-bank financial companies.
- The model leverages the Support Vector Machine (SVM) and Linear Regression algorithms to analyze borrower data and predict the likelihood of loan approval or default.

**GPU Accelerated Brain Simulations with Procedural Connectivity**

September 2024 - February 2025

- The project accelerates the Spiking Neural Network (SNN) simulation of Biological Neural Networks, on a single GPU system, using parallel computing methods and simulate the trials using GeNN simulator.

**Stock Price Prediction Using Gen AI and Opinion Mining**

March 2024 - November 2024

- The model predicts the market value of stocks using a generative AI model, such as a transformer model to perform opinion mining and judge the value of the stock. A financial RNN, FinBERT, which is a Transformer based model and VADER, was used for the opinion mining.

## PROFESSIONAL EXPERIENCE

**Zoho Corporation, Chennai, India**

December 2023 - January 2024

*Internship*

- As a part of the Key Manager Plus team, I worked with SSL and SSH key certificates and developed a system of storing and verifying security certificates using Java and JavaScript

**PurpleSlate Ltd, Chennai, India**

May 2024 - June 2024

*Internship*

- Worked mainly with Generative AI, Large Language Models, Prompt Engineering and Transformer model based chatbots, with the ParrotGPT team. Built a custom chatbot [[Github](#)]

## LEADERSHIP POSITIONS

**English Literary Club, Shiv Nadar University, Chennai**

July 2022 - May 2025

*Club President*

- Organized several events and club activities, implemented and managed the club structure.

**SSN-SNUC MUN, Shiv Nadar University, Chennai**

August 2024 - September 2024

*Under Secretary General of Technology*

- Worked along with my team of developers to redesign the conference website and worked on a lunch verification application.

**B****Banesh, Divya**

to ahrens@lanl.gov, laidlaw.david@gmail.com, David, Sir, Aditi, me ▾

Wed, Sep 17, 11:05 PM ☆ ↶ ⋮

Hello all,

It was nice to meet you today. As we discussed, here are a few resources to get your started on your project proposals:

1. The LANL websites are currently being updated. Here is the new link to the 2022 SciVis Contest: <https://www2.lanl.gov/projects/sciviscontest2022/index.html>
2. This website above links to both the document regarding the data details and the open data server, named Oceans11, that hosts the data.
3. As we started discussing today, the wildfire data can be visualized in Paraview. Paraview can be downloaded here: <https://www.paraview.org/>
4. The tutorial to help you get started with Paraview can be found here: <http://docs.paraview.org/en/latest/Tutorials/SelfDirectedTutorial/index.html>
5. Attached is a .pvsm file. It's called a Paraview State File and allows us to pass visualization workflows from one person to another to help share some initial information about how to get started. To use this particular state file:
  - a. Download the data from here: [output.70000.vts](#)
    - i. Right click and "Save As" and name it: backCurve40.70000.vts
  - b. Open Paraview, go to "File", "Load State", and navigate to the location of the pvsm from this email you've downloaded.
  - c. Once you select the pvsm, it should open another window that allows you to Load State Options.
  - d. In the drop down, select "Choose File Names" and select the vts file you've just downloaded. Then hit Ok at the bottom.
  - e. It might take a few minutes, but at the end, you should see the view attached.
  - f. This now visualizes the landscape, the vegetation, the fire and the smoke. From here, you should be able to do more interesting things for your project.

Good luck and let me know if you have questions.

# From Scattered Clusters to Similarity Structures: A Graph-Based Spherical Tiling for Genotype Visualization

Ottocaro Weiss  
PI

`ottocaro-weiss@brown.edu`

Alex Diaz-Papkovich  
Collaborator

`alex_diaz-papkovich@brown.edu`

October 20, 2025

## Abstract

We propose a novel clustering and visualization pipeline for population genetics data. Standard approaches such as UMAP scatterplots emphasize differences while encoding non-informative global distances, which undermines interpretability and seeds unscientific claims. Our method instead grounds the representation in genetic similarity, arranging samples into planar constructions that maximize neighborhood connections and shared relatedness. Projecting these embeddings onto a sphere enables refinement of edge lengths to reflect similarity, yielding a globally stable, crossing-free layout. Through the barycentric dual, this process produces a spherical tiling in which polygon shape, size, and color convey both global and local structure.

# 1 Aims

The goal of this project is to develop, implement, and evaluate a novel graph-based visualization pipeline for population genetics data. Specifically, we aim to:

- **Experiment and evaluate the application of Planar Maximally Filtered Graphs (PMFGs) and related weight maximal, triangulated, non-crossing graphical constructions for the analysis of genotype similarity matrices.**
- **Develop approaches that balance near-optimal edge selection (i.e. high overlap with MSTs and PMFGs) with the geometric flexibility required for encoding visual features and maintaining interpretability.**
- **Generate interpretable spherical tilings based on the duals of these graphical constructions, employing layout optimization techniques to drive correlation between similarity and edge length.**
- **Experiment and evaluate coloring and edge thickness schemes to communicate gene flow, groupings, centrality, and other aspects of the network to increase interpretability and limit misuse.**
- **Compare with previous population genetics visualization methods such as UMAP, PCA, and TMAP through statistical metrics and user studies.**

# 2 Motivation

Recent critiques and debates underscore **methodological and ethical shortcomings in current approaches to human population genetics visualization**. Standard methods—t-SNE, PCA, and UMAP—exaggerate separation, distort global distance, and invite over-interpretation. PCA scatter-plots are sensitive to sampling and scaling and can yield contradictory narratives [12], while UMAP and t-SNE optimize local neighborhoods at the expense of meaningful global geometry [2]. Coupled with categorical labels and color schemes, the resulting figures emphasize difference and separation among populations, fueling narratives with roots in scientific racism and eugenics [5]. Recent cases show genetics figures appropriated or altered online to promote this sense of division founded on misinterpretations of the underlying data [15, 6]. Some have proposed guidelines and optimization techniques to ameliorate these issues [8, 14], but the overarching DR-based approach persists.

Standard DR-based methods represent individual samples in two-dimensional and three-dimensional embedding space, endowing points with Euclidean ( $\ell_2$ ) distance metrics; however, the underlying high-dimensional genotype data comprise discrete—and relatively rare—variations of allelic states. We argue that the limitations of current approaches stem from a mismatch between data and representation—e.g., (i) visualizing  $k$ -nearest-neighbor graph-based embeddings as scatter plots in  $\mathbb{R}^2/\mathbb{R}^3$  (UMAP/t-SNE), and (ii) projecting samples that are  $\sim 99.9\%$  identical onto axes of maximal variance (PCA) [21]. These issues motivate designs that foreground continuity and relatedness, and that communicate uncertainty and gradients rather than categorical separations [5].

### 3 Approach and Related Work

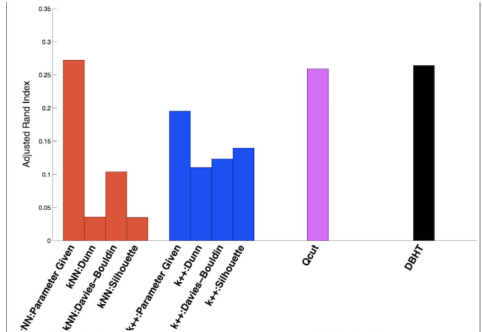
Graph-centric approaches illustrate networks of connectivity and similarity for analysis and visualization. Findings from across population genetics suggest graph-based methods are a natural fit for genetic data. Our work will directly build on TMAP, which popularized an MST-first mapping for high-dimensional genetic datasets via LSH/MinHash neighbor search and Kruskal-based spanning trees [23]. Even while enforcing a single-path connectivity structure, TMAP outperforms UMAP on local neighborhood preservation when applied to the ChEMBL dataset.

Population Graphs model conditional genetic covariance and prune redundant links, recovering a sparse network of effective connectivity rather than an arbitrary spatial embedding [11, 10]. For human genetics data, spectral and graph-partitioning approaches to individual-level similarity graphs detect substructure without projecting onto axes with uninterpretable distances [16, 28]. At the population scale, admixture graphs such as TreeMix make drift and mixture explicit in an interpretable topology [22], and IBD networks in large cohorts recover fine-scale, demographically coherent communities [20, 3].

This past literature provides grounding for a more expansive graph-centric approach that couples visual and analytical expression and precision. We borrow TMAP’s scalable neighbor construction but visualize a maximal planar mesh instead, preserving cycles/triangles critical for expressing overlapping affinities, while retaining the option to overlay an MST skeleton for overview. In our pipeline, similarity defines the object of inference and the canvas alike: we first construct a sparse connectivity graph that preserves conditional relationships, then arrange nodes to maximize adjacency among genetically similar samples. To initialize and stabilize this topology at scale, we filter the full similarity matrix with a planar backbone—starting from a maximum-similarity spanning tree and expand via similarity and geometric quality towards a complete planar triangulation [29, 18]. Independent evaluations show that planar graph filters (PMFG/TMFG) retain substantially more edge-level agreement with the full network than MST and support both intra-cluster and inter-cluster hierarchies [27, 19].

DBHT is a hierarchical clustering algorithm that takes advantage of this exact feature, exploiting separating 3-cliques in PMFG/TMFG to induce a deterministic, parameter-free hierarchy of “bubble” clusters [27, 26]. As shown in the figure on the right, DBHT provides competitive accuracy while avoiding tunable parameters, aligning with our goal of interpretable, cycle-preserving structure.

PMFG and TMFG yield maximal planar triangulations with up to  $3n - 6$  edges, which imposes an extremely rigid topology. Much of our work will involve testing and implementing strategies to select and align edges with preferred geometric qualities at construction or through edge-swapping. In our final step, we will take the barycentric dual tiling of our crossing-free spherical layout, encoding multi-scale structure—polygon adjacency for local neighborhoods; polygon shape/area/color for gradients of relatedness and uncertainty. This similarity-first, graph-grounded visualization preserves the analytic virtues of network models while providing a compact, interpretable display that resists spurious geometric inference. The resulting visualizations would emphasize connections and overlap rather than distances and separation, while also fitting better to the underlying data. The visualization below shows a direct TMFG applied on the sphere: starting from the canonical tetrahedral seed ( $K_4$ ), vertices are inserted into triangular faces and all edges are rendered as minor geodesic arcs, which is

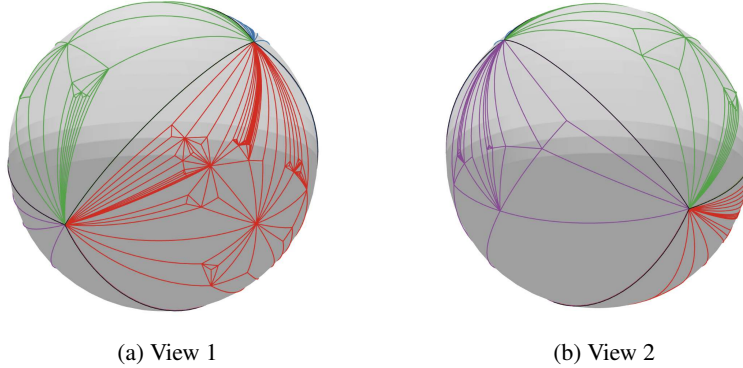


**Mean Adjusted Rand Index (ARI) over multiple clustering techniques on *cDNA microarray* datasets [27].**



essential for downstream layout optimization tasks.

## 4 Preliminary Results



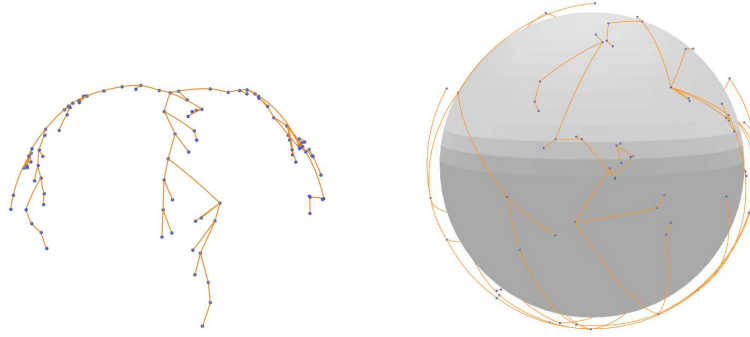
**TMFG non-crossing spherical embedding with 1,000 samples from the 1000 Genomes Project with tetrahedral faces distinguished by color.**

The visualization above does not incorporate these additional layout optimizations for aligning arc lengths with the underlying similarity values; nevertheless, it performs comparably to UMAP even without explicit distance calibration.

| Method             | Global        |               |               | Local (k=25)  |               |               | Local (k=5)   |               |               | Time (s)    |
|--------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|-------------|
|                    | Stress        | Mantel $r$    | Spearman      | Trust         | Cont.         | Jacc.         | Trust         | Cont.         | Jacc.         |             |
| TMFG (Euclidean)   | 0.6068        | <b>0.7702</b> | <b>0.8627</b> | 0.9112        | 0.9172        | 0.0640        | 0.9154        | 0.9282        | 0.0690        | <b>1.42</b> |
| TMFG (Topological) | 0.6076        | 0.7688        | 0.8598        | 0.9118        | 0.9191        | <b>0.0850</b> | 0.9265        | 0.9323        | <b>0.1148</b> | <b>1.42</b> |
| UMAP (Spherical)   | <b>0.4376</b> | 0.1068        | 0.0688        | 0.7783        | 0.6442        | 0.0501        | 0.8022        | 0.6502        | 0.0169        | 21.37       |
| UMAP (2D)          | 0.5347        | 0.6848        | 0.5372        | <b>0.9409</b> | <b>0.9580</b> | 0.0846        | <b>0.9376</b> | <b>0.9648</b> | 0.0349        | 7.48        |

Comparison of raw TMFG construction (**no arc length or distance optimization**), Spherical UMAP, and 2D UMAP on similarity metrics on the complete 1000 Genomes Project (3450 samples). Similarity matrix from GRM pairwise distances. Spherical UMAP uses the haversine distance, as specified in documentation.

Our research will involve testing and implementing strategies to select and align edges with preferred geometric qualities at construction or through edge-swapping. To control geometric quality, we experiment with post-selection layers that prioritizes short geodesic edges and favor non-separating 3-cycles, as well as constraints during construction. Since the final graph forms a fully triangulated mesh, we can also make use of algorithms from computer graphics, such as SDEM [17, 1]. Another potential approach involves adding a loss term to a stress optimization objective to penalize crossings. Then, starting with the MST, we can add batches of highest similarity edges and remove those that still remain after a set number of iterations.



(a) MST Seed

(b) During Training

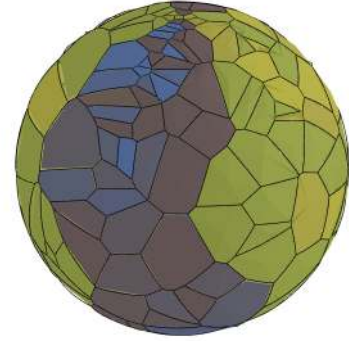
**Iterative Non-Crossing Layout Optimization with MST Backbone.**

These steps all contribute towards the final spherical tiling, which we realize by taking the barycentric dual of the triangulated, planar graph. The tiling therefore encodes a multiscale structure: polygon adjacency denotes local neighborhoods; polygon area reflects variation and centrality; color represents gradients of similarity and other correlative measures; and further emergent features arise naturally. This similarity-first, graph-based visualization preserves the analytic virtues of network models while providing a compact, interpretable display that resists spurious geometric inference. The resulting visualizations emphasize connection and overlap rather than distance and separation, aligning more closely with the structure of the underlying data. The final visualizations would likely resemble the figure below.

## 5 Significance

**Clustering and visualization as a single construct.** In our approach, clustering and visualization are the same operation: the similarity graph we construct is precisely the structure we render. This collapses the usual pipeline (cluster  $\rightarrow$  embed  $\rightarrow$  plot) into a single, connectivity-first representation in which adjacencies directly encode relatedness, reducing the interpretive gap between analysis and figure and mitigating common misreadings of global distances found in scatterplots (e.g., PCA/UMAP) [8, 9].

**Building on DBHT (clustering on maximal planar substrates).** DBHT derives a deterministic, parameter-free hierarchy from a filtered maximal planar graph, using separating 3-cliques to define “bubble” clusters and assemble a directed bubble tree [29, 18, 27]. We adopt the same filtration step to obtain a maximal planar substrate but preserve the full mesh instead of outputting only a dendrogram. The result is a visual realization of DBHT’s clusters in which the cluster structure and the visualization coincide, modified only insofar as such adjustments improve interpretability of the underlying data [26, 24].



**Illustrative Final Result with Synthetic Data.**

**Graph drawing and planar spherical embeddings.** Tutte’s barycentric embedding guarantees a planar, crossing-free drawing inside a convex boundary; related planar/spherical layout techniques and circle-packings are surveyed in the graph drawing literature [13, 7, 30]. A fundamental finding in graph theory is that all planar graphs can be embedded on a sphere, and this proposal seeks to explore stochastic and deterministic approaches for realizing these constructions. Further layout optimization yields stable, reproducible meshes that balance graph theoretic and geometric glyph-related optimums. Advances in the drawing of planar graphs could help address the ever-present “hair-ball” problem in graph visualization [25].

**From mesh to tiling via the barycentric dual.** From the crossing-free spherical triangulation, we construct the barycentric dual by connecting triangle centroids across shared edges. On the sphere, this yields a boundary-free tiling with one polygon per sample whose boundary follows the cyclic order of its one-ring neighbors. The dual preserves adjacency semantics and solves overplotting/labeling issues inherent to scatter-plots. At the cell level, polygon area can encode local similarity density; shape captures directional neighborhood structure; and color conveys metadata or centrality—making overlap, continuity, and admixture visually explicit [4, 31].

**Impact on Population Genetics Visualizations and Beyond.** Together, these choices replace distance-first scatter-plots with a stable, connectivity-first spherical tiling. This emphasis on networks of connection and diversity could lead to new avenues of population genetics research and fewer cases of harmful misinterpretation and misuse. As previously argued, network-based visualizations may prove a better fit for population genetics data. This approach allows for us to perform an efficient, graph-theoretic alternative to dimensionality reduction, which could be expanded on further. The visualization technique is generalizable to any similarity matrix, so other scientific domains could also benefit.

## 6 Research Plan

- **Weeks 1–2 — Robust, flexible end-to-end pipeline:** Build a documented, reproducible pipeline with clear explanations and visuals for each stage: data loading → similarity matrices (e.g., GRM, Jaccard, LSH) → planar substrate (PMFG/TMFG or order-based triangulation) → planar embedding and spherical lift → no-crossing geodesic optimization → barycentric-dual tiling and encodings.
- **Weeks 3–4 — Applications on real datasets:** Apply the pipeline to human population datasets and animal species cohorts; stress-test scalability, run parameter sweeps, and collect qualitative feedback from domain users. Evaluate case-studies with human testers with before/after visuals and notes on interpretability and stability.
- **Weeks 5–6 — Final evaluation, testing, and deliverables:** Quantitatively compare against PCA/UMAP (stability under resampling,  $k$ NN neighborhood preservation, triangle quality/area uniformity, planarity invariants, and agreement with known structure). Conduct a small human evaluation with domain users (task accuracy/time on structure-identification tasks; ratings of interpretability/trust; general feedback). Finalize artifacts: open-source repository with documentation, figures, slide deck, and illustrative examples.

## References

- [1] N. Aigerman, S. Z. Kovalsky, and Y. Lipman. Spherical orbifold tutte embeddings. *ACM Trans. Graph.*, 36(4), July 2017.
- [2] C. J. Battey, G. C. Coffing, and A. D. Kern. Visualizing population structure with variational autoencoders. *G3: Genes, Genomes, Genetics*, 11(1), 2021.
- [3] G. M. Belbin, S. Cullina, S. Wenric, E. R. Soper, B. S. Glicksberg, D. Torre, and et al. Toward a fine-scale population health monitoring system. *Cell*, 184(8):2068–2083.e11, 2021.
- [4] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. ESRI Press, Redlands, CA, 2010.
- [5] K. A. Bird and J. Carlson. Typological thinking in human genomics research contributes to the production and prominence of scientific racism. *Frontiers in Genetics*, 15:1345631, 2024.
- [6] J. Carlson, B. M. Henn, D. R. Al-Hindi, and S. Ramachandran. Counter the weaponization of genetics research by extremists. *Nature*, 610(7932):444–447, 2022.
- [7] C. R. Collins and K. Stephenson. A circle packing algorithm. *Computational Geometry*, 25(3):233–256, 2003.
- [8] A. Diaz-Papkovich, L. Anderson-Trocmé, C. Ben-Eghan, and S. Gravel. Umap reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLOS Genetics*, 15(11):e1008432, 2019.
- [9] A. Diaz-Papkovich, L. Anderson-Trocmé, and S. Gravel. A review of UMAP in population genetics. *Journal of Human Genetics*, 66(1):85–91, 2021.
- [10] R. J. Dyer. Population graphs and landscape genetics. *Annual Review of Ecology, Evolution, and Systematics*, 46:327–342, 2015.
- [11] R. J. Dyer and J. D. Nason. Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, 13(7):1713–1727, 2004.
- [12] E. Elhaik. Principal component analyses (pca)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, 12(14683), 2022.
- [13] H. Gibson, J. Faith, and P. Vickers. A survey of two-dimensional graph layout techniques for information visualisation. *Information Visualization*, 12(3-4):324–357, 2013.
- [14] H. Huang, Y. Wang, C. Rudin, and E. P. Browne. Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. *Communications Biology*, 5(719), 2022.
- [15] M. Kozlov. ‘all of us’ genetics chart stirs unease over controversial depiction of race. *Nature*, Feb 2024<sub>02</sub>.
- [16] A. B. Lee, D. Luca, and K. Roeder. A spectral graph approach to discovering genetic ancestry. *The Annals of Applied Statistics*, 4(1):179–202, 2010.
- [17] Z. Lyu, L. M. Lui, and G. P. T. Choi. Spherical density-equalizing map for genus-0 closed surfaces. *SIAM Journal on Imaging Sciences*, 17(4):2110–2141, 2024.

- [18] G. P. Massara, T. Di Matteo, and T. Aste. Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks*, 5(2):161–178, 2017.
- [19] T. Millington. An investigation into the effects and effectiveness of correlation network filtration methods with financial returns. *PLOS ONE*, 17(9):e0273830, 2022.
- [20] J. Nait Saada, G. Kalantzis, D. Shyr, F. Cooper, M. Robinson, A. Gusev, and P. F. Palamara. Identity-by-descent detection across 487,409 british samples reveals fine-scale population structure and ultra-rare variant associations. *Nature Communications*, 11(1):6130, 2020.
- [21] National Human Genome Research Institute. Genetics vs. genomics fact sheet.
- [22] J. K. Pickrell and J. K. Pritchard. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, 8(11):e1002967, 2012.
- [23] D. Probst and J. Reymond. Visualization of very large high-dimensional data sets as minimum spanning trees. *Journal of Cheminformatics*, 12:12, 2020.
- [24] S. Raphael and J. Shun. Faster parallel triangular maximally filtered graphs and hierarchical clustering. 2024.
- [25] H.-J. Schulz and C. Hurter. Grooming the hairball - how to tidy up network visualizations? In *INFOVIS 2013, IEEE Information Visualization Conference*, Atlanta, United States, Oct. 2013. IEEE. Tutorial.
- [26] W.-M. Song, T. Di Matteo, and T. Aste. Nested hierarchies in planar graphs. *Discrete Applied Mathematics*, 159(17):2135–2146, 2011.
- [27] W.-M. Song, T. Di Matteo, and T. Aste. Hierarchical information clustering by means of topologically embedded graphs. *PLOS ONE*, 7(3):e31929, 2012.
- [28] S. Sridhar, S. Rao, and E. Halperin. An efficient and accurate graph-based approach to detect population substructure. In T. Speed and H. Huang, editors, *Research in Computational Molecular Biology (RECOMB 2007)*, volume 4453 of *Lecture Notes in Computer Science*, pages 503–517, Berlin, Heidelberg, 2007. Springer.
- [29] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426, 2005.
- [30] D. Vrajitoru and J. DeBonis. Consistent weighted graph layouts. In *Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications (AICCSA 2005)*, Cairo, Egypt, 2005. IEEE.
- [31] A. Yavari. On geometric discretization of elasticity. *Journal of Mathematical Physics*, 49, 01 2008.

# OHCA-Vis: Interactive Visualization of Determinants in Pediatric Out-of-Hospital Cardiac Arrest

Ryan Huang  
PI

`ryan_y_huang@brown.edu`

Latha Ganti  
Collaborator

`latha_ganti@brown.edu`

October 12, 2025

## Abstract

We present a novel method called *OHCA-Vis* for visualizing pediatric out-of-hospital cardiac arrest (OHCA) data using interactive alluvial plots. The method overcomes a limitation in registry dashboards that lack data insights and standard alluvial plots that become overly convoluted. To illustrate the interpretability of *OHCA-Vis*, we will conduct user studies and compare findings with state-of-the-art methods.

# 1 Response to Reviewer Comments

We appreciate the constructive feedback on the manuscript and have revised the proposal accordingly. Below we provide an itemized list of reviewer comments and summarize how each has been addressed in the revised version.

## Reviewer 1 (R1)

[Link to R1](#)

### **[R1.1] The interpretability metrics could be more detailed.**

*Response:* We have expanded the discussion of interpretability in the *Color-encoded Feature Importance Integration* section. The revised text now describes how color and opacity encodings will represent feature importance and confidence levels based on perceptual uniformity principles. We also note that we are going to compute feature importance scores using random forest or SHAP-based models.

### **[R1.2] Limited discussion of data.**

*Response:* We have elaborated on the dataset description within the *Aims* section, describing the size of the data and presenting a snapshot of the dataset in Figure 1.

### **[R1.3] Evaluation relies only on qualitative user feedback; consider quantitative metrics.**

*Response:* The *Evaluation* section has been substantially expanded to include both quantitative and qualitative measures. In addition to clinician feedback and usability observations, quantitative metrics such as task completion time, accuracy of insights, and visual clarity metrics have been added to assess system performance.

### **[R1.4] The proposal's novelty may be somewhat incremental, combining existing visualization and interpretability techniques.**

*Response:* The revised manuscript clarifies that the novelty lies in dynamically integrating feature importance with interactive alluvial visualization. *OHCA-Vis* uniquely allows users to add, remove, and filter variables in real time, with corresponding visual and computational updates to the importance encodings. These visual additions have never been investigated before.

## Reviewer 2 (R2)

[Link to R2](#)

### **[R2.1] The proposal acknowledges scalability and readability issues with alluvial plots but does not offer a concrete strategy to address them.**

*Response:* The revised *Interactive Plot Implementation* and *Color-encoded Feature Importance Integration* sections describe several strategies for maintaining readability and scalability. These include allowing users to add or remove variables dynamically to manage visual complexity, applying edge overlap minimization algorithms such as NeighborNet and Force-Directed Layouts, and using opacity to reduce clutter in less influential pathways.

**[R2.2] The real-world clinical impact remains speculative, with no clear evidence it will change decision-making or outcomes.**

*Response:* We have clarified in the revised *Evaluation* section that the study’s immediate goal is to enhance interpretability and hypothesis generation for clinicians. The tool is intended to support exploratory data analysis and pattern recognition, providing a foundation for future research rather than immediate changes in clinical decision-making.

**[R2.3] The proposal combines known methods (alluvial plots + feature importance) rather than introducing a genuinely new visualization or analysis paradigm.**

*Response:* As mentioned in the *Aims* and *Significance* sections, the novelty of this work lies in the integration and interactivity of these methods. Unlike previous static approaches, *OHCA-Vis* dynamically recomputes and encodes feature importance values as users adjust filters or variables.

### Reviewer 3 (R3)

Link to R3

**[R3.1] The novelty of the project could be elaborated. Are there visualization elements that have not been tried before in other fields?**

*Response:* The revised manuscript expands on novelty in both the *Significance* and *Visualization Design* sections. The contribution of *OHCA-Vis* lies in two major design innovations:

- **Dynamic feature-importance recalibration:** Importance scores and visual encodings are updated in real time as users modify filters or variable selections.
- **Interactive alluvial manipulation:** Users can add, remove, or refine variable ranges to explore clinically relevant subgroups, minimizing clutter while preserving interpretability.

These features extend beyond traditional visualization practices in clinical data exploration.

**[R3.2] Are there examples of explicit relationships that this visualization enables which are not visible in other methods?**

*Response:* We have clarified in the revised text that *OHCA-Vis* enables identification of complex relationships such as:

- **Interaction effects:** How combinations of factors—such as bystander CPR and early AED placement—affect survival likelihood more strongly than either factor alone.
- **Feature Importance:** By visualizing the feature importance of all variables of interest, the user can easily determine which factor is most associated with the survival.

These examples illustrate how the visualization uncovers multi-factor patterns not apparent in static summaries or univariate analyses.





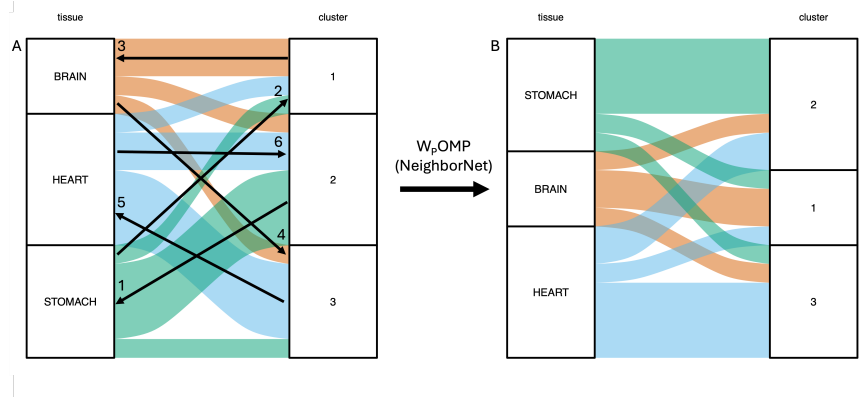


Figure 2: NeighborNet Edge Overlap Minimization

### 2.3 Evaluation

An extensive evaluation will be conducted with EMS providers and pediatric emergency physicians to assess the usability, interpretability, and clinical relevance of the *OHCA-Vis* system. The evaluation will employ a mixed-methods approach, combining quantitative performance metrics with qualitative feedback.

Participants will complete a series of task-based assessments designed to measure how effectively they can extract insights from the visualization compared to state-of-the-art methods like standard alluvial plots. Tasks will include identifying key predictors of survival, comparing intervention outcomes, and filtering by demographic subgroups. Several quantitative metrics will be collected including task completion time, accuracy of insights, and visual clarity metrics. On the qualitative side, we will ask users to describe perceived reduction in visual clutter, clinical utility, and clarity of color encodings.

## 3 Significance

Out-of-hospital cardiac arrest in children is a rare but devastating event, with only a 2–9.6% survival rate [7]. Improving outcomes requires understanding which patient characteristics and interventions most strongly influence survival. Although national and regional registries capture rich event-level data, current reporting methods rely on static tables and summary statistics that obscure variable interactions and prevent exploration of how different factors shape outcomes. As a result, valuable information for guiding clinical decision-making, protocol design, and quality improvement often remains hidden.

This project addresses that gap by introducing *OHCA-Vis*, an interactive visualization tool that uses alluvial plots to reveal which clinical and demographic factors influence outcomes. There are two key visualization components as mentioned in Section 2.

**Interactive Plot.** By allowing users to dynamically select variables and refine their ranges, the system promotes targeted exploration and hypothesis generation of key outcome determinants without overwhelming visual clutter. This approach can help clinicians identify meaningful trends, compare interventions, and better understand patient subgroups.

**Color-encoded Feature Importance.** Embedding feature importance directly within alluvial plots will allow users not only to visualize patient trajectories but also to visually identify predictors of survival. Moreover, the use of color encodings for feature importance has never been done before.

Overall, the significance of this work lies in its ability to shift registry data from a retrospective reporting

mechanism into an active, exploratory resource for hypothesis and insight generation. For EMS agencies and pediatric emergency departments, this approach can uncover critical delays, identify effective interventions, and ultimately inform strategies to improve survival in one of the most challenging clinical contexts.

## 4 Related Work

In clinical and biomedical contexts, prior work has focused on static dashboards, tables, timelines, or survival summaries that aggregate patient outcomes by individual factors such as age, sex, or intervention type [2, 3]. These methods are limited because they obscure interactions between variables, prevent exploration of how combinations of factors affect outcomes, and do not allow dynamic comparison of different predictors.

Alluvial and Sankey-style visualizations have been used to summarize categorical flows and transitions in various domains, including bibliometrics and demographic studies [10]. Moreover, alluvial plots have been used to evaluate factors in relation to COVID-19 survival rates [8]. However, there has been no work in making these plots interactive or exploratory for users to evaluate and choose which factors/variables to visualize. Given the complexity and magnitude of EMS data, these plots can be convoluted and difficult to understand.

It has been shown that the complexity of alluvial plots (number of flows) are inversely correlated with the understandability of the diagram [1]. With the amount of data that exists in EMS data, the number of flows would inevitably increase. Thus, a traditional alluvial plot would not suffice. The introduction of an editable and interpretable plot is hypothesized here to improve user understanding, as the plot would be easily simplified and user-modifiable.

While interpretability techniques such as partial dependence plots and feature importance rankings can identify influential predictors, they are typically applied in isolation to predictive models rather than in conjunction with patient event data [4]. The integration of alluvial plots and feature importance has been done as shown in Figure 3 below [6]. Yet, it doesn't consider user inputs and interactivity and, as previously mentioned, is created in conjunction with the alluvial plot.

By embedding these methods within interactive and editable alluvial plots, the proposed approach aims to provide clinicians with both interactive visualization and interpretable insights, addressing the limitations of existing static methods.

Together, these limitations motivate the need for a tool that can both visualize patient pathways and integrate interpretable analyses of factor influence, enabling clinicians to explore outcomes in a flexible and actionable way.

## 5 Research Plan

The project will be completed over six weeks, progressing from data preparation to system implementation, integration of interpretability features, and a structured evaluation with clinicians. Each stage will build upon the previous one to ensure both technical robustness and clinical relevance.

### 1. Week 1: Data Preparation and Initial Development

Clean and preprocess pediatric OHCA registry data by standardizing variable names, encoding categorical fields, and filtering incomplete records. Implement a base alluvial plot in Plotly or D3.js to visualize patient pathways across sequential EMS and hospital events (e.g., CPR provider, AED use, initial rhythm, and outcome).

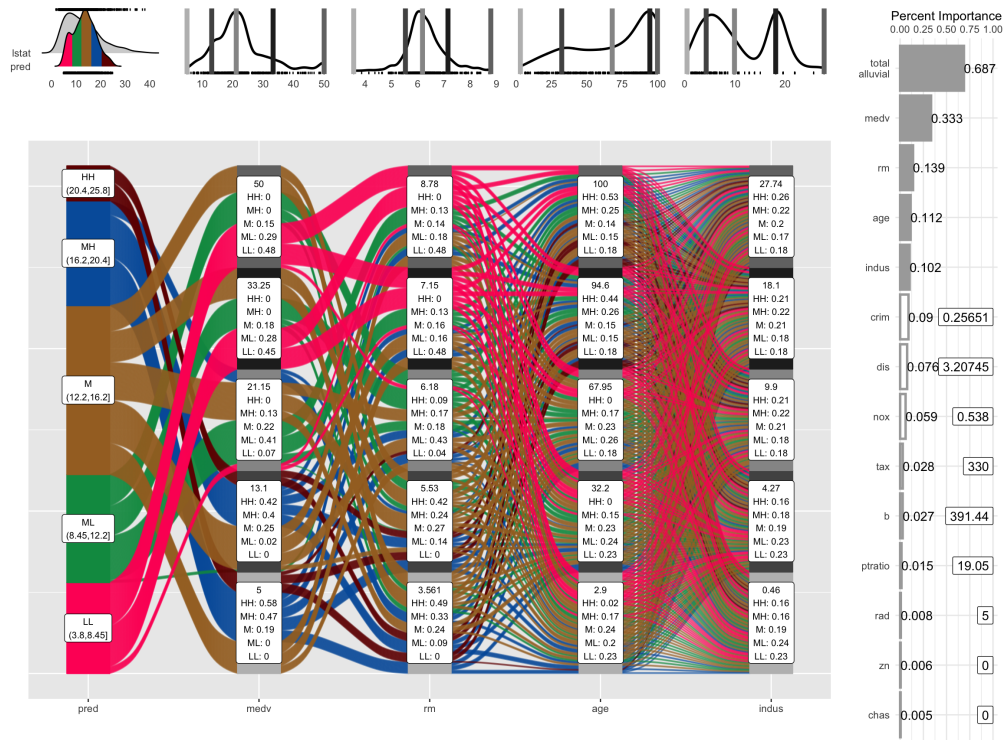


Figure 3: Alluvial Plot with Feature Importance Bar Graph for Housing Prices

## 2. Week 2: Add Interactivity to Plot

Integrate interactive controls that allow users to dynamically select and remove variables of interest, as well as filter variable ranges (e.g., age groups, intervention times).

## 3. Week 3: Integrate Color-encoded Feature Importance

Compute feature importance scores using random forest or SHAP-based models trained on the dataset to identify key predictors of survival. Integrate these scores into the alluvial plot using perceptually uniform color gradients and opacity scaling to indicate feature relevance and confidence.

## 4. Week 4: System Refinement

Implement edge overlap minimization algorithms such as Force-Directed Layout or NeighborNet to preserve interpretability of overlapping flows. Refine visual encoding schemes to ensure consistent interpretability across monitors and color vision profiles. Prepare finalized interface for clinician evaluation.

## 5. Week 5: User Study and Evaluation

Conduct structured user studies with EMS providers and pediatric emergency physicians. Participants will perform standardized tasks (e.g., identifying key predictors of survival, comparing intervention outcomes) using both *OHCA-Vis* and baseline tools (e.g., standard alluvial plots). Quantitative metrics—task completion time, accuracy, and visual clarity—will be collected, alongside qualitative feedback on usability, interpretability, and perceived clinical utility through post-task questionnaires and semi-structured interviews.

## 6. Week 6: Analysis, Reporting, and Dissemination

Analyze both quantitative and qualitative data to assess visualization effectiveness. Statistical tests will evaluate improvements in task performance and insight accuracy. Qualitative results will be coded for recurring themes related to interpretability, cognitive load, and perceived impact on clinical decision-making.

## 6 Proof of Collaboration

L

**Ganti, Latha MD**  
to Juliana, Korey, Timothy, Shravya, me ▾

Tue, Sep 30, 8:28 PM (2 days ago) ☆ ↶ ⋮

Hi Ryan & team,  
here is the data for adult cardiac arrest years 2022-2024.

here is the pediatric cardiac arrest data. 2017-2024. The data from 2012-2016 is nice and cleaned up, (2nd file), and that was the data from which we wrote the paper I had sent the group.  
It would be great if we can put the 2017-2024 data into the cleaned up 2012-2016 sheet. When they are in a single sheet, I can import into JMP and run stats.

I would suggest maybe you split up and one group work on adult cardiac arrest and one group work on peds arrest.  
We will be writing at least 1 publication from each dataset, and you are welcome to co-author the manuscripts with my team and I.

We also collect sepsis and STEMI data, I just haven't seen that to be as useful, but I will try to do another look tomorrow and see if there's anything worth exploring.

Also FYI, another one of your classmates not in this group also asked me about the PCFR data, and I sent him the stroke data.

...

4 Attachments • Scanned by Gmail ⓘ

RAW ADULT Rese...

adult cardiac arre...

Pedi Cardiac Arre...

2012-2016 peds c...

Figure 4: Email with Latha

## References

- [1] A. Arunkumar. Bayesian modelling of alluvial diagram complexity. manuscript, Arizona, United States of America, 2021.
- [2] J. Chishtie. Interactive visualization applications in population health and health services research: Systematic scoping review. manuscript, Toronto, Canada, 2022.
- [3] O. M. Christen. Dashboard visualization of information for emergency medical services. manuscript, Biel, Switzerland, 2020.
- [4] M. Hasan. Pre-hospital prediction of adverse outcomes in patients with suspected covid-19: Development, application and comparison of machine learning and deep learning methods. manuscript, Sheffield, United Kingdom, 2022.
- [5] S. G. Kobourov. Force-directed drawing algorithms. manuscript, Arizona, 2012.
- [6] B. Koneswarakantha. easyalluvial: Generate alluvial plots with a single line of code. manuscript, 2025.
- [7] A. Kämäräinen. Out-of-hospital cardiac arrests in children. manuscript, Tampere, Finland, 2010.
- [8] W. C. K.-T. T. Po-Tsung Yen, Tsair-Wei Chien. Using the alluvial diagram to display variable characteristics for covid-19 patients and research achievements on the topic of covid-19, epidemiology, pathogenesis, and vaccine (cepv): Bibliometric analysis. manuscript, Tainan, Taiwan, 2023.
- [9] J. Rich. Optimizing alluvial plots. manuscript, Pasadena, California, 2025.
- [10] A. W. K. Yeung. Data visualization by alluvial diagrams for bibliometric reports, systematic reviews and meta-analyses. manuscript, Hong Kong, Hong Kong, 2018.

# Ryan Yuki Huang

[ryan\\_y\\_huang@brown.edu](mailto:ryan_y_huang@brown.edu) | [ryanhuang.dev](http://ryanhuang.dev) | 858-519-2280

## EDUCATION

### Brown University

Providence, RI | Expected Graduation: May 2026

*Bachelor of Science in Computer Science*

GPA: 4.0/4.0

- **Relevant Courses:** Data Structures & Algorithms, Deep Learning, Computer Systems, Software Engineering, Computer Vision, Graphics, Statistical Inference, Computer Security, Discrete Mathematics, Linear Algebra
- **Honors/Awards:** Brown Undergraduate Research Award, Hack@Brown 2024 Winner, Regeneron STS Semifinalist

## TECHNICAL SKILLS

**Languages:** Python, Java, JavaScript, TypeScript, C/C++, HTML/CSS, SQL, Go, R

**Technologies:** React, Node, Flask, Docker, MongoDB, AWS, PyTorch, MySQL, PostgreSQL, Git, Bash, Linux, Blender

## EXPERIENCE

### Google, Software Engineer Intern, YouTube Team

San Bruno, CA | May 2025 – Aug 2025

- Enhanced the YouTube comments section by migrating translation state management from a global storage to local state in **TypeScript**, improving efficiency and reducing translation response payloads by **53%**
- Built infrastructure for Google DeepMind's new **machine translation** model for YouTube comments in **C++**
- Expanded translation feature to both video and notification comments, supporting design for all **2.5 billion users** using **iOS, Android, and web** clients across all apps including YouTube Main App, YT Studio, and YT Music

### Amazon Robotics, Software Development Engineer Intern, Movement Team

Boston, MA | Jan 2025 – May 2025

- Developed a **React** app to dynamically monitor and configure **10,000+ autonomous robots** in Fulfillment Centers
- Designed new **RESTful APIs** that retrieve and modify robot configurations (e.g. path and speed) in the **DynamoDB** database, including endpoints for real-time data monitoring, comparing robot models, and editing configurations
- Created scalable infrastructure in **Java** using **AWS Lambdas** to call the APIs and **AWS SNS topics** and **SQS queues** to get real-time metadata updates, allowing administrators to retrieve data **62%** faster

### ModernVivo, AI/ML Software Engineer Intern

Seattle, WA | Jan 2024 – Jan 2025

- Created and deployed an LLM-guided system in **Python** that semantically clusters texts from the **PostgreSQL** database with **97%** accuracy, allowing customers to find clinical trials (**Best Paper Runner-up at GLOW @ IJCAI 2024**)
- Utilized **multithreading** to run **OpenAI API** requests in the system concurrently, decreasing clustering times by **55%**
- Incorporated **Google BigQuery** to store large text datasets, and utilized it to test the clustering system

### Brown Visual Computing Lab, Machine Learning Research Intern

Providence, RI | Sep 2022 – Present

- Created a shape editing system in **Python** that lets users augment 3D furniture meshes with natural language by leveraging Chain-of-Thought prompt requests to the **OpenAI API** (**Accepted to SIGGRAPH Asia 2024**)
- Rendered edited meshes into **Blender** with **Python** scripts and a web app via **WebGL** with **JavaScript** for visualization
- Contributed to developing a novel 3D mesh generation method using **Python** and **R** (Published in [biorXiv](#))

## PERSONAL PROJECTS

### MemoTech

*Python, Flask, TypeScript, React, Redis, MongoDB*

- **Won 1st Place out of 300+ Participants at Hack@Brown 2024** for a **React** application that provides detailed feedback for users' answers to flashcards using the **OpenAI API**
- Made flashcard recommendation by storing embeddings in **MongoDB** and creating **Flask** vector search endpoints
- Deployed database onto **AWS EC2** Instance through MongoDB Atlas, supporting backend for **200+ users**

### Pokémon GAN (Gotta generate 'em all)

*Python, PyTorch, Spark*

- Created a DCGAN in **PyTorch** with **Python** to generate synthetic Pokémon images with an FID of 126.7
- Employed **Apache Spark** for distributed preprocessing of large Pokémon datasets, enhancing training efficiency

### Brown Puzzlehunt

*React, TypeScript, Python, PostgreSQL*

- Developed a website using **Django** and **React** for Brown's annual puzzlehunt with over **5,000+ participants**
- Set up **RESTful API** hooks for users to submit puzzle answers, which were securely validated on the backend

# PointPipe: A Point-Cloud classification pipeline for Archaeology in Dense Vegetation

Sir Dylan Allotey  
PI

`sir_dylan_allotey@brown.edu`

Charly Castillo  
Co-PI

`charly_castillo@brown.edu`

Parker VanValkenburgh  
Collaborator  
`parker_vanvalkenburgh@brown.edu`

October 14, 2025

## Abstract

Most LiDAR processing in archaeology focuses on identifying low, mounded features, but existing workflows frequently mistake structures for trees when there is substantial standing architecture. Currently, the only effective approach is manual editing. This study seeks to develop PointPipe, a pipeline combining LiDAR data cleaning and tailored deep learning models to reduce vegetation-structure confusion and improve the reliability of point-cloud classification for archaeological sites in the Andes.



Link to the manuscript pdf:

[https://drive.google.com/file/d/1df2E\\_Cr30NIZMcCRXS37q99z833IDpB/view?usp=sharing](https://drive.google.com/file/d/1df2E_Cr30NIZMcCRXS37q99z833IDpB/view?usp=sharing)

## Response Letter

Dear Editor and Reviewers,

We appreciate the constructive feedback on the manuscript, and we have prepared a revision accordingly. Please see below for an itemized list of your feedback and a brief description of how each is addressed in the revised manuscript.

R1: <https://drive.google.com/file/d/1ISRPjqy1I9T0DigziagxplSeTDX8J4nQ/view?usp=sharing>

[R1.1] Lack of explanation of visualization components or novelty in model/dataset methods.

**Response:** We added a section describing three visualization tools: a semantic labeling viewer, an uncertainty heatmap, and a change-detection overlay. These will help archaeologists interpret model outputs interactively, which is not available in standard LiDAR workflows. We also introduce a geometric-prior fusion layer tailored for archaeological scans.

[R1.2] Expand on how the proposed method differs from previous work.

**Response:** PointPipe fuses geometric covariance features (curvature, planarity) with PointNet++ embeddings using a late-fusion design. Unlike prior urban-focused models, ours is trained on irregular, vegetation-heavy archaeological data. This adaptation directly targets false positives from tree trunks and dense foliage.

[R1.3] Uncertain about dataset cleaning and feasibility

**Response:** We now outline a two-stage cleaning workflow using PDAL and CloudCompare. Automatic filtering will handle most vegetation noise, with manual validation on 5–10% of points. The process fits comfortably within the first two weeks of the project timeline.

[R1.5] Lack of detail of datasets

**Response:** We will be using a data set provided by our collaborator as well as open source sets that we can use to train our program on.

R2: <https://drive.google.com/file/d/159DCLS8rQilZkyHOBajbBe2gHLkz6XSw/view?usp=sharing>

[R2.1] Clarify scientific merit and anthropological relevance

**Response:** The visualizations help archaeologists interpret the cleaned models interactively, offering insight into spatial organization and human modification of the landscape.

[R2.2] How does annotated data look

**Response:** Each point will be annotated as one of five classes: vegetation, bare ground, structure, rock, or artifact cluster. We added sample annotation visualizations (Section 5) to demonstrate labeling density and class balance across scenes.

[R2.3] How is proposal different than SOTA methods

**Response:** PointPipe will incorporate a feature that reweights geometric features based on canopy density. This will hopefully improve segmentation in densely vegetated areas.

R3: <https://drive.google.com/file/d/1H4ZI5MLZ1L5TShxX3b6Xr7WH60-UUoED/view?usp=sharing>

[R3.1] Clarify visualization goals related to reducing false positives

**Response:** We have expanded a section that details how confidence-based overlays and heatmaps will help archaeologists identify and correct false positives interactively.

[R3.2] The significance of the proposal could be more clearly articulated

**Response:** We have revised the introduction and conclusion to emphasize how PointPipe will reduce manual cleaning time for archaeologists.

R4: <https://drive.google.com/file/d/1D4iNHARl91stZ6xdUUY52Du-zus5h2jE/view?usp=sharing>

[R4.1] Clarify how the method will advance disciplines involved

**Response:** We have expanded the broader impacts section to specify that PointPipe contributes to 3D vision through denoising for Lidar data and to archaeology by creating an automatic workflow that replaces manual terrain modeling.

[R4.2] Clarify how to prove improvement over baseline statistics

**Response:** We now specify that we will use semantic [3D.net](#) metrics (IoU) to compare PointPipe to standard baselines like PointNet++.

[R4.3] Visualization utilities seem under developed

**Response:** We have added a visualization module to display confidence-weighted classifications and a false positive heat map. This will help archaeologists interpret the model's outputs more intuitively.

R5: <https://drive.google.com/file/d/1-ZpdLU86D32TToQR-ahpOiPQbydLGZHv/view?usp=sharing>

[R5.1] Consider adding ablations and target priors

**Response:** We now include ablation studies to measure the impact of geometric priors on trunk-wall disambiguation and preprocessing steps. This will clarify which components contribute most to improved performance.

[R5.2] Data scarcity and domain may limit generalization

**Response:** We describe a workflow that could be used in other domains, and we plan to use data sets from [semanti3D.net](#) to limit domain shift.

[R5.3] Reproducibility of the pipeline

**Response:** We plan to package all scripts PDAL pipeline, environmental files, and model weights into a reproducible release, ensuring that the workflow can be replicated for future archaeological sites.

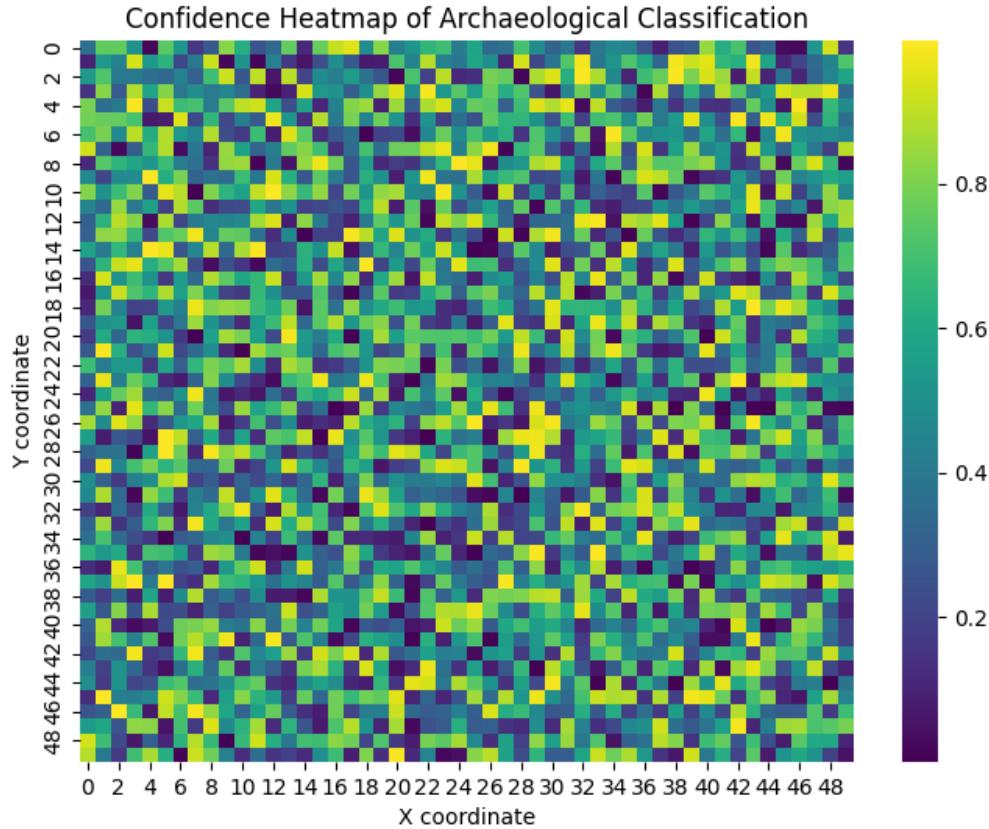


Figure 1: Figure 1 depicting example of Confidence Heatmap

## 1 Aims

1. Develop a reproducible pre-processing pipeline for cleaning raw LiDAR point-clouds from archaeological sites using PDAL and CloudCompare, including automated vegetation noise removal and normalization.
2. Train and evaluate a point-cloud classification model that integrates geometric covariance features with PointNet++ embeddings to reduce misclassifications between vegetation and walls.
3. Compare model performance to baseline semantic3D.statistics using Intersection-over-Union (IoU) and conduct ablation studies to quantify significance of geometric priors.
4. Produce a small reliable dataset with five semantic classes (vegetation, bare ground, structure, rock, artifact cluster) and visualization tools such as uncertainty heatmaps (similar to Figure 1) and confidence overlays to assist archaeological interpretation.
5. Measure expert trust in PointPipe having them complete manual labeling, labeling with PointPipe, and leading point-cloud classification using semantic3D.net.

| Method   | $\overline{IoU}$ | $OA$         | $t[s]$  | $IoU_1$      | $IoU_2$      | $IoU_3$      | $IoU_4$      | $IoU_5$      | $IoU_6$      | $IoU_7$      | $IoU_8$      |
|----------|------------------|--------------|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| TMLC-MSR | <b>0.542</b>     | <b>0.862</b> | 1800    | <b>0.898</b> | <b>0.745</b> | 0.537        | <b>0.268</b> | <b>0.888</b> | <b>0.189</b> | <b>0.364</b> | <b>0.447</b> |
| DeepNet  | 0.437            | 0.772        | 64800   | 0.838        | 0.385        | <b>0.548</b> | 0.085        | 0.841        | 0.151        | 0.223        | 0.423        |
| TML-PCR  | 0.384            | 0.740        | unknown | 0.726        | 0.73         | 0.485        | 0.224        | 0.707        | 0.050        | 0.0*         | 0.15         |

Figure 2: Figure 2 depicting IoU Scores [5]

## 2 Significance

This project addresses an enduring challenge for remote sensing in archaeology: distinguishing man-made structures from dense vegetation in LiDAR data. By introducing a reproducible, geometry-aware workflow, PointPipe reduces manual cleaning and labeling time, making studies that were not feasible before, namely large-scale landscape analysis, possible.

Beyond improving efficiency, PointPipe enhances interpretability and trust in automated classification. The integrated visualization tools such as confidence weighted overlays, uncertainty heatmaps, and change detection maps allow archaeologists to interactively evaluate model predictions and identify potential false positives. This transparency supports informed decision making during planning and site validation, allowing archaeologists to save time, money, and effort during site tours.

Furthermore, the pipeline contributes to 3D vision research by addressing noise and irregular point distributions in natural terrain, a domain often neglected in rural-centric LiDAR studies. The inclusion of geometric priors enables more robust classification under vegetation occlusion, offering insights that can be generalized to forestry, environmental monitoring, and heritage conservation.

Figure 2 compares IOU scores for baseline models across urban and rural environments that have varying degrees of vegetation density. The low score for IoU3 shows a sharp decline in accuracy for densely vegetated areas; this low score underscores the importance of having a specialized model like PointPipe to detect structures in populated areas.

An archaeologist might use PointPipe by loading a LiDAR scan of a densely vegetated Andean site. This system allows interactive inspection of uncertain classifications via heatmaps and confidence overlays, enabling the expert to quickly flag candidate walls or other structures for further study. This workflow reduces the need for labor-intensive manual labeling while improving interpretability and trust in automated classification.

By bridging archaeological field practice and modern 3D computer vision, this project lays the groundwork for reproducible open-source tools that publicizes access to large scale interpretable LiDAR analysis in cultural heritage research.

## 3 Related Work

Point-cloud classification has been studied extensively in computer vision, remote sensing, and archaeology. Early deep learning approaches such as PointNet [7] and its successors demonstrated that point sets

with neural networks can outperform traditional voxel or image-based methods. Building on this, Uy et al. [9] introduced real-world benchmark datasets to evaluate generalization, while Zhang et al. [11] provide a systematic survey of deep learning-based 3D point-cloud methods and future outlooks.

Alternative representations have also been explored. Ben-Shabat et al. [3] proposed 3DmFV, which encodes local point distributions into Fisher Vectors for efficient convolutional processing, achieving real-time classification. More recently, Diab et al. [4] applied deep learning directly to LiDAR data in remote sensing contexts, showing the benefits of combining geometric features with deep models.

Traditional methods still play a role, especially when annotated data is scarce. Rutzinger et al. [8] pioneered object-based analysis of full waveform airborne LiDAR for urban vegetation classification, while Aljumaily et al. [1] demonstrated voxel-based random forest classification with engineered attributes. These methods highlight the importance of feature design when data is noisy or domain-specific.

Within archaeology, airborne LiDAR has opened new possibilities for site detection under dense vegetation. Balsi et al. [2] presented a UAV-borne LiDAR survey for archaeological site mapping, and Vinci et al. [10] offered a systematic review of LiDAR applications in archaeology, emphasizing vegetation filtering challenges. Poux et al. [6] further advanced integration of 3D point-clouds with archaeological knowledge, focusing on quasi-planar structures common in built heritage.

Despite these advances, a consistent limitation persists: false positives in vegetated archaeological contexts. Tree trunks are often misclassified as walls or other human-made structures, particularly when vegetation is dense. Current deep learning approaches [7, 9, 11, 3, 4] are optimized for generic urban or natural scenes, and traditional methods [8, 1] depend heavily on hand-crafted features. Archaeological applications [2, 10, 6] recognize the issue but rely largely on manual verification or site-specific heuristics.

Our contribution, PointPipe, directly addresses this gap by creating a tailored classification pipeline for archaeological sites under vegetation cover. Unlike prior work, we combine LiDAR data cleaning (ground normalization, vegetation-aware preprocessing) with a customized classification model that integrates geometric priors specifically designed to reduce tree trunk-wall confusions. This approach builds on deep learning advances while grounding them in domain specific archaeological constraints, enabling automated labeling at scales not feasible with manual effort.

## 4 Research Plan

### Week 1: Data Acquisition

- Acquire or request LiDAR/ point-cloud datasets of archaeological sites from collaborators and open sources.
- Set up Python development environment and configure PDAL and CloudCompare for preprocessing.

### Week 2: Data Cleaning and Preprocessing Pipeline:

- Implement data cleaning process: noise removal, ground normalization, and vegetation segmentation using PDAL filters.

- Validate cleaning quality through manual inspection of small subset and generate visualizations showing vegetation vs. structure segmentation.
- Milestone - show cleaned dataset and initial visualization that distinguishes vegetation from structures.

### **Week 3: Baseline Model Training**

- Train baseline classifier using cleaned data with PointNet++.
- Evaluate model's ability to distinguish between vegetation, ground, walls, and structures using standard metrics like IoU.
- Evaluate the performance in distinguishing walls versus tree trunks. Milestone - present accuracy metrics of classifications.

### **Week 4: Model Improvement**

- Integrate geometric priors into the classification model and conduct ablation studies to measure their contribution.
- Add visualization utilities such as uncertainty maps and confidence weighted overlays to help archaeologists interpret outputs.

### **Week 5: Evaluation and Error Analysis**

- Conduct testing on multiple sites to assess generalization and domain robustness.
- Analyze error using heatmaps and propose improvements.
- Begin drafting abstract and final report. Milestone - present confusion score, underscoring improvement in detecting archaeological structures.

### **Week 6: Finalization and Presentation**

- Finalize cleaned dataset, model weights, and PDAL pipeline scripts for reproducibility and package them in a markdown file.
- Write the final presentation summarizing contributions, results, and archaeological significance.
- Prepare slides and walkthrough.
- Milestone - Final presentation with a demo of PointPipe.

### **Facilities to Be Used:**

- Computing: Personal machine with CUDA-enabled GPU
- Software: Python, PDAL for LiDAR processing, CloudCompare for manual visualization
- Data: Requested datasets and open-access datasets of archaeological sites

## References

- [1] H. Aljumaily, D. F. Laefer, D. Cuadra, and M. Velasco. Point cloud voxel classification of aerial urban lidar using voxel attributes and random forest approach. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 118, Article 103208, Elsevier, 2023. doi:10.1016/j.jag.2023.103208.
- [2] M. Balsi, S. Esposito, P. Fallavollita, M. G. Melis, and M. Milanese. Preliminary archaeological site survey by uav-borne lidar: A case study. *Remote Sensing*, Vol. 13, No. 3, p. 332, MDPI, 2021. doi:10.3390/rs13030332.
- [3] Y. Ben-Shabat, M. Lindenbaum, and A. Fischer. 3dmfv: Three-dimensional point cloud classification in real-time using convolutional neural networks. *IEEE Robotics and Automation Letters*, Vol. 3, No. 4, pp. 3145–3152, IEEE, 2018. doi:10.1109/LRA.2018.2850061.
- [4] A. Diab, R. Kashef, and A. Shaker. Deep learning for lidar point cloud classification in remote sensing. *Sensors*, Vol. 22, No. 20, p. 7868, MDPI, 2022. doi:10.3390/s22207868.
- [5] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys. Semantic3d.net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*, ETH Zurich, 2017. URL: <http://arxiv.org/abs/1704.03847>.
- [6] F. Poux, R. Neuville, L. V. Wersch, G.-A. Nys, and R. Billen. 3d point clouds in archaeology: Advances in acquisition, processing and knowledge integration applied to quasi-planar objects. *Geosciences*, Vol. 7, No. 4, p. 96, MDPI, 2017. doi:10.3390/geosciences7040096.
- [7] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *arXiv preprint arXiv:1704.03847*, Stanford University, 2017. <https://arxiv.org/abs/1704.03847>.
- [8] M. Rutzinger, B. Höfle, M. Hollaus, and N. Pfeifer. Object-based point cloud analysis of full-waveform airborne laser scanning data for urban vegetation classification. *Sensors*, Vol. 8, No. 8, pp. 4505–4528, MDPI, 2008. doi:10.3390/s8084505.
- [9] M. A. Uy, Q.-H. Pham, B.-S. Hua, T. Nguyen, and S.-K. Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, 2019.
- [10] G. Vinci, F. Vanzani, A. Fontana, and S. Campana. Lidar applications in archaeology: A systematic review. *Archaeological Prospection*, Vol. 32, No. 1, pp. 81–101, Wiley, 2025. doi:10.1002/arp.1931.
- [11] H. Zhang, C. Wang, S. Tian, B. Lu, L. Zhang, X. Ning, and X. Bai. Deep learning-based 3d point cloud classification: A systematic survey and outlook. *Displays*, Volume 79, Article 102456, Elsevier, 2023. doi:10.1016/j.displa.2023.102456.

# SIR DYLAN ALLOTEY

Bronx, New York, 10453 | [sirdylan.allovey1@gmail.com](mailto:sirdylan.allovey1@gmail.com) | (646) 773-9172 | [LinkedIn](#)

## EDUCATION

**Brown University**, Providence, RI **Expected Graduation May 2026**  
*Master of Science in Computer Science*  
**Brown University**, Providence, RI **Graduated May 2025**  
*Bachelor of Science in Computer Science* GPA: 3.75  
Honors: Sydney Frank Scholarship (2021), Zeta Boule Scholarship (2021)

## PROFESSIONAL EXPERIENCE & LEADERSHIP

**Modern Railway Systems**, Littleton, CO; Windsor, CA; Los Angeles, CA **May 2023 – Present**  
*Systems/Software Engineering Intern*

- **Fall 2025** - Maintained backend systems for applications managing live trains
  - Designed and optimized database update mechanisms
  - Implemented access controls and safety restrictions
- **Summer 2025** – Contributed to multibillion-dollar new rail infrastructure project for LA Metro
  - Performed duct bank cross-section calculations
  - Coordinated with contractors, stakeholders, and multidisciplinary teams to support engineering tasks
- **Summer 2024** - Used data collections to design and test hardware systems for use in the field
  - Developed software in Python to model aspects of vehicles and city plans such as approach velocity and warning time
  - Designed and wired several routers and switches to achieve redundancy and cyclic connections in the network
- **Summer 2023** - Worked with engineering teams to model several road conditions to create safe track speed restrictions for trains using kinematics
  - Designed router systems and CAD drawings to test and improve the current system's redundancy
  - Developed software in Python to represent data including time, costs, and resources a project

## PROJECTS

**Ray Tracer** **December 2023**  
*Developer*

- Designed ray tracer to build 3d scenes like Maya
- Rendered navigable scenes using primitive shapes (cone, cylinder, cube, sphere)
- Implemented Phone lighting reflection model
- Coded in C++ using the QT creator framework and the Open Graphics Library

**Computer Shell** **December 2023**  
*Developer*

- Coding in C with Ubuntu and Docker
- Featured memory allocation, extensive error-checking, background processes, and signal handling

## SKILLS

**Computer:** Proficient in Microsoft Applications (Excel, Word, & PowerPoint), Intermediate in Blender and Maya, Building Computers

**Languages:** Basic in Korean

**Coding Languages:** Korean Basic (Writing, reading, and speaking), Proficient in Java, Python, C, C++



# CHARLY CASTILLO

Weehawken, NJ | (201) 682-5540 | [charly\\_castillo@brown.edu](mailto:charly_castillo@brown.edu) | [LinkedIn](#)

## EDUCATION

**Brown University**, Providence, RI

Expected Graduation 05/2026

- *Sc.B. in Statistics and International and Public Affairs, Security*
- **D.E. Shaw Discovery Fellow:** 1 of 49 sophomores nationwide selected for all-expenses paid program at DESCO HQ
- **Undergraduate Teaching and Research Award:** Fall '23; 1 of ~100 Brown students awarded \$1,200 research stipend
- **Relevant Coursework:** *Undergraduate:* Stochastic Calculus, Computational Probability, Modeling, Optimization, Theoretical Linear Algebra, Scientific Computing, Regression; *Graduate:* Scientific Visualization, Cybersecurity

## WORK EXPERIENCE

**Bank of America**, *Quantitative Analyst Intern*, New York, NY

06/2025–08/2025

- Evaluated 2024 policy replacing aging fees with desk-specific aged inventory limits, applying ITS analysis to measure shifts across BofA's 6 Global Credit flow desks (ASIA, CEEMEA, EMEA, LATAM, US HY, US IG).
- Found 15-37% reductions on EMEA and CEEMEA desks; normalized aged inventory by book size across 12 US IG traders using coefficients of variation; proposed orthogonal random forest to isolate overlapping policy effects.

**JPMorgan Chase & Co.**, *Investment Banking Summer Analyst*, New York, NY

06/2024–07/2024

- Analyzed \$11B cross-border merger, assessing strategic fit by through revenue segmentation and market scope.

**Brown University Department of Biostatistics**, *Research Intern*, Providence, RI

09/2023–12/2023

- Benchmarked 46 distance and similarity metrics using philentropy to evaluate accuracy in matching multi-part names across datasets without unique identifiers; simulated 10 data corruption scenarios (substitutions, insertions, deletions, transpositions) on 1,000+ records using stringr and purrr to assess robustness under realistic data issues.

**New Jersey Cybersecurity and Communications Integration Cell**, *SECOPS Intern*, Newark, NJ

06/2023–07/2023

- Developed logistic regression pipeline with scikit-learn to classify 10,000+ algorithmically generated domains (AGDs), engineering 10 predictive linguistic features (e.g., consonant length, domain length, vowel/consonant ratio).
- Achieved 64% precision in correctly identifying AGDs; improved model performance by 8% (F1 score) through hyperparameter tuning with GridSearchCV; benchmarked against Random Forest and Gradient Boosting classifiers.

## LEADERSHIP EXPERIENCE

**Brown Office of Undergraduate Admissions**, *Student Head*, Providence, RI

02/2024–Present

- Lead campus tour program and conduct 3 weekly campus tours for audiences of up to 180; represent Brown as a keynote speaker at ADOCH, the official [virtual tour](#), and on Brown's official Instagram account (25k+ followers).

**Brown Statistics Department Undergraduate Group**, *President*, Providence, RI

10/2023–Present

- Organize academic events (e.g., alumni panels, course registration Q&As) attended by 200+ students annually; manage \$1,500 budget and negotiate program funding through UFunds to expand department programming.

## PROJECTS

**Total Variation Image Denoising via ADMM**, *Class Project*

04/2025–05/2025

- Applied ADMM with Gauss-Seidel iterations to Rudin-Osher-Fatemi total variation denoising; evaluated convergence (L2 norms, iteration counts) and output quality across eight regularization parameters on grayscale images.

**Estimating U.S. Inflation Divergence from G7 Economies**, *Independent Project*

05/2024–09/2024

- Analyzed U.S. inflation trends using Difference-in-Differences, Bayesian Structural Time Series (CausalImpact), and Synthetic Control (Synth); processed IMF data in R using fixed effects regressions and counterfactual simulations.
- Estimated inflationary effect of 30-140 basis points attributable to U.S. policies, including global spillover effects.

## SKILLS

- *Technical Skills:* Python (pandas, NumPy, scikit-learn, statsmodels, JAX, SciPy, Plotly), R, MATLAB, Julia, SQL
- *Tools:* IBM SPSS Statistics, Excel (PivotTables, VBA), Bloomberg, Git, LaTeX, Google Workspace
- *Languages:* English (native), Castilian Spanish (native), Galician (native), French (intermediate), Portuguese (basic)

## NATHANIEL PARKER VANVALKENBURGH

128 Hope St., Providence, RI / parker\_vanvalkenburgh@brown.edu / Phone (401) 863-3251

---

### ACADEMIC EMPLOYMENT

- 2015- **Brown University**, Providence, RI  
Associate Professor of Anthropology, 2022-  
Elected Fellow, Institute at Brown for Environment and Society, 2018-  
Faculty Fellow, Joukowsky Institute for Archaeology and the Ancient World, 2022-  
Stanley J. Bernstein Assistant Professor of Social Sciences, 2020-22  
Assistant Professor of Anthropology, 2015-2022
- 2013-15 **University of Vermont**, Burlington, VT  
Assistant Professor of Anthropology
- 2012-13 **Washington University in St. Louis**, St. Louis, MO  
Andrew Mellon Postdoctoral Fellow in Modeling Interdisciplinary Inquiry  
Research Associate, Department of Anthropology

### EDUCATION

- 2012 **Harvard University**, Cambridge, MA  
Ph.D., Anthropology
- 2005 **Institute of Latin American Studies, University of London**, UK  
M.A., Latin American Studies
- 2004 **University of Cambridge**, Cambridge, UK  
M.Phil., Archaeology, with Distinction
- 2003 **Stanford University**  
B.A., Anthropological Sciences; Honors, with Distinction

### ADDITIONAL EDUCATION

- 2013 **Washington U. in St. Louis** - Advanced GIS; Database Design, Spatial Modeling
- 2009 **Mellon Summer Seminar in Spanish Paleography**, Pasadena, CA

### PRIMARY RESEARCH PROJECTS

- 2017- **Principal Investigator**, Paisajes Arqueológicos de Chachapoyas (PACHa)  
Regional archaeological project focused on long term histories of (Inka and Spanish) colonialism in the Middle Utcubamba Valley. Survey conducted in 2017; excavations in Choctamal (La Fila, Llaucán, and Changalí), lidar survey at Kuelap in 2018; excavations in Purun Llaqta del Maino, 2019. Collaborators include Zachary Ahonen-Critchley, Matthias Kucera, Megann Phillips, Daniel Plekhov, Alexis Reátegui Diaz, Sophie Reilly, Carol Rojas Vega, J. Marla Toyne, and Bethany Whitlock.

- 2016- **Co-Principal Investigator**, GeoPACHA: Geospatial Platform for Andean Colonial History and Archaeology  
Development of a browser-based, edited geospatial platform for discovering and mapping archaeological sites in the Andean region; designed to facilitate the identification of archaeological sites through “virtual survey” of satellite and aerial imagery. Co-PI - Steven Wernke (Vanderbilt University).
- 2019- **Co-Principal Investigator**, Mapping Historical Trauma in Tulsa, 1921-2021  
Collaborative, community-based research project on the Tulsa Race Massacre of 1921 and its legacies, including geophysical survey, excavations, digital storytelling and museum exhibition. Also includes development of Greenwood Historical GIS, a series of resources for visualizing the effects and experiences of the Tulsa Massacre. Co-PI - Alicia Odewale (University of Tulsa).
- 2008-17 **Principal Investigator**, Proyecto Arqueológico Zaña Colonial (PAZC)  
Regional survey, excavation, and archival project examining Indigenous experience and landscapes in Peru’s Zaña and Chamán valleys. Regional survey, 2008-10; From 2012-2016, excavations concentrated at the site of Carrizales, focusing on the effects of forced resettlement and environmental change on domestic economies and residential life. Served as umbrella for series of additional subprojects, including chemical characterization of ceramics, collaboration with Federated Archaeological Information Management System, isotopic studies. Collaborators included Katherine Chiou, Natalia Guzmán Requena, Sarah Kelloway, Sarah Kennedy, Alexis Reátegui Diaz, Carol Rojas Vega, Rocío Torres Mora, Chester Walker, and Daniela Zevallos Castañeda.
- 2011-15 **Principal Investigator**, Archaeometry of Early Green Glazed (EGG) Ware  
Investigation of glazed vessels made by Indigenous potters on Peru’s north coast in 16<sup>th</sup> century. Thin section petrography, Instrumental Neutron Activation Analysis (INAA), Laser-Ablation Inductively Coupled Mass Spectrometry (LA-ICP-MS), and re-firing studies using Scanning Electron Microscopy (SEM). With Laure Dussubieux (Field Museum), Sarah Kelloway (University of New South Wales), Jeffrey Quilter (Harvard), Bill Sillar (UCL), Karen Privat (UNSW), and Michael Glascock (MURR).

#### OTHER RESEARCH

- 2022-23 **Co-Organizer**, Deep Displacement: Excavating Histories of Migration and Settlement  
Planned conference in Spring 2022, focusing on comparative approaches to displacement, migration and settlement in archaeology, leading to edited volume. Co-organized with Dr. Melissa Rosenzweig (Northwestern University). Preparing Workshop Grant resubmission for Wenner-Gren Foundation for Anthropological Research, May 1, 2021.
- 2019-22 **Steering Committee**, FAIMS 3.0 Electronic Field Notebooks: A Platform for Field Research Supporting Digital Data Capture and Management in Diverse Situations  
Software development program for mobile digital recording platforms for archaeological research. Funded by Australian Research Data Commons. As steering committee member, provide design input and testing.

- 2016- **Project Member**, Colonial Modernity in the Andes: A Comprehensive Study of Viceroy Toledo's General Resettlement
- Collaborative research project on Indigenous forced resettlement in colonial Peru, linked to two books and several conferences. Japan Society for the Promotion of Science KAKENHI Grant Number 15H01911. PI Akira Saito, National Museum of Ethnology, Japan. Have participated in three workshops and contributed two research papers.
- 2021-23 **Partner Investigator**, Archaeology, Material Culture and Identity in Colonial Peru
- Research project focused on the study of transformations in material culture change (primarily ceramics) at a series of distinct Peruvian sites. Chief Investigators - Sarah Kelloway, James Flexner, and Fernanda Peñaloza (University of Sydney). Proposal currently under review with Australian Research council. Would contribute previously excavated samples for analysis, assist in interpretations, and contribute to publications.

#### EXTERNAL RESEARCH GRANTS

##### **For Paisajes Arqueológicos de Chachapoyas (PACHa)**

- 09/2021 "Collaborative Research: Adaptive Strategies under Empire Transitions" - NSF Senior
- 08/2025 Archaeology Award 2114106, co-PI J. Marla Toyne, University of Central Florida
- 04/2020 "Imperial Transitions in the Upper Amazon Basin: Forced Resettlement, Diet, and Political
- 09/2022 Subjectivity at Purun Llaqta del Maino, Peru" - American Philosophical Society Franklin Research Grant
- 05/2018 "'Warriors Above the Clouds': Exploring Settlement History and Landscape Change in
- 10/2018 Peru's Chachapoyas Region Using Airborne LiDAR and Photogrammetry." National Geographic Committee for Research and Exploration (grant HJ-044R-17)
- 2017 "Employing TanDEM-X to Analyze Settlement History and Landscape Change in Peru's Chachapoyas Region" - TanDEM-X Science Grant
- 2016 "Settlement History in Peru's Chachapoyas Region" - Digital Globe Foundation Grant

##### **For Geospatial Platform for Andean Culture, History, and Archaeology (GeoPACHA)**

- 10/2021 "Collaborative Research: RI: Medium: Deep Generative Modeling for Urban and
- 09/2024 Archaeological Recovery" (IIS - Robust Intelligence Grant 2106766) PI - Daniel Aliaga, Purdue University; co-PIs Rajesh Salnayam, Purdue University; Ian Lindsay, Purdue University; Steven Wernke, Vanderbilt University
- 01/2018 "Extending GeoPACHA: Geospatial Platform for Andean Culture, History, and
- 06/2021 Archaeology" - ACLS Digital Extension Grant, w/ Steven Wernke
- 01/2017 "Developing GeoPACHA: Geospatial Platform for Andean Culture, History, and
- 12/2018 Archaeology" - Spatial Archaeometry Research Collaborations (SPARC) grant - University of Arkansas, with Steven Wernke

##### **For Mapping Historical Trauma in Tulsa, 1921-2021**

- 6/2022 "Archaeology of Greenwood: Bearing Witness to a Century of Resilience," Co-PI with Dr.
- 6/2024 Alicia Odewale, National Geographic Azimuth Grant NGS-94600R-22

01/2019 "Mapping Historical Trauma in Tulsa, 1921-2021" - Tulsa Race Massacre Centennial  
08/2021 Commission, with Alicia Odewale

**For Deep Displacement Conference**

01/2022 "Deep Displacement: Excavating Histories of Migration and Settlement" - Wenner Gren  
06/2022 Foundation Workshop Grant, with Melissa Rosenzweig (Northwestern University)

**For Proyecto Arqueológico Zaña Colonial (PAZC)**

10/2014 "Imperial Households: Forced Resettlement, Diet and Daily Life at Carrizales, Peru."  
03/2018 National Endowment for the Humanities Collaborative Research (grant RZ-51748-14)  
07/2013- "The Archaeology of Forced Resettlement and Daily Life at Carrizales and Conjunto 131,  
08/2014 Zana, Valley, Peru." National Geographic Committee for Research and Exploration (grant  
9334-13)  
04/2012 "Reducción and Policía: Spanish Colonial Forced Resettlement and Daily Praxis at  
12/2012 Carrizales (Zaña Valley, Peru)" Wenner-Gren Post-Ph.D. Research Grant & Osmundsen  
Supplement  
2010 Program for Cultural Cooperation (Spain's Ministry of Culture)  
2009 International Dissertation Research Fellowship, SSRC

**For Archaeometry of Colonial Chimu-Inka Glazed Ceramics (EGG Ware)**

2014 "A SEM microstructural investigation of Peruvian EGG ware: exploring potting practices  
in 16th Century Peru Wainwright Analytical Centre Research Grant, U. of New South  
Wales, with Sarah Kelloway and Karen Privat  
2014 "Characterization of Early Green Glazed Wares and Tin-Enameled Wares from Colonial  
Peru, Using Instrumental Neutron Activation Analysis" - National Science Foundation  
Subvention, U. Missouri Research Reactor  
2010 "LA-ICP-MS analysis of Chimu-Inka glazed ware," NSF Subvention Field Museum  
Analytical Research Facility

**INTERNAL RESEARCH GRANTS**

2019 Summer Undergraduate Teaching and Research Award (UTRA x 3)  
2020 Richard B. Salomon Faculty Research Award  
2020 Institute at Brown for Environment and Society Seed Grant, with Stephen Porder  
2019 Faculty Development Funding, Brown University Dean of Faculty  
2019 Social Science Research Institute Seed Grant  
2019 Summer Undergraduate Teaching and Research Award (UTRA x 2)  
2018 Interdisciplinary Team UTRA Grant (3 students)  
2017 Social Science Research Institute Seed Grant

- 2016 Brown University Interdisciplinary Team UTRA Grant (4 students)
- 2009 Frederick Sheldon Traveling Fellowship, Harvard University
- 2008 Summer Research Travel Grant, David Rockefeller Center, Harvard University
- 2006 Graduate Society Summer Research Grant, Harvard University

#### FELLOWSHIPS AND AWARDS

- 2021 William G. McLoughlin Award for Excellence in Teaching in the Social Sciences (Brown)
- 2017 American Council of Learned Societies Fellowship
- 2017 Visiting Fellow, McDonald Institute of Archaeology, University of Cambridge
- 2017 Cogut Center for the Humanities Faculty Fellowship (deferred to 2018-19)
- 2016 Haffenreffer Museum of Anthropology Faculty Fellow
- 2014 Kroepsch-Maurice Excellence in Teaching Award (Nominee), UVM
- 2012 Harvard Extension School Student Choice Teaching Award Nominee
- 2010 Dumbarton Oaks Junior Fellowship
- 2010 Cora DuBois Dissertation Completion Fellowship
- 2008 Harvard University Certificate of Distinction in Teaching
- 2007 Harvard University Certificate of Distinction in Teaching
- 2008 Graduate Society Merit Fellowship, Harvard University
- 2008 Jens Aubrey Westengard Scholarship, Harvard University
- 2004 Fellow, Cambridge Overseas Trust Society
- 2003 Phi Beta Kappa, Stanford University
- 2003 Firestone Medal for Undergraduate Research, Stanford University
- 2003 British Marshall Scholarship

#### WORK UNDER REVIEW OR CONTRACT

- Under Contract* **VanValkenburgh, Parker** and Steven A. Wernke (joint first authors). *The Archaeology of Spanish Colonialism in the Central Andes*. Cambridge University Press World Archaeology Series.
- Under Review* Smit, Douglas K. and **Parker VanValkenburgh**. Indigenous labor and the Circulation of Majolicas in the Colonial Andes. Submitted to *Journal of Archaeological Science: Reports*
- Under Review* Jiachen Xu; James Zimmer-Dauphinee; Quan Liu; Yuxuan Shi; Junlin Guo; **Parker VanValkenburgh**; Steven A. Wernke; Yuankai Huo. Submitted to *International Journal of Remote Sensing*.
- Under* Morrow, Giles S; Steven Wernke; **Parker VanValkenburgh**., Christopher Wai. "Seeking

- Review* Refugia on the North Coast of Peru: Expanding the 'Total Coverage' Archaeological Surveys of the Jequetepeque and Lower Santa Valleys through systematic analysis of satellite imagery using the GeoPACHA." Submitted to *Antiquity*.
- Under Review* Wernke, Steven A; **Parker VanValkenburgh**; et al. "GeoPACHA: An online platform for collaborative imagery-based survey in the Andes" Submitted to *Antiquity*.
- Under Review* Marcone, Giancarlo; Steven Wernke; James Zimmer-Dauphinee; **Parker VanValkenburgh** and Justin Moat "Lomas in the mist: Late pre-Hispanic Fog oasis settlements and the long-term human occupations on the Peruvian Central Coast from satellite imagery"

## BOOKS

- 2021 **VanValkenburgh, Parker**. *Alluvium and Empire: The Archaeology of Colonial Resettlement and Indigenous Persistence on Peru's North Coast*. Tucson: University of Arizona Press.
- 2013 Osborne, James F. and **Parker VanValkenburgh**, eds. *Territoriality in Archaeology*. Archaeological Papers of the American Anthropological Association 22. Washington, D.C.

## THEMED JOURNAL ISSUES

- 2020 VanValkenburgh, Parker and Andrew Dufton. Archaeological Vision in the Age of Big Data Themed Issue of *Journal of Field Archaeology*. Volume 45, Supplement 1.
- 2016 Traslaviña, T. Abel, Chase, Zachary, **VanValkenburgh, Parker** and Brendan J.M. Weaver, eds. *Arqueología Histórica en el Perú*. Boletín de Arqueología PUCP. Volumes 20 y 21.

## PEER-REVIEWED JOURNAL ARTICLES

- 2021 Plekhov, Daniel, **Parker VanValkenburgh**, Paul Abrams, Amanda Cutler, Justin Han, Alexis Reátegui Díaz, Bryn Sullivan, Steven Wernke. "Linking Past and Present Land-use Histories in Southern Amazonas, Peru." *Remote Sensing* accepted / in press.
- 2020 **VanValkenburgh, Parker** and Andrew Dufton. "Big Archaeology: Horizons and Blindspots." *Journal of Field Archaeology* 45: S1-S7. [10.1080/00934690.2020.1714307](https://doi.org/10.1080/00934690.2020.1714307)
- 2020 **VanValkenburgh, Parker**, K. C. Cushman, Luis Jaime Castillo Butters, Carol Rojas Vega, Carson Roberts, Charles Kepler and James Kellner. "Lasers Without Lost Cities: Using Drone-mounted LiDAR to Capture Architectural Complexity at Kuelap, Amazonas, Peru" *Journal of Field Archaeology* 45: S75-S88 [10.1080/00934690.2020.1713287](https://doi.org/10.1080/00934690.2020.1713287)
- 2020 Wernke, Steven, **VanValkenburgh, Parker** and Akira Saito. "Interregional Archaeology in the Age of Big Data: Building Online Collaborative Platforms for Virtual Survey in the Andes" *Journal of Field Archaeology* 45: S61-S74. [10.1080/00934690.2020.1713286](https://doi.org/10.1080/00934690.2020.1713286)
- 2019 Kelloway, Sarah; **VanValkenburgh, Parker**; Astuhamán González, César; González Lombardi, Andrea; Bedoya Vidal, Diego. "International Pots of Mystery: using PXRF to identify the provenance of *botijas* from 16<sup>th</sup> Century sites on Peru's north coast." *Journal of Archaeological Science Reports* 27 [10.1016/j.jasrep.2019.101974](https://doi.org/10.1016/j.jasrep.2019.101974)

- 2019 **VanValkenburgh, Parker.** "The Past, Present and Future of Transconquest Archaeology in the Andes." *International Journal of Historical Archaeology* 23 (4): 1063–1080 [10.1007/s10761-018-0484-z](https://doi.org/10.1007/s10761-018-0484-z)
- 2019 Kennedy, Sarah, Chiou, Katherine and **Parker VanValkenburgh.** "Inside the Reducción: Crafting Colonial Foodways at Carrizales and Mocupe Viejo, Zaña Valley, Peru (1570-1700) *International Journal of Historical Archaeology* 23(4): 980–1010 [10.1007/s10761-018-0481-2](https://doi.org/10.1007/s10761-018-0481-2)
- 2019 Turner, Bethany; **VanValkenburgh, Parker;** Lee, Kristina; and Benjamin Schaefer. "Paleodiet among Late Prehispanic and Early Colonial Human Remains from Carrizales, Zaña Valley, Peru." *International Journal of Osteoarchaeology* 29(4): 560-573. [10.1002/oa.2752](https://doi.org/10.1002/oa.2752)
- 2018 **VanValkenburgh, Parker;** Silva, Luiza O. G.; Repetti-Ludlow, Chiara; Crook, Jackson; Gardner, Jake; Ballsun-Stanton, Brian. 2018. "Mobilization as Mediation: implementing a tablet-based recording system for ceramic classification." *Advances in Archaeological Practice* 6(4): 342-356 [10.1017/aap.2018.12](https://doi.org/10.1017/aap.2018.12)
- 2018 Kelloway, Sarah; **VanValkenburgh, Parker (joint first author);** Iñáñez, Javier; Dussubieux, Laure; Quilter, Jeffrey; and Michael D. Glascock. 2018. "Identifying Andean-made Majolica from 16th-18th Century Sites on Peru's North Coast." *Journal of Archaeological Science: Reports* 17: 311–324. [10.1016/j.jasrep.2017.10.008](https://doi.org/10.1016/j.jasrep.2017.10.008)
- 2017 **VanValkenburgh, Parker,** Kelloway, Sarah, Privat, Karen, Sillar, Bill, and Jeffrey Quilter. Rethinking Cultural Hybridity and Technology Transfer: SEM Microstructural Analysis of Lead Glazed Ceramics from Early Colonial Peru. *Journal of Archaeological Science* 82: 17-30. [10.1016/j.jas.2017.04.007](https://doi.org/10.1016/j.jas.2017.04.007)
- 2017 **VanValkenburgh, Parker.** Unsettling Time: Persistence and Memory in Spanish Colonial Peru. *Journal of Archaeological Method and Theory* 24(1): 117-148. [10.1007/s10816-017-9321-7](https://doi.org/10.1007/s10816-017-9321-7)
- 2016 Weaver, B.J.M., Traslaviña, A., **VanValkenburgh, P.** and Z. Chase. "Arqueología histórica en el Perú: La sociedad andina en la transición económica, política y social." *Boletín de Arqueología PUCP* 21: 5-12 [10.18800/boletindearqueologiapucp.201602.001](https://doi.org/10.18800/boletindearqueologiapucp.201602.001)
- 2016 **VanValkenburgh, Parker.** "Produciendo Chérrepe: reducción, etnia y performance en los valles de Zaña y Chamán, siglos 16 y 17" *Boletín de Arqueología PUCP* 20: 139-158. [10.18800/boletindearqueologiapucp.201601.007](https://doi.org/10.18800/boletindearqueologiapucp.201601.007)
- 2016 **VanValkenburgh, Parker,** Chase, Zachary J., Weaver, Brendan J.M. and Teddy Abel Traslaviña Arias. "Arqueología Histórica en el Perú: Posibilidades y Perspectivas." *Boletín de Arqueología PUCP* 20: 5-24. [10.18800/boletindearqueologiapucp.201601.001](https://doi.org/10.18800/boletindearqueologiapucp.201601.001)
- 2016 Kennedy, Sarah and **Parker VanValkenburgh.** "Zooarchaeology and Changing Food Practices at Carrizales, Peru." *International Journal of Historical Archaeology* 20(1): 73-104. [10.1007/s10761-015-0319-0](https://doi.org/10.1007/s10761-015-0319-0)
- 2016 Kelloway, Sarah J., Ferguson, Timothy Iñáñez, Javier G., **VanValkenburgh, Parker,** Roush, Cody C., Gibbs, Martin, and Michael D. Glascock. "Sherds on the Edge: Characterisation of 16<sup>th</sup> Century Colonial Spanish Pottery Recovered from the Solomon Islands." *Archaeometry* 58(4): 549–573. [10.1111/arc.12187](https://doi.org/10.1111/arc.12187)



- 2015 **VanValkenburgh, Parker**, Kelloway, Sarah, Laure Dussubieux, Quilter, Jeffrey, and Michael Glascock. 2015. "The Production and Circulation of Indigenous Lead-Glazed Ceramics in Northern Peru During Spanish Colonial Times." *Journal of Archaeological Science* 61: 172-185 [10.1016/j.jas.2015.06.006](https://doi.org/10.1016/j.jas.2015.06.006)
- 2015 **VanValkenburgh, Parker**, Chester P. Walker, and Jennie O. Sturm. 2015. "Gradiometer and Ground-Penetrating Radar Survey of two *Reducción* Settlements in the Zaña Valley, Peru" *Archaeological Prospection* 22 (2): 117-129. [10.1002/arp.1499](https://doi.org/10.1002/arp.1499)
- 2013 **VanValkenburgh, Parker**. 2013. "Creolization, Hybridity and Mestizaje: A Comment." *Archaeological Review from Cambridge* 28: 301-322. [LINK](#)
- 2013 **VanValkenburgh, Parker** and James F. Osborne. 2013. "Home Turf: Archaeology, Territoriality, and Politics." *Territoriality in Archaeology*, ed. Osborne, James F. and Parker VanValkenburgh, eds. *Archaeological Papers of the American Anthropological Association* 22: 1-27. [10.1111/apaa.12000](https://doi.org/10.1111/apaa.12000)

#### PEER-REVIEWED BOOK CHAPTERS

- Subm.* **VanValkenburgh, Parker**. Submitted. "The Houses and House-lives of *Indios Reducidos*: Domestic Space, Food Sharing, and the Politics of Dwelling at Carrizales (Zaña Valley, Peru)," in Akira Saito and Steven Wernke, eds. *Unsettling Resettlement*
- Subm.* **VanValkenburgh, Parker**. Submitted. "Seeing Through the Trees: Lidar Landscapes and the Visual Culture of Archaeology" in Godfrey, Anne C., ed., *Active Landscape Photography: Diverse Practices* (Routledge)
- 2021 **VanValkenburgh, Parker**. "Colonial Ceramics: Commerce and Consumption," in Jeffrey Quilter, ed. *Magdalena de Cao Viejo: An Early Colonial Town on the North Coast of Peru*. Cambridge, MA: Peabody Museum Press. Pp. 209-229.
- 2020 **VanValkenburgh, Parker**, Kennedy, Sarah, Rojas Vega, Carol and Gabriel Hassler. "El Contrato del Mar: Colonial Life and Maritime Subsistence at Carrizales, Zaña Valley, Peru." Gabriel and Daniel Sandweiss, eds. *Maritime Communities of the Ancient Andes*. Gainesville, FL: University Press of Florida. Pp. 366-396.
- 2018 **VanValkenburgh, Parker**. "Whither the History of Forced and Undocumented Migration? Notes for Genealogical and Comparative Approaches." In Hamilakis, Y. (ed.). *The New Nomadic Age: Archaeologies of Forced and Undocumented Migration*. Sheffield: Equinox. Pp. 231-238. [DOI: 10.1558/equinox.34658.Nov.2018](https://doi.org/10.1558/equinox.34658.Nov.2018)
- 2017 **VanValkenburgh, Parker**. "Historias galonadas: la genealogía de reducción, paisaje y población en los valles de Zaña y Chamán, Perú," en Saito, Akira y Claudia Rosas Lauro, eds. *Reducciones: La Concentración Forzada de las Poblaciones Indígenas en el Virreinato del Perú*. Lima: Fondo Editorial PUCP. Pp. 223-260. [LINK](#)
- 2016 Sobotkova, A., Ross, S., Ballsun-Stanton, B., Fairbairn, A., Thompson, J., and **VanValkenburgh, P.** Measure twice, cut once: cooperative deployment of a generalised, archaeology-specific field data collection system. In E. W. Averett, J. M. Gordon, & D. B. Counts (Eds.), *Mobilizing the Past: Recent Approaches to Archaeological Fieldwork in the Digital*

Age University of North Dakota Digital Press in conjunction with Mukurtu 2.0 (Center for Digital Archaeology). Pp. 337-374. [LINK](#)

#### ARTICLES IN NON-PEER REVIEWED VENUES

- 2020 Sarmiento, Fausto O., M.B. Bush, W. Church, **P. VanValkenburgh**, M. Oliva, E. Delgado, S. Fernandez and N. Rojas. 2020. "Mountain science helps ecotourism in the cloud forests of Peru." PAGES Magazine. 28(1), 22-23. DOI [10.22498/pages.28.1.22](#)
- 2018 **VanValkenburgh, Parker**; Sarah Kelloway, Daniela Zevallos Castañeda y Diego Bedoya Vidal. 2018. "Caracterización química de cerámica colonial temprana en el sitio de Carrizales," Lambayeque, Perú. Actas del III Congreso Nacional de Arqueología (Lima, Peru).
- 2017 Marwick, B. d'Alpoim Guedes, J.; Barton, C.M.; Bates, L.A.; Baxter, M.; Bevan, A.; Bollwerk, E.A.; Bocinsky, R.K.; Brughmans, T.; Carter, A.; Conrad, C.; Contreras, D.A.; Costa, S.; Crema, E.R.; Daggett, A.; Davies, B.; Drake, B.L.; Dye, T.S.; France, P.; Fullagar, R.; Giusti, D.; Graham, S.; Harris, M.D.; Haws, J.; Heath, S.; Huffer, D.; Kansa, E.C.; Kansa, S.W.; Madsen, M.E.; Melcher, J. Negre, Joah; Neiman, F.D.; Opitz, R.; Orton, D.C.; Przstupa, P.; Raviele, M. Riel-Salvatore, J.; Riris, P.; Romanoska, I. Smith, J.; Strupler, N.; Ullah, I.I., Van Vlack, H.G.; **VanValkenburgh, N.**; Watrall, E.E.; Webster, J.W.; Winters, J. and C.D. Wren. 2017. "Open Science in Archaeology." SAA *Archaeological Record* 17: 8-14. [DOI 10.17605/OSF.IO/3D6XX](#)
- 2013 **VanValkenburgh, Parker**, Carol Rojas Vega and Rocío Torres Mora. 2013. "Arqueología Histórica en los Valles de Zaña y Chamán." *Revista Argumentos*, año 7, n° 3. Lima: IEP.

#### BOOK REVIEWS

- 2014 VanValkenburgh, Parker. Review of *The Archaeology of Hybrid Material Culture*, ed. Jeb J. Card. *American Antiquity* 79(4): 802-803.
- 2014 VanValkenburgh, Parker. Review of *Vertical Empire: The General Resettlement of Indians in the Colonial Andes* by Jeremy Mumford (Duke, 2012) and *Negotiated Settlements: Andean Communities and Landscapes under Inka and Spanish Colonialism* by Steven Wernke (Florida 2013). *Colonial Latin American Review* 23(2): 280-294
- 2012 VanValkenburgh, Parker. Review of *Converting Words: Maya in the Age of the Cross*, by William Hanks. *Journal of Latin American and Caribbean Anthropology* 351-353.
- 2008 VanValkenburgh, Parker. Review of *Heads of State: Icons, Power and Politics in the Ancient and Modern Andes*, by Denise Arnold and Christine Hastorf. *Arqueolog.com*.

#### TECHNICAL REPORTS

- 2021 Parker VanValkenburgh and Alicia Odewale. "Report to the Tulsa Development Authority on Geophysical Survey Conducted by the Mapping Historical Trauma in Tulsa, 1921-2021 project in October 2020."
- 2021 Alicia Odewale and Parker VanValkenburgh. "Report to Oklahoma State University – Tulsa on Geophysical Survey Conducted by the Mapping Historical Trauma in Tulsa project, in October 2020"

- 2019 Carol Rojas Vega and Parker VanValkenburgh. Paisajes Arqueológicos de Choctamal - Chachapoyas – Fase I, Temporada 3 – Excavaciones en Purun Llaqta del Maino - (Informe Técnico). Ministerio Nacional de Cultura (Peru).
- 2018 Carol Rojas Vega and Parker VanValkenburgh. Paisajes Arqueológicos de Choctamal - Chachapoyas – Fase I, Temporada 2 – Pozos de Prueba - (Informe Técnico). Ministerio Nacional de Cultura (Peru).
- 2017 Carol Rojas Vega and Parker VanValkenburgh. Paisajes Arqueológicos de Chachapoyas – Fase I – Prospección - (Informe Técnico). Ministerio Nacional de Cultura (Peru).
- 2016 VanValkenburgh, Parker and Daniela Zevallos Castañeda. Proyecto Arqueológico Zaña Colonial – Fase III – Investigaciones de Materiales en Laboratorio – Temporada 2016 (Informe Técnico). Ministerio Nacional de Cultura (Peru).
- 2015 VanValkenburgh, Parker and Daniela Zevallos Castañeda. Proyecto Arqueológico Zaña Colonial – Fase IIc - Excavaciones en el Sitio de Carrizales (LA-057), Temporada 2015 (Informe Técnico). Ministerio Nacional de Cultura (Peru).
- 2014 VanValkenburgh, Parker and Rocío Torres Mora. Proyecto Arqueológico Zaña Colonial – Fase IIb - Excavaciones en el Sitio de Carrizales (LA-057), Temporada 2014 (Informe Técnico). Ministerio Nacional de Cultura (Peru).
- 2012 VanValkenburgh, Parker and Natalia Guzmán Requena. Proyecto Arqueológico Zaña Colonial – Fase IIa - Excavaciones en el Sitio de Carrizales (LA-057), Temporada 2012 (Informe Técnico). Ministerio Nacional de Cultura (Peru).
- 2010 VanValkenburgh, Parker and Carol Rojas Vega. Mapeo, Prospección y Pozos Iniciales en el Valle de Zaña. Temporada 2009-10 (Informe Técnico). Ministerio Nacional de Cultura (Peru).

#### INVITED PRESENTATIONS

- 2022 “Rethinking the Archeology of Settlement” - Boston University Archaeology Lecture Series, December 1<sup>st</sup>
- 2022 "South American Archaeology at Scale: the Geospatial Platform for Andean Culture, History, and Archaeology (GeoPACHA). Archaeological Institute of America “Quick Takes:” Archaeology and Big Data.
- 2021 “Archaeology, Big Data, and Big Questions about lidar in the Andean Cordillera and Beyond” - Texas Tech / American Institute of Archaeology, Lubbock, TX (remote)
- 2021 “Lasers Below the Clouds: Mapping Kuelap with Drone-mounted and Terrestrial Lidar,” in Dumbarton Oaks Research Library and Collections, June (remote)
- 2021 “South American Archaeology at Scale?: The Geospatial Platform for Andean Culture, History, and Archaeology (GeoPACHA),” Archaeology Seminar - Harvard University, February (remote)
- 2020 “Title TBD.” Colorado State University Department of Anthropology and Geography Seminar Series, September 25th (canceled due to Covid-19)
- 2020 “Webinar – GeoPACHA: Geospatial Platform for Andean Culture, History and Archaeology.” Universidad de Ingeniería y Tecnología, Lima, Peru (remote), June

- 2019 "Arqueología Aérea en la Cuenca del Río Utcubamba: Paisajes Arqueológicos de Chachapoyas," AGOPI, Chachapoyas, August
- 2019 "Conferencia Magestral: Proyecto Arqueológico Choctamal – Chachapoyas." INAAK and Universidad Nacional Toribio Rodríguez de Mendoza, Chachapoyas, Peru, May
- 2019 Discussant, "Lost in Transition" symposium, organized by Clea Moulin, Abel Traslaviña, and James Zimmer-Dauphinee Society for American Archaeology Annual Meetings, April
- 2019 "Mapping Kuelap Using UAV-mounted LiDAR and Hyperspectral Imagery," Harvard University Andean Seminar, April
- 2019 "LOGAR and GeoPACHA: Progress and Prospects of Collaborative Digital Humanities Tools for the Archaeology and Ethnohistory of the Andes." With Steven Wernke and Akira Saito. Research Institute for Humanity and Nature, Kyoto, Japan, March
- 2019 "Las Reducciones en Chachapoyas," National Museum of Ethnology, Osaka, Japan, March
- 2018 Discussant, "#MeToo and Archaeology," panel organized by Jason De Leon and Chelsea Fischer, American Anthropological Association Annual Meetings, November
- 2018 "Becoming Indios: Genealogy, Landscape and Political Subjectivity in the Zaña Valley, Peru, 15th-18th Centuries CE." Department of Archaeology, Leiden University, Netherlands, October
- 2018 "LOGAR y GeoPACHA: Logros y posibilidades de humanidades digitales para el estudio de la Reducción General de Indios." Un Imperio, Múltiples Espacios Conference Hosted by Qhapac Ñan Project, Ministerio de Cultura, Lima, Peru, August
- 2018 "LOGAR and GeoPACHA: Progress and Prospects of Collaborative Digital Humanities Tools for the Archaeology and Ethnohistory of the Andes." by Steven A. Wernke, Parker VanValkenburgh, and Akira Saito. Invited talk presented at the workshop / Andean Archaeologists of Central New York/, Syracuse University, April
- 2018 "Implementing Collaborative Digital Humanities Tools for the Andean Archaeology and History," by Steven A. Wernke, Parker VanValkenburgh, and Akira Saito. Invited presentation for "Thinking Through the Archives," a Digital Humanities Best Practices Workshop, University of Southern California, April
- 2018 "Domestic Life in Toledan Reducciones and the Case of Carrizales, Zaña Valley, Peru" *Unsettling Resettlement* conference, Vanderbilt University, February
- 2017 "The Edges and Ends of Empire in Northern Peru, 16<sup>th</sup>-18<sup>th</sup> Centuries CE" University of Frankfurt and Max Planck Institute of European Legal History, Frankfurt, Germany
- 2017 "Building Indios: A Genealogy of Landscape and Political Subjectivity in Peru's Zaña Valley, 12th-18th Centuries CE" McDonald Institute of Archaeology, Cambridge University, Garrod Seminar, October
- 2017 "Site-seeing: Archaeology, Aeriality and Objectivity in coastal Peru," Smith College, April
- 2017 Forum Participant, "Afro-Latin American Archaeology," Society for American Archaeology Annual Meeting
- 2017 "Surf and Turf: Maritime Foodways and the Challenge of Spanish Empire at Carrizales, Peru - 12th to 17th Centuries CE. Boston University Department of Archaeology Brown Bag Lecture Series, March

- 2017 "El Contrato del Mar: Diet, Daily Life and Political Subjectivity at Carrizales, Peru 12th-16th centuries CE. Yale University Department of Anthropology Archaeology Brown Bag Series, February
- 2017 "Building Indios: A Genealogy of Landscape and Political Subjectivity in Peru's Zaña Valley, 12th-18th Centuries C.E." University of Michigan, Center for Latin American and Caribbean Studies, January
- 2016 Discussant - "Ethnohistorical Uses of Geographical Information Systems," Panel at American Society for Ethnohistory Annual Meeting, November
- 2016 "'Comparative Perspectives on the Impact of Reducción on Indigenous Domestic Life'" Colonial Modernity in the Andes: A Comprehensive Study of Viceroy Toledo's General Resettlement," hosted by Dr. Akira Saito - La Paz, Bolivia, December
- 2015 "Archaeological Perspectives on the *Reducción* Process: Considering Interregional Variation in the Movement's Dynamics and its Afterlives." Vanderbilt University, November
- 2014 "To See and Be Seen: Towards an Archaeology of Scopic Regimes." *Artifacts as Evidence: The Material Record of Politics* Symposium at Washington University in St. Louis, October
- 2014 "Drones and Bones: Digital Approaches to Archaeology in Northern Peru," Second Annual James B. Petersen Memorial Lecture, University of Vermont, September
- 2013 "Regional Approaches to Reducción in Northern Peru: Examples from the Zaña and Chamán Valleys." Vanderbilt University, November

#### JURIED PRESENTATIONS AND POSTERS

- 2022 "The Sustainability of Agricultural Terraces: Three Scales of Evaluation from Chachapoyas, Peru" - in "Lessons from the Past: Archaeological Insights on Environmental, Economic, and Social Sustainability" - Society for American Archaeology Annual Meeting, Chicago, IL; Daniel Plekhov, Parker VanValkenburgh, Carol Rojas Vega, Alexis Reátegui Díaz
- 2022 "Deploying GeoPACHA: An Overview of Results from Stage One," in GeoPACHA: Expanding scales of archaeological analysis through imagery-based survey in the Andes; Society for American Archaeology Annual Meeting, Chicago, IL
- 2021 "Examining Long-term Histories of Land Use in Peru's Chachapoyas Region" with Mark Bush, Courtney Shadik, Daniel Plekhov and Bethany Whitlock, Society for the Anthropology of Lowland South America, 6/2021
- 2020 Roundtable Chair, "Collaborative Virtual Survey in the Andes: The Geospatial Platform for Andean Culture, History, and Archaeology" Society for American Archaeology Annual Meeting, 4/2020 (cancelled)
- 2019 Roundtable Presenter, "Disrupting Power: Between and Beyond Hierarchy and Egalitarianism in the Community," American Anthropological Association Annual Meeting, 11/2019
- 2019 "Here's Looking at You: The Politics of UAV-based and satellite-based aerial survey in the Peruvian Andes," Society for American Anthropology Annual Meeting, 4/2019

- 2018 Roundtable Presenter, "Mobility, Resilience and Resistance: Migration in Historical Perspective." American Anthropological Association Annual Meeting, 11/2018
- 2017 "Spanish Colonial Impacts on Foodways and Diet in the Zaña Valley of Peru: A Multi-Isotopic Reconstruction" second author, with Bethany Turner-Livermore and Benjamin Schaefer. American Association of Physical Anthropologists, 4/2017
- 2017 "Site-seeing: archaeology, aeriality and objectivity in coastal Peru" (submitted; session organized by M. Llobera), Society for American Archaeology Annual Meetings, Vancouver, BC, 4/2017
- 2017 "*Chacras* in the Clouds: Documenting High-Altitude Agricultural Landscapes in the Tambillo Valley of Chachapoyas, Amazonas, Peru" - Society for American Archaeology Annual Meetings, poster presentation, second author, with Joseph Cronin and Anna Guengerich 4/2017
- 2016 "Forced Resettlement and Governmentality in Spanish Colonial Latin America: Comparative Approaches from Peru and New Spain" American Anthropological Association Annual Meeting 11/2016
- 2016 "Landscape, Labor and the Dialectics of Difference in Zaña, Peru during the 16<sup>th</sup>-18<sup>th</sup> and 21<sup>st</sup> Centuries, American Society for Ethnohistory Annual meetings, 11/2016"
- 2016 "Caracterización química de cerámica colonial temprana en el sitio de Carrizales, Lambayeque, Peru" Congreso Nacional de Arqueología, 9/2016, with Diego Bedoya Vidal and Sarah Kelloway.
- 2016 "Diet, Status, and Identity in Colonial Peru: Investigations at Carrizales" (2nd author, w/ S. Kennedy). Society for American Archaeology Annual Meetings, 4/2016
- 2016 "Landscape, Labor, and the Production of Difference in Colonial Peru: *Indios* and *Negros* in the Zaña Valley, 16<sup>th</sup> through 18<sup>th</sup> centuries C.E." Society for American Archaeology Annual Meetings, 4/2016
- 2015 "Forced Resettlement, Domestic Space and Maritime Subsistence at Carrizales (Zaña Valley, Peru), 13<sup>th</sup>-17<sup>th</sup> centuries C.E." Northeast Conference in Andean Archaeology and Ethnohistory, Toronto, ON, 10/2015
- 2015 "Gardens and Forking Paths: A Genealogy of Landscape and Subject Formation in the Zaña Valley, Peru." Society for American Archaeology Annual Meetings, 4/2015
- 2015 "Architecture in Negative: Mapping Social Space at Carrizales, Peru Using Low Altitude Aerial Photography and Photogrammetry." Society for American Archaeology Annual Meetings, with Mark Willis and Chester P. Walker, 4/2015
- 2014 "Unsettling Subjects: Strategies of Amnesia and Tactics of Remembrance in Colonial Peru." American Anthropological Association Annual Meetings, 12/8/2014
- 2014 "Putting FAIMS (Federated Archaeological Information Management System) to use in Coastal Peru" Australian Archaeological Association Annual Meetings - Remote Presentation, 12/4/2014
- 2014 "El Contrato del Mar: Forced Resettlement and Maritime Subsistence at Carrizales and Conjunto 131, Zaña Valley, Peru" (with Sarah Kennedy, Carol Rojas Vega, and Gabriel Hassler) Society for American Archaeology Annual Meeting, Austin, TX, 4/2014

- 2014 "Chérrepe in Fragments: Time, Place and Representation in Andeanist Historical Archaeology," Society for Historical Archaeology Annual Meeting, Quebec, QC, Canada, 1/2014
- 2013 "Probing the analytical potential of low altitude aerial photography and digital photogrammetry in the lower Zaña Valley, Perú." (with Chester P. Walker) Northeast Conference in Andean Archaeology and Ethnohistory, New Haven, CT, 11/2013
- 2012 "The World on a Plate: Consumption and Commerce At Magdalena De Cao Viejo Through the Lens of Ceramic Analysis." American Anthropological Association Annual Meetings, 11/2012.
- 2012 "Regional Archaeology in Coastal Peru: Adapting New Approaches to Solve Old Problems" – Society for American Archaeology Annual Meeting 4/2012
- 2011 "Urbs and Civitas: Landscapes of Forced Resettlement in the Zaña and Chamán Valleys, Peru – 16<sup>th</sup> through 18<sup>th</sup> Centuries C.E.," Dumbarton Oaks, 3/2011 and (with C. Rojas Vega), Society for American Archaeology Annual Meeting 4/2011
- 2009 "Home Turf: Archaeology, Territoriality, and Politics" (with J. Osborne), Society for American Archaeology Annual Meeting, Atlanta, GA, 4/2009
- 2008 "Continuity and Transformation in Indigenous Households During the Colonial Period: Recent Insights from Magdalena de Cao Viejo, Peru" (First Author; Second Author - J. Quilter) Society for American Archaeology Annual Meeting, Vancouver, BC, Canada, 3/2008
- 2005 "The Sound of Interregionalism in the Late Initial Period and Early Horizon: Strombus Trumpets from Chavín de Huántar, Peru" Society for American Archaeology Annual Meeting, Salt Lake City, 4/2005

#### CONFERENCE SESSION ORGANIZER / CHAIR

- 2022 GeoPACHA: Expanding scales of archaeological analysis through imagery-based survey in the Andes; Society for American Archaeology Annual Meeting, Chicago, IL; co-organized with Steven Wernke
- 2020 Collaborative Virtual Survey in the Andes: The Geospatial Platform for Andean Culture, History and Archaeology; Society for American Archaeology Annual Meeting, Austin, TX (canceled); co-organized with Steven Wernke
- 2019 Archaeological Vision in the Age of Big Data; Society for American Archaeology Annual Meeting Albuquerque, NM; co-organized with Andrew Dufton
- 2018 Executive Session - Mobility, Resilience and Resistance: Migration in Historical Perspective; American Anthropological Association Annual Meeting, San Jose, CA
- 2016 Deep Displacement: Excavating the History of Forced Migration and Forced Resettlement; American Anthropological Association Annual Meeting, Minneapolis, MN
- 2016 The Spatial Production of Blackness in the Spanish Americas; American Society for Ethnohistory, Nashville, TN; co-organized with Brendan J.M. Weaver
- 2014 Las Preguntas Que Cuentan; Society for Historical Archaeology Annual Meeting, Quebec, QC; co-organized with Ross Jamieson

- 2012 Frogs Crossing the Pond: New (and Old) Frontiers in Regional Archaeology; Society for American Archaeology Annual Meeting, Memphis, TN; co-organized with Thomas Leppard and Bradley Sekadat
- 2009 Territoriality in Archaeology; Society for American Archaeology Annual Meeting, Atlanta, GA; co-organized with James Osborne

## TEACHING

### **Instructor of Record at Brown University**

- 2021- Anthropology 2202 - Advanced GIS and Spatial Analysis
- 2019 Humanities 1973J - How to Things with Maps: Cartography, Power and Political Imagination, from Gilgamesh to Google
- 2017- Anthropology 2501 - Space, Power and Politics
- 2016- Anthropology 2201 - Archaeology in the Digital Age
- 2016- Anthropology 1505 - Vertical Civilization: Andean Archaeology from Monte Verde to the Inkas
- 2016- Anthropology 0500 - Past Forward: Introducing Anthropological Archaeology
- 2015- Anthropology 1201 - Introduction to GIS and Spatial Analysis

### **Instructor of Record at University of Vermont**

- 2014-15 Anthropology 196 - Landscape Anthropology and Spatial Analysis (S2014, S2015)
- 2014 Anthropology 95B - Conquistadors and Cannibals, University of Vermont
- 2014 Anthropology 210 - Archaeological Theory, University of Vermont (S2014)
- 2013-15 Anthropology 24 - Introduction to Prehistoric Archaeology (F2013, S2015)

### **Instructor of Record at Washington University in St Louis**

- 2013 Anthropology 4092 - Vertical Civilization: Andean Archaeology and Andean Environments, from Monte Verde to the Inkas
- 2012 Interdisciplinary Program in the Humanities 405 - History and Anthropology: Facets of a Complex Dialogue

## NATIONAL/INTERNATIONAL ACADEMIC SERVICE

- 2021- **Editorial Board**, Journal of Field Archaeology
- 2021 **Program Committee**, Society for American Archaeology Annual Meeting
- 2020 **Program Committee**, Society for American Archaeology Annual Meeting
- 2020 **Review Panelist**, National Endowment for the Humanities
- 2019 **Host/Organizer**, Northeast Conference on Andean Archaeology and Ethnohistory. Two-day conference hosted November 2019 at Brown University; 120 attendees
- 2015 **Review Panelist**, National Endowment for the Humanities



- 2014 **Host/Organizer**, Northeast Conference on Andean Archaeology and Ethnohistory. Two-day conference hosted November 2014 at University of Vermont; 80 Attendees
- 2014- **Editorial Board**, Historical Archaeology in South America Series, University of Alabama Press
- 2012- **Ad Hoc Reviewer** - Advances in Archaeological Practice, American Anthropologist, Antiquity, Anthropologica (Peru), Boletín de Arqueología PUCP, Cambridge Archaeological Journal, Colonial Latin American Review, Current Anthropology, Field Museum Analytical Research Facility (NSF), Journal of Archaeological Science, Journal of Archaeological Science: Reports, Journal of Archaeological Research, Journal of Field Archaeology, Journal of Archaeological Method and Theory, Latin American Antiquity, National Geographic Committee on Research and Exploration, National Science Foundation, PLOS One, SSRC International Dissertation Research Fellowship, University of Utah Press, University Press of Florida

#### UNIVERSITY ACADEMIC SERVICE AT BROWN UNIVERSITY

##### University Committees

- 2020-22 **Faculty Executive Committee**
- 2019-20 **Standing Committee on Academic Code**

##### Advisory / Executive Committees for Academic Programs

- 2022- **Faculty Fellow, Joukowsky Institute of Archaeology and the Ancient World**
- 2021- **Campus Advisory Board, Brown Data Science Initiative**
- 2019- **Academic Advisory Committee, Spatial Structures in the Social Sciences**
- 2018- **Academic Advisory Committee, Program in Early Cultures**
- 2018-19 **John Carter Brown Library, Academic Advisory Committee**
- 2016-17 **John Carter Brown Library, Academic Advisory Committee**

##### Review Panels for University Awards

- 2022 **Center for Latin American and Caribbean Studies Dissertation Prize Committee**
- 2022 **Spring Application Review Committee**
- 2019 **UTRA Application Review Committee**
- 2018 **UTRA Application Review Committee**
- 2017 **Rhodes, Marshall and Mitchell Scholarships Review and Interview Committee**
- 2016 **Rhodes, Marshall and Mitchell Scholarships Review and Interview Committee**
- 2015 **Rhodes, Marshall and Mitchell Scholarships Review and Interview Committee**

##### Undergraduate Advising

- 2020-21 **First-Year Advisor and Second-Year Advisor**

2018-19 **First-Year Advisor**

2016-17 **First-Year Advisor**

DEPARTMENT ACADEMIC SERVICE AT BROWN UNIVERSITY

2022-23 **Director of Graduate Admissions, Department of Anthropology**

2022 **Watson Smith Prize Review Committee**, Department of Anthropology

2022-23 **Faculty Supervisor, Anthropology Department Colloquium**

2021-22 **Faculty Supervisor, Anthropology Department Colloquium**

2020-21 **Faculty Supervisor, Anthropology Department Colloquium**

2020-21 **Department Liaison, Sheridan Center for Teaching**

2018-19 **Faculty Supervisor, Anthropology Department Colloquium**

2019 **Graduate Curriculum Review Committee**, Department of Anthropology

2016 **Watson Smith Prize Review Committee**, Department of Anthropology

2017 **Watson Smith Prize Review Committee**, Department of Anthropology

2015- **Director, Brown University Digital Archaeology Lab**. Hold lab meetings and workshops; administer GIS computer cluster; provide support to student researchers

LECTURE SERIES ORGANIZED

2020-21 "The Uses of Space in Andean History and Archaeology," with Jeremy Mumford

2019-20 "City Rhythms and Urban Forms," with Rebecca Carter, A Suitor, M. Paulk

2018-19 "Performance, Embodiment and Identity in the Andes," with Jessaca Leinaweaver

2018-19 "Precolumbian Dialogues," with Andrew Scherer

2016-17 "Precolumbian Dialogues," with Andrew Scherer and Stephen Houston

2016-17 "Approaching Race and the Politics of Indigeneity in the Andes," with Jeremy Mumford

OTHER SERVICE

2019- **Board Member**, Public Archaeology Laboratory, Pawtucket, RI

MENTORSHIP

**Graduate Student Primary Advisees**

Matthew Balance (Brown University, Anthropology – entering fall 2022)

Alyssa Bolster (Brown University, Anthropology; co-advisor – entering fall 2022)

Daniel Plekhov (Brown University, Joukowsky Institute; co-advisor – Ph.D. 2022)

Daiana Rivas-Tello (Brown University, Anthropology, year 4)

Bethany Whitlock (Brown University, Anthropology, year 5)

**Graduate Student Committees**

Mark Agostini (Brown Anthropology) – Preliminary exam committee

Omar Alcover (Brown Anthropology) – Dissertation committee, grad. 2019  
Morgan Clark (Brown Anthropology) – M.A. Committee  
Eve Dewan (Brown Anthropology) - Dissertation committee, grad. 2020  
Fabiola Hernandez (Brown Anthropology [sociocultural]) - M.A. Committee, grad. 2019  
Nkem Ike (University of Tulsa, Anthropology) - Dissertation Committee  
Megann Phillips (University of Central Florida) – Dissertaiton Committee  
Daniela Raillard (Northwestern University, Anthropology) - Dissertation Committee  
Sophie Reilly (Northwestern University, Anthropology) - Dissertation Committee  
Alejandra Roche Recinos (Brown Anthropology) – Dissertation Committee, grad. 2021  
Miriam Rothenberg (Brown, Joukowsky Inst. of Archaeology) – Diss. Committee, grad. 2021  
Joshua Schnell (Brown Anthropology) – Dissertation Committee  
Maria Smith (Syracuse University, Anthropology) - Dissertation Committee  
Dé Leonel Soares (Universidade de Sao Paulo, Brazil) – Dissertation Committee, grad. 2022  
Anabelle Sutor (Brown Anthropology [sociocultural]) – Dissertation Committee  
Rachel Thimmig (Brown University, Anthropology) – M.A. Committee  
Jennifer Thum (Brown, Joukowsky Institute of Archaeology) - Dissertation Committee, grad. 2019

## Project Proposal Inbox x



**Allotey, Sir Dylan**

Thu, Oct 9, 10:12 AM (5 days ago) ☆

Good morning Professor Parker, I have worked on my proposal, and I wanted to hear your thoughts before I revised it. I t...



**VanValkenburgh, Parker**

Thu, Oct 9, 1:13 PM (5 days ago) ☆ ↩ ⋮

to me ▾

Dear Sir Dylan -- thanks so much for reaching out and I'm excited to hear that you're planning on working on this project! The LiDAR data I have will be ideal for it.

All the best,  
Parker VanValkenburgh



--

**Parker VanValkenburgh** (he/him/his)  
Associate Professor of Anthropology and Archaeology  
Director of Graduate Studies, Department of Anthropology

# Uncovering patterns of Institutional trust and perception

Shravya Munugala

PI

shravya\_munugala@brown.edu

Gordan Milovac

Co-PI

gordan\_milovac@brown.edu

Bertram Malle

Collaborator

bfmalle@brown.edu

October 14, 2025

## **Abstract**

We present a novel tool for visualizing perceptions of government agencies across multiple trust and style dimensions. The tool overcomes a limitation in existing two dimensional charts, which oversimplify complex patterns. In a planned evaluation study, participants will compare this system with existing two dimensional visualizations, and we expect the results to show that our approach is easier to interpret and better at supporting independent insight generation.

# 1 Response to Reviewers

## 1.1 Reviewer: Sir Dylan Allotey

### **Overall: 2**

Strong, clearly written, and well-motivated proposal. It articulates practical contributions.

### **Interdisciplinary: 1**

Combines psychology, data science, and interactive visualization.

### **Scientific: 2**

The evaluation design with baseline 2D plots ensures measurable outcomes, and the data preprocessing follows sound scientific reasoning.

### **Visualization: 2**

The visualization aims to represent trust data in an easy-to-understand way and goes beyond static 2D representations. The plan for clustering, filtering, and profile inspection may allow for insightful visualizations.

### **Significant: 3**

The potential impact is high for both research and practice. It could influence journalism and public communication.

### **Novel: 2**

Combining multidimensional trust metrics with interactive visualization is innovative. Prior visualizations such as Chernoff faces have not been applied in this domain.

### **Goals Clearly Stated: 1**

Clear articulation of objectives, methods, and outcomes. The stages of the research plan are specific, achievable, and logically ordered.

### **Likelihood of Success: 2**

Given the PI and Co-PI's technical skills and prior experience, this project is highly feasible. The timeline may be a little tight.

### **Strengths:**

- Well-written and structured proposal with a clear research trajectory.
- Strong interdisciplinary integration between social science, computer science, and visualization.
- Clear evaluation plan and realistic milestones.
- Team's technical background is extremely relevant.

### **Weaknesses:**

- Limited discussion of data sources and how trust and style dimensions are measured.  
*Response: The data for this project come from an existing survey collected by the collaborator, capturing public perceptions of approximately 30 U.S. government agencies. Trust is measured across*

*five dimensions—competence, reliability, ethical integrity, transparency, and benevolence—while interpersonal style is captured along two axes: warmth and dominance. All responses have been anonymized and structured into a consistent dataset format for analysis.*

- Success depends heavily on user interface design quality, which is not deeply detailed.  
*Response: The current draft emphasized conceptual design, but the revised version now includes more detail on interface planning. Specifically, we outline design principles of clarity, accessibility, and intuitiveness, and describe how elements such as interactive clustering, hover-based agency profiles, and dynamic filtering will be implemented using Plotly and Dash. This clarifies how interface design decisions support interpretability and usability goals.*
- Novelty leans on combination and application rather than innovation.  
*Response: Our aim is to bring together tools from machine learning and visualization with ideas from social perception research in a way that hasn't been done before. The innovation lies in how these techniques interact to produce new forms of interpretability and user engagement, helping people see patterns and relationships in complex trust data that would otherwise stay hidden. This revision makes clear that the project's strength lies in its interdisciplinary approach and its connection of technical methods with real-world understanding.*

## **1.2 Reviewer: Skylar Walters**

### **Overall: 3**

The proposal has incredibly strong interdisciplinary and scientific contributions. However, the proposal would benefit from a more robust description of the proposed visualization—what does the program explicitly build on from dimensionality reduction? How does it achieve interpretability? How will users interact with the visualization? Perhaps a mock-up would be beneficial in the next iteration.

*Response: Thank you for the feedback! We agree that the visualization needs a clearer description. It builds on dimensionality reduction (mainly UMAP) to position agencies based on similarity across trust and style dimensions, making clusters and differences easier to see. Users will be able to hover or click to view detailed profiles and use filters to explore specific dimensions. We will also include a simple mock-up in the next version to better illustrate how the tool supports interpretability and interactive exploration.*

### **Interdisciplinary: 1**

The proposal clearly outlines the integration of visualization techniques with social sciences. Outside of the given project, the author also discusses areas in the social sciences more broadly where this method could be applied.

### **Scientific: 2**

The proposal defines that the project will enable more robust knowledge of government agency perception, which is currently lacking due to limited visual distinction in imagery.

### **Visualization: 5**

The project outlines what is wrong with current visualization methods in the field. However, the actual description of the visualization was lacking. Example images of the display could be helpful for understanding what additional elements this work incorporates.

**Significant: 2**

With political divisiveness seemingly increasing by the day, having a robust understanding of public perception of the government is essential. Creating a tool for visualization of social metrics allows exploration of these ideas within the project and in a larger context.

**Novel: 5**

This builds on past work in dimensionality reduction, interactivity, and social metric tracking. However, this appears to be the first method to include both interaction and interpretation in a visualization for social metrics. The novelty score would benefit from a more robust description or mock-up image describing the software in greater depth.

**Goals Clearly Stated: 3**

Goals are clear but could benefit structurally from an explicit goal list, though this may be a matter of personal preference.

**Likelihood of Success: 3**

The PI has a background in software engineering, which will be helpful in designing interactivity. Additionally, the timeframe is logical, though more time may be needed to apply changes from the interview stage into the app.

**Strengths:**

- Scientific contribution and significance are very strong. The list of potential applications outside of the project conveys this effectively.
- The current figure in the proposal emphasizes the current limitations of the state-of-the-art visualization.
- The project is clearly interdisciplinary and, as noted above, can be applied beyond this project and more broadly across the social sciences.

**Weaknesses:**

- A mock-up or more in-depth description of the proposed project would greatly benefit understanding of the project's differentiation from other work and emphasize its novelty.  
*Response: Please refer to the first reviewers response[R1.3] and [R2.1]. The mock-up will be present in the final draft.*

**1.3 Reviewer: Yonas Amha****Overall: 3**

Great proposal.

**Interdisciplinary: 3**

Combines visualization with government agency perceptions.

**Scientific: 3**



The proposal highlights a novel way of visualizing government agency perceptions that builds upon the previous state of the art.

**Visualization: 4**

The visualization goals seem to be straightforward.

**Significant: 4**

The project seeks to change how government agencies are interpreted.

**Novel: 4**

The visualization extends previous behavioral visualization methods by showing five trust dimensions.

**Goals Clearly Stated: 3**

The goals of the proposal are all clearly communicated.

**Likelihood of Success: 2**

**Strengths:**

- The proposal presents a very clear plan and accurately addresses the shortcomings of the previous state of the art.

**Weaknesses:**

- The overall goals for the visualization, besides changing the trust dimensions, are a little vague and could be expanded upon further.

*Response: Thank you for pointing that out. We will expand upon the visualization aspect in the final draft. Below are the features we plan to integrate:*

- 1. **Pattern Discovery:** Highlight how users can identify clusters of agencies with similar trust profiles or interpersonal styles.*
- 2. **Comparisons Across Agencies:** Show how the tool allows side-by-side comparisons of agencies on trust dimensions and interpersonal metrics.*
- 3. **Interactive Exploration:** Explain interactive features like filtering, hovering, and selecting specific agencies to inspect detailed data.*
- 4. **Insight into Relationships:** Demonstrate how overlaying interpersonal style (warmth and dominance) on trust dimensions can reveal correlations or trends.*

## **1.4 Reviewer: Yiyang Zhang**

**Overall: 3**

Well-motivated problem with a feasible plan that blends social science constructs and modern visualization, though evaluation rigor and dimensionality reduction (DR) transparency need tightening.

**Interdisciplinary: 2**

Strong mix of social psychology, HCI, visualization, and machine learning with clear cross-domain applicability and an expert collaborator.

**Scientific: 4**

The approach is well described, but the study needs specified inclusion criteria, target sample size, defined tasks and metrics, and guardrails against DR-induced distortion to support valid inferences.

**Visualization: 3**

Interactive multi-dimensional exploration is appropriate. Could add stability and uncertainty cues, legends for dimensions and weights, and controls to mitigate clutter.

**Significant: 3**

The project could materially improve public-facing analyses of institutional trust and support evidence-based discussion, with transferability to other perception datasets.

**Novel: 4**

The integration is thoughtful, though the components (dimensionality reduction, clustering, and interactive profiles) are established. Novelty hinges on the specific trust/style synthesis and evaluation.

**Goals Clearly Stated: 3**

Timeline and features are clear, though success criteria, preregistered hypotheses, and data-governance details should be made explicit.

**Likelihood of Success: 3**

With a fixed dataset of 30 agencies and mature tooling (Python + Plotly/Dash), a working interactive system and pilot study are very achievable in six weeks.

**Strengths:**

- Clear motivation and a compact dataset enabling rapid iteration.
- The interface concept invites discovery while aligning with contemporary HCI guidance on interactivity.

**Weaknesses:**

- Evaluation plan lacks sampling and analysis specifics. Dimensionality reduction may mislead without stability and neighborhood-preservation reporting.

*Response: Thank you for your feedback! The evaluation plan has been updated to specify sampling details, analysis methods, and dimensionality reduction stability checks. It now reports neighborhood-preservation metrics (trustworthiness and continuity) and includes tests across multiple UMAP parameter settings to ensure consistent, interpretable results.*

**Other Comments for Discussion:**

- Consider defining study tasks and metrics such as accuracy, time on lookup, and setup comparison tasks. Report dimensionality reduction parameters and stability (seed/split sensitivity, trustworthiness, and continuity).

*Response: The evaluation section has been revised to include clear task definitions. Dimensionality reduction parameters and stability measures are now explicitly documented to enhance transparency and reproducibility.*

## 2 Aims

The main aim of this project is to develop an interactive visualization tool that helps people make sense of how government agencies are perceived across multiple dimensions of trust and interpersonal style. Existing approaches often condense such information into a single score or a static two-dimensional chart, which oversimplifies complex patterns and obscures meaningful differences between agencies. Our tool addresses this limitation by projecting agencies into a reduced-dimensional space where similarities and differences can be explored intuitively through interactive, human-centered design.

The dataset for this project comes from an existing survey collected by our collaborator, capturing public perceptions of approximately 30 U.S. government agencies. Trust is measured across five dimensions—competence, reliability, ethical integrity, transparency, and benevolence—while interpersonal style is represented along two axes: warmth and dominance. All responses have been anonymized and structured into a consistent dataset format for analysis.

The visualization supports multiple forms of exploration. Users can identify clusters of agencies with similar trust profiles, compare them side-by-side across dimensions, and examine individual profiles through hover- and click-based interactions. By overlaying interpersonal style on trust dimensions, the system reveals correlations and patterns that would otherwise remain hidden.

To further enhance interpretability, the visualization will include glyph-based representations that encode multiple agency characteristics simultaneously. Visual parameters such as color, size, and shape will be mapped to different trust and style dimensions, allowing users to intuitively perceive complex relationships at a glance. For instance, competence and benevolence—dimensions that most clearly differentiate agencies in this dataset—will be emphasized using vivid, saturated colors. The interface will also support flexible parameter mapping, enabling researchers to reassign visual encodings (e.g., color, size, orientation) to variables of their choice. This adaptability makes the system more generalizable to other datasets while empowering users to experiment with alternate mappings to uncover new insights.

Guided by the principles of clarity, accessibility, and intuitiveness, the interface—implemented using Plotly and Dash—aims to make complex data interpretable for both experts and non-experts.

To evaluate the tool’s effectiveness, we will conduct user studies and gather expert feedback to assess interpretability, insight generation, and usability. Ultimately, our goal is to demonstrate how an interactive, multidimensional approach can transform the way people explore and reason about government data.

## 3 Significance

This project addresses a fundamental challenge in understanding public perceptions of government agencies: representing complex, multidimensional trust data in a way that is both accurate and accessible. Traditional approaches often reduce trust to a single score or present it in simple two-dimensional charts (refer Figure 4), which obscure important differences between agencies and fail to capture nuances in interpersonal style.

Our tool integrates social science insights on trust and interpersonal style with techniques from human-computer interaction, machine learning, and information visualization to create an interactive, multidimensional representation. By combining five trust dimensions—competence, reliability, ethical integrity, transparency, and benevolence—with two interpersonal style dimensions, warmth and dominance, the tool provides a richer understanding of how agencies are perceived. Users can explore patterns, uncover clusters of agencies with similar profiles, and generate insights independently, making the tool valuable not only for researchers but also for policymakers, journalists, students, and engaged citizens.

The interdisciplinary nature of this work is a key strength. Social science theories guide the selection



Figure 1: Current 2D visualization of agency trust data

and interpretation of dimensions, while machine learning techniques such as dimensionality reduction and clustering allow visualization of complex relationships. Principles from human-computer interaction inform the interactive design, ensuring that exploration and comparison of agencies is intuitive even for non-experts. This integration across fields enables both technical rigor and human-centered interpretability, a combination that is rare in current trust data visualization approaches.

The novelty of this project lies in synthesizing these multiple perspectives rather than introducing a new algorithm. By bridging quantitative modeling with social perception theory and interactive visualization, the approach reveals patterns and relationships that traditional methods obscure, empowering users to generate meaningful insights independently.

Understanding relationships within government trust data is critical because trust underpins democratic governance and effective policy implementation. When trust declines, cooperation diminishes, misinformation spreads, and policy outcomes suffer. By making trust patterns visible, this tool can guide communication strategies, highlight areas needing reform, and enable evidence-based decisions that strengthen public confidence in institutions.

Beyond government applications, this framework is broadly applicable to any domain where multidimensional perception data is important, from healthcare and corporate reputation to education and policy evaluation. By combining interdisciplinary insights with interactive, quantitative analysis, this project demonstrates how complex datasets can be transformed into actionable understanding, advancing both scientific methodology and practical engagement with social perception.

## 4 Related Work

High-dimensional data visualization has a long history. Parallel coordinates, introduced by Inselberg and Wegman [1, 2], were among the first techniques for representing multidimensional datasets in two dimensions. While useful for expert analysis, these methods are mostly static and require significant training, which makes them difficult for general users to adopt. Similarly, metaphorical approaches like Chernoff faces [3] or symbolic representations such as trees and castles [4] make complex data more interpretable,

but they are still static, can only handle a limited number of dimensions, and do not scale well to larger datasets. These foundational methods are valuable, but they do not support interactive exploration or the simultaneous investigation of multiple trust dimensions, such as warmth and dominance.

Dimensionality reduction techniques, like Sammon mapping [5] and UMAP [6], offer ways to project high-dimensional data into lower-dimensional spaces while keeping the underlying structure intact. They make patterns and clusters easier to see, but in most prior applications, the resulting visualizations are static. Users cannot interact with the data, filter it, or explore patterns dynamically, which limits their usefulness for tasks that rely on human judgment or nuanced interpretation.

Our visualization builds directly on the interpretive capacity of dimensionality reduction (specifically UMAP) by projecting agencies based on similarity across trust and style dimensions. To maintain analytical validity, we will report measures of stability and neighborhood preservation—such as trustworthiness and continuity—to ensure local relationships are accurately represented. Unlike most static UMAP projections, this tool integrates dynamic overlays and clustering features that enhance interpretability and transparency of reduced-dimensional representations.

Recent research emphasizes the role of human-centered design and interactivity in making visualizations more understandable, reducing mental effort, and producing actionable insights [7, 8]. Techniques such as filtering, interactive exploration, and detailed profile inspection have been shown to improve user performance and reduce cognitive load, especially for non-experts. Our project builds on these principles by creating an interactive, multidimensional tool for trust data, where users can explore clusters, filter by dimensions, and examine detailed agency profiles. This approach directly addresses the shortcomings of previous methods and aims to make patterns more discoverable, interpretable, and meaningful.

By combining dimensionality reduction with interactivity and the layering of multiple trust dimensions, our tool fills an important gap in the literature. To our knowledge, no existing approach allows users to simultaneously explore multiple trust dimensions in a way that is both interactive and interpretable. We will evaluate the effectiveness of our tool through user studies and expert feedback, comparing it against baseline 2D visualizations and static projections. This will provide evidence for whether interactive multidimensional visualization can make trust relationships easier to understand, more actionable, and ultimately more insightful than current approaches.

## 5 Research Plan

Outlined below is the proposed research plan, covering all stages from preparing the dataset to completing the final proposal.

- **Week 1–2: Data Preparation and Exploration** The first two weeks will be devoted to cleaning and organizing the trust and interpersonal style dataset for all 30 government agencies. Exploratory analysis will be performed to examine distributions and relationships, followed by preliminary dimensionality reduction to identify clustering patterns. By the end of this stage, the dataset will be prepared for visualization and an initial 2D plot will be available.
- **Week 3–4: Prototype Development** Weeks three and four will focus on building the interactive visualization. The prototype will represent agencies in reduced-dimensional space, with hover or click functionality to display agency-specific details. Optional filtering and grouping by trust dimensions will be added, and visual clarity of clusters will be improved. This two-week period ensures a robust and usable prototype is in place.

- **Week 5: Iteration and Feature Refinement** During the fifth week, feedback will be gathered from collaborators or a small group of users, while continued development and refinement of the tool takes place. New features, usability improvements, and adjustments based on feedback will be implemented to enhance interpretability, interactivity, and user experience.

- **Week 6: Evaluation and Finalization**

The final week will be dedicated to structured evaluation of the tool as detailed in Section 5.1, completing documentation, and preparing a concise abstract and presentation/demo. At this stage, the functional prototype will be finalized and ready for submission.

## 5.1 Evaluation

The evaluation phase will involve approximately 20–30 participants, including students and early-career researchers familiar with social data. Each participant will use both the new visualization and a baseline 2D plot to perform comparison and interpretation tasks. We will define measurable outcomes such as task accuracy, completion time, and the number of insights reported.

The evaluation will use trust and style data from approximately 30 U.S. government agencies. For each agency, responses contributing to the five trust dimensions and two interpersonal style metrics will be aggregated and standardized prior to analysis. We will explicitly document how many responses contribute to each dimension and report these statistics in the evaluation summary. Dimensionality reduction outputs will be validated using neighborhood-preservation metrics such as trustworthiness and continuity, ensuring that the 2D projection preserves meaningful agency relationships. To confirm stability, UMAP will be re-run under multiple parameter settings and random seeds, and consistency of clustering patterns will be quantified across runs.

In designing the study, we will explicitly test several criteria that align with our visualization goals: (1) users’ feeling of understanding- how clearly they perceive trust and style relationships; (2) likelihood of discovery- whether the tool enables users to uncover previously unnoticed patterns; (3) accuracy of interpretation- the degree to which users correctly identify underlying relationships; and (4) enjoyability and engagement- how intuitive and satisfying the interactive experience feels. Selecting these criteria early allows us to prioritize design features that enhance interpretability and exploration from the start.

Evaluation will also include survey-based measures of clarity, interpretability, and user satisfaction. Dimensionality reduction metrics- particularly local neighborhood preservation—will be calculated to ensure relationships in high-dimensional space are faithfully represented. Consistency will be checked across different UMAP parameterizations and subsampling runs to validate robustness.

Success will be defined through a combination of these criteria: users’ ability to identify clusters and relationships accurately, high ratings of clarity and usefulness, stability of structure under parameter variation, and overall satisfaction with interactivity and design.

**Facilities and Resources** Development and analysis will be conducted on a personal computer using Python for data processing and dimensionality reduction, with Plotly and Dash for interactive visualization. GitHub will handle version control, and cloud resources may be used if additional computational capacity is needed.

## References

- [1] Inselberg, Alfred, and Bernard Dimsdale. "Parallel coordinates: a tool for visualizing multi-dimensional geometry." Proceedings of the first IEEE conference on visualization: visualization90. IEEE, 1990.
- [2] Wegman, Edward J. "Hyperdimensional data analysis using parallel coordinates." Journal of the American Statistical Association 85.411 (1990): 664-675.
- [3] Chernoff, Herman. "The use of faces to represent points in k-dimensional space graphically." Journal of the American Statistical Association 68.342 (1973): 361-368.
- [4] Kleiner, Beat, and John A. Hartigan. "Representing points in many dimensions by trees and castles." Journal of the American Statistical Association 76.374 (1981): 260-269.
- [5] Sammon, John W. "A nonlinear mapping for data structure analysis." IEEE Transactions on Computers 100.5 (2006): 401-409.
- [6] McInnes, Leland, John Healy, and James Melville. "UMAP: Uniform manifold approximation and projection for dimension reduction." arXiv preprint arXiv:1802.03426 (2018).
- [7] Franconeri, Steven L., et al. "The science of visual data communication: What works." Psychological Science in the Public Interest 22.3 (2021): 110-161.
- [8] Van Berkel, Niels, et al. "Impact of interaction technique in interactive data visualisations: A study on lookup, comparison, and relation-seeking tasks." International Journal of Human-Computer Studies 192 (2024): 103359.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
| <div>SHRAVYA MUNUGALA</div> <div>(401) 215-3306 — shravya_munugala@brown.edu — linkedin.com/in/shravya29M — github.com/Shravya29M</div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                             |
| EDUCATION                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                             |
| <div>Brown University</div> <div>Master of Science in Computer Science</div> <div> <ul style="list-style-type: none"> <li>Relevant Coursework: Deep Learning, Interdisciplinary Scientific Visualization</li> </ul> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <div>Expected May 2027</div> <div>Providence, RI</div>      |
| <div>Vellore Institute of Technology, Chennai</div> <div>Bachelor of Technology in Computer Science, GPA: 3.97/4</div> <div> <ul style="list-style-type: none"> <li>Relevant Coursework: Data Structures and Algorithms, Operating Systems, AI, DBMS, Software Engineering</li> </ul> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <div>May 2025</div> <div>Chennai, India</div>               |
| EXPERIENCE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                             |
| <div>Cloud Computing Intern</div> <div>Pentabrain Solutions Pvt Ltd</div> <div> <ul style="list-style-type: none"> <li>Assisted in building <b>serverless applications</b> with AWS Lambda and API Gateway, optimizing DynamoDB pipelines to <b>reduce data access latency</b> by 20%.</li> <li>Implemented <b>CI/CD pipelines</b> with GitHub Actions and CloudFormation for <b>automated deployments</b>.</li> <li><b>Collaborated with cross-functional teams</b> through sprint planning and standups, leveraging Jira to support event-driven workflows and <b>improve collaboration</b> in a fast-paced environment.</li> <li>Maintained documentation in GitHub, <b>enhancing knowledge transfer</b> and accelerating team onboarding.</li> </ul> </div> | <div>April 2025 - July 2025</div> <div>Chennai, India</div> |
| <div>Research Intern</div> <div>Center for Healthcare Advancement, Innovation and Research</div> <div> <ul style="list-style-type: none"> <li>Developed a <b>custom dual deep learning framework</b> to classify schizophrenia from 3D sMRI data. Optimized a CNN with 5-fold cross-validation, boosting accuracy by 13% across datasets.</li> <li>Adapted the model to diverse medical imaging datasets, maintaining <b>high clinical reliability and scalability</b>.</li> </ul> </div>                                                                                                                                                                                                                                                                       | <div>May 2024 - Dec 2024</div> <div>Chennai, India</div>    |
| PROJECTS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                             |
| <div>Elyssa – AI Mental Health Assistant   Python, Gradio, LLM, Huggingface, Kubernetes</div> <div> <ul style="list-style-type: none"> <li>Built Elyssa, a novel AI mental-health assistant fusing facial-expression analysis, sentiment-shift detection, and confidence-weighted response generation.</li> <li>Developed a <b>scalable, load-balanced system</b>, also improving response quality by 10.46% (ROUGE F1) and ensuring reliable, context-aware interactions.</li> </ul> </div>                                                                                                                                                                                                                                                                    | <div>Feb 2025 – July 2025</div>                             |
| <div>AutoNewsSummarizer   PyTorch, BART, BERT, BeautifulSoup, Scikit-learn, FastAPI</div> <div> <ul style="list-style-type: none"> <li><b>Engineered two novel NLP models:</b> a sentiment- and topic-aware summarizer for emotionally rich content, and a hierarchical long-form summarizer with iterative refinement.</li> <li><b>Deployed a real-time URL-to-summary pipeline</b>, generating context-aware summaries from live web content.</li> <li>Outperformed ROUGE and BLEU, <b>producing coherent</b>, recall-optimized summaries beyond standard models.</li> </ul> </div>                                                                                                                                                                           | <div>Dec 2023 – March 2024</div>                            |
| TECHNICAL SKILLS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                             |
| <div>Languages: Python, Java, C, C++, SQL, R, MATLAB</div> <div>Frameworks: TensorFlow, PyTorch, Keras</div> <div>Developer Tools: Docker, Google Cloud Platform, AWS (EC2, S3), Git, Github, Linux</div> <div>Libraries &amp; Specializations: Scikit-learn, Pandas, NumPy, OpenCV, Deep Learning, NLP, Data Structures &amp; Algorithms</div> <div>Certifications: AWS Cloud Practitioner (2025), Google Cloud Digital Leader (2025)</div>                                                                                                                                                                                                                                                                                                                    |                                                             |
| PUBLICATIONS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                             |
| <ul style="list-style-type: none"> <li>Shravya Munugala, et al. "Alzheimer's Disease Classification Using CUDA-Accelerated DenseNet201 and VGG16", IEEE ICICNIS, 2024.</li> <li>H.P, Shravya Munugala, et al. "GenoCryptNet: The Future of Secure Communication Through DNA", IEEE ICSTSDG, 2024.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                             |
| LEADERSHIP                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                             |
| <div>Head of HR and Management at PlaceXP Club</div> <div> <ul style="list-style-type: none"> <li>Led 7+ events and alumni panels, engaging 100+ students on college and internship opportunities.</li> <li>Coordinated a fundraising drive with 3 clubs, raising Rs. 15,000 for the education of rural children.</li> </ul> </div>                                                                                                                                                                                                                                                                                                                                                                                                                             | <div>Feb 2024 – Feb 2025</div>                              |

Figure 2: Resume of PI



# Gordan Milovac

+1 (401) 471-9866 | [gordan\\_milovac@brown.edu](mailto:gordan_milovac@brown.edu) | [linkedin.com/in/gmilovac](https://www.linkedin.com/in/gmilovac) | [gordanmilovac.com](https://gordanmilovac.com)

## Education

**Brown University**, Providence, RI

*Master of Science in Computer Science – AI/ML Track*

Sep 2025 - May 2026 (Expected)

*Bachelor of Science in Computer Science and Economics - Professional Track*

Sep 2021 - May 2025

- Sigma Xi Honors Society, 2023 LINK Award recipient, GAMMA, GPA: 3.7/4.0

## Experience

**Teaching Assistant**, [Brown University Department of Computer Science](#)

Jan 2025 - Present

- Assisted developing course content, Held labs and office hours, Graded and overseen 200+ students for Data Science

**Sales Engineer Intern**, [InterSystems Corporation](#)

May 2024 - Aug 2024

- Built an AI powered Synthetic Patient Data Generator, Gave over 5 Sales Presentations for the HealthShare team  
- Received special recognition from the department VP, Director and Manager, alongside a part-time extension offer

**Research Assistant**, [Brown University School of Engineering](#)

Jan 2024 - May 2024

- Devised techniques for aerial and ground robots to detect and map dangerous yet unexplored areas (FastSLAM)

**iOS App Developer**, [ScanEat](#)

Jun 2023 - Sep 2023

- Gained startup experience, Developed and tested iOS apps using Swift, Built UI features on a 100,000+ user app

**SWE and Multimedia Intern**, [DUMP Split](#)

Sep 2020 - Dec 2020

- Practiced full-stack web development (HTML/CSS, JS, React.js), Gained experience in photo and video editing  
- Helped organizing the annual DUMP Days conference, attracting 2000+ followers and max live appointments

## Projects

**AI-SDA Generator**, InterSystems Corporation

May 2024 - Aug 2024

- Website utilizes AI to generate and populate SDA medical records, reducing creation time from 40h to 5min  
- Integrated GPT-4o interaction commands using precise prompt engineering and parallel API calls (React.js)

**PeachInnovators**, Brown University

April 2024 - May 2024

- Gathered telemetry data for Brown Crew and modeled it to see which factors contribute to the boat speed (SQL)  
- Utilized the Pearson Correlation test to analyze the relationship of several key Peach system variables (Python)  
- Implemented a KERAS model utilizing a straightforward sequential architecture and 4-fold cross validation training

**AiRobic**, Brown University

April 2023 - May 2023

- Website generates a user-specific plan of rowing workouts, using model generation on the backend server (Java)  
- Algorithm utilizes Markov Model with a matrix between a set of hidden states, workouts stored in Firebase database

## Leadership

**D1 Athlete**, Brown Men's Varsity Crew

Sep 2021 - May 2025

- 25+ hours weekly commitment, Represented Brown at the Henley Royal Regatta and IRA National Championship  
- 2024 Ivy League Champion, Head of The Charles Regatta course record holder, 2024 Sri Chinmoy Marathon runner

**National Team Athlete**, Croatian Rowing Federation

Sep 2019 - Aug 2021

- 3x National champion, Indoor champion and Junior record holder, World and European Rowing Championship  
- Rowing Coach, Croatian Youth Olympic Team member, Awarded "Accomplished athlete" Trophy in 2020 and 2021

## Coursework & Skills

Intro to Programming (OOP), Program Design: Data Structures & Algorithms, Discrete Structures and Probability, Software Engineering, Computer Systems, Computer Graphics, Artificial Intelligence, Machine Learning (Capstone), Data Science, Cybersecurity and International Relations, Intermediate Calculus, Linear Algebra, Statistical Inference

**Languages:** Python, Java, C++, C, Swift, SQL, JavaScript, TypeScript, HTML, CSS, InterSystems ObjectScript

**Frameworks:** React, Next.js, Angular, Jest, NumPy, Pandas, PyTorch, Spark, JUnit, Firebase, Qt, OpenGL

**Tools:** Git, GitHub, Perforce, IRIS, VSCode, JetBrains IDE's, XCode, Adobe CC, PowerPoint, Excel, LaTeX, Linux

Figure 3: Resume of Co-PI

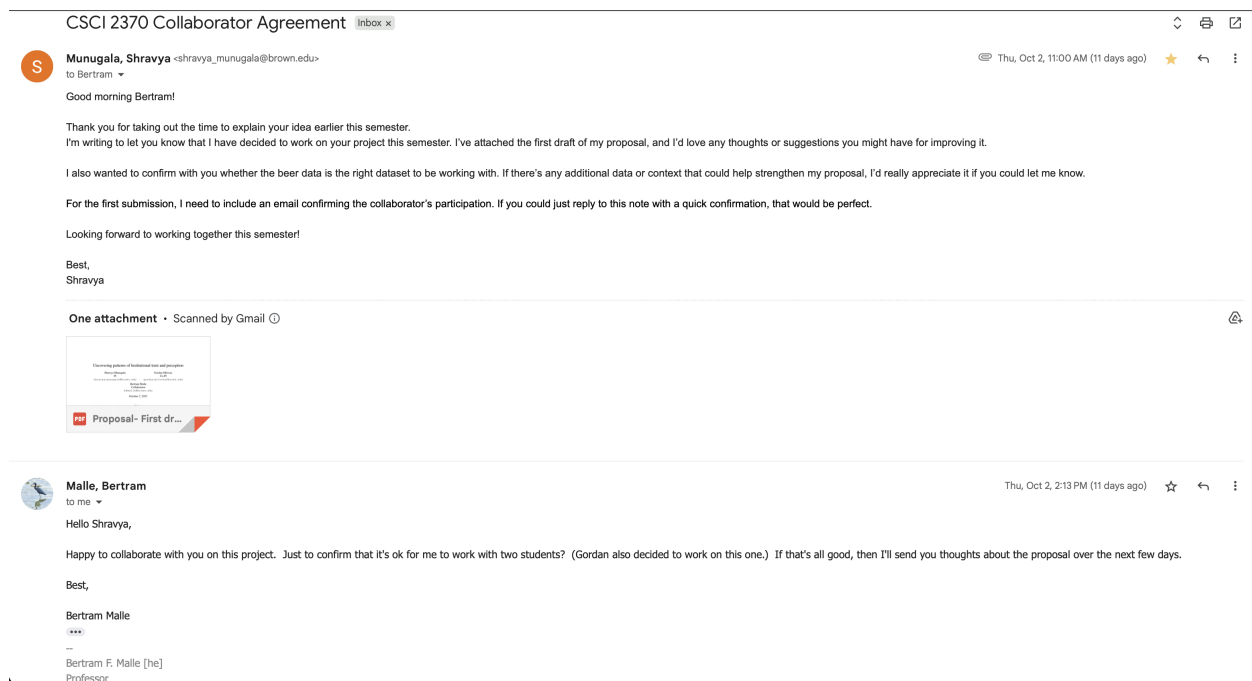


Figure 4: Collaborator Agreement Email

Link to the manuscript pdf:

[https://drive.google.com/file/d/1F0UOEe5Dlpmy99\\_j7rm8kpYZ8xE2KXrP/view?usp=drive\\_link](https://drive.google.com/file/d/1F0UOEe5Dlpmy99_j7rm8kpYZ8xE2KXrP/view?usp=drive_link)

## Response Letter

Dear Editor and Reviewers,

We appreciate the constructive feedback on the manuscript, and we have prepared a revision accordingly. Please see below for an itemized list of your feedback and a brief description of how each is addressed in the revised manuscript.

[R1]:

[https://drive.google.com/file/d/1aqSkIMnX73ONtoIMo0yttD9K5qx4h76L/view?usp=drive\\_link](https://drive.google.com/file/d/1aqSkIMnX73ONtoIMo0yttD9K5qx4h76L/view?usp=drive_link)

*[R1.1] Limited detail on IRB/ethics and survey methodology*

**Response:** The goals of the survey have been updated so as not to need IRB approval. First, all scientists surveyed are surveyed in the capacity that they are providing professional/design feedback, rather than studying some notion of human behavior. Second, all non-scientists surveyed will be taken from 1) the CSCI2370 course or 2) BIOL1430 course (population genetics). These courses are responding with feedback and opinions about their view of “better” visualization practice. Since these results are strictly qualitative and feedback-based, we do not need IRB approval. Should an advising board (ie., Prof. Laidlaw) think this would need IRB approval, we will submit an application at the soonest possible time.

*[R1.2] Feasibility of SHAP for nonparametric UMAP is nontrivial*

**Response:** The proposal was edited to add further clarity on how SHAP will be applied to UMAP values. More specifically, we plan to create [parametric UMAP embeddings](#) of the initial data— in effect, creating a neural network that can approximate the UMAP embedding for the data— and calculate SHAP values from these results. This enables us to conduct SHAP-based analysis “on” the UMAP approximation, providing greater insight into biological drivers.

*[R1.3] The timeline may be a bit tight for tool building with human-subjects evaluation.*

**Response:** The timeline has been adjusted to begin reaching out to human subjects earlier. Further, the exact groups targeted for interviews were clarified and the role of the IRB was discussed.

*[R1.4] Consider adding a minimal statistical analysis plan, pre-registering survey prompts/metrics, and defining success criteria.*

Added discussion on prompts and success criteria to survey discussion. Due to qualitative nature of survey, no statistical analyses were included.

[R2]:

[https://drive.google.com/file/d/11titolzeAjnxOGFPoOJFKclzpC8gUUaN/view?usp=drive\\_link](https://drive.google.com/file/d/11titolzeAjnxOGFPoOJFKclzpC8gUUaN/view?usp=drive_link)

*[R2.1] The general plan for the proposal seems a little vague on some of the weeks.*

Added detail to the weekly plan, including:

1. Subgoals in both the UMAP and PCA creation pipelines
2. Subgoals in the multiview discussion
3. Questions to include in the surveying process

[R3]:

[https://drive.google.com/file/d/1lr2QXoVIQhOBxeGXO1Ncil06puT0hbHd/view?usp=drive\\_link](https://drive.google.com/file/d/1lr2QXoVIQhOBxeGXO1Ncil06puT0hbHd/view?usp=drive_link)

*[R3.1] Limited innovation in visualization techniques.*

Visualization advances were clarified, particularly in comparing the built-in visual interpretations to present options.

[R4] [https://drive.google.com/file/d/1m4lVsT2ZKcBj49lDb02kXoP6bddq-lvE/view?usp=drive\\_link](https://drive.google.com/file/d/1m4lVsT2ZKcBj49lDb02kXoP6bddq-lvE/view?usp=drive_link)

*[R4.1] Only the visualization is novel*

Adds discussion on novelty aspect of the guidelines for ethical visualization, adds discussion on the novel discoveries that can be driven in the biology field (ex. sharing preliminary findings on drivers of differentiation in the Y chromosome)

*[R4.2] Authors could think a bit deeper about what their "very broad" impacts may be. In terms of the public eye, what makes the visualization of this data so important?*

We appreciated this comment. We greatly expanded discussion of the project's societal relevance. Draft revision:

Visualizations of human genetic variation have powerful cultural and ethical implications, as they often shape how both scientists and the public perceive population differences.

Misinterpretations of such plots have historically fueled deterministic or racialized narratives. By developing interactive, context-rich visualization tools that highlight uncertainty and explain underlying features, our work promotes accurate understanding of human diversity and responsible communication of genomic data. These visualizations therefore serve not only as analytical tools for researchers but also as a safeguard against public misuse or oversimplification of population structure.

*[R4.3] Figure 1 could be bigger*

Figure 1 was made larger (ie., now 80% of page width).

[R5] (additional student review?)

[https://drive.google.com/file/d/1wQmAKpwVlyFPVepVIWXUyRDGKnav9vUI/view?usp=drive\\_link](https://drive.google.com/file/d/1wQmAKpwVlyFPVepVIWXUyRDGKnav9vUI/view?usp=drive_link)

*R[5.1] Survey analysis could be described more rigorously*

See [R1.1], [R1.4].

*R[5.2] Ethical outcomes depend on the quality and diversity of survey participants*

The authors made explicit note of this relationship between survey quality and participant diversity and engagement.

# Socially-Responsible Visualization: Creating Context-Driven, Multi-View Clustering Methods for Population Genetics

Skylar Sargent Walters  
PI

skylar@brown.edu

Kyle Wisialowski  
Co-PI

kyle.wisialowski@brown.edu

Sohini Ramachandran  
Collaborator

sramachandran@brown.edu

Alex Diaz-Papkovich  
Collaborator

alex\_diaz-papkovich@brown.edu

October 13, 2025

## Abstract

We present a novel method for visualizing human population genomic data through interpretable, multi-view PCA/UMAP plots, integrating scientific context and minimizing risk of misinterpretation. We incorporate interactive annotations and alternative-view dashboards to highlight contributing variation and clustering variability, enabling responsible exploration. Effectiveness was measured via surveys of researchers and non-scientists, resulting in a new guideline for human population visualization.

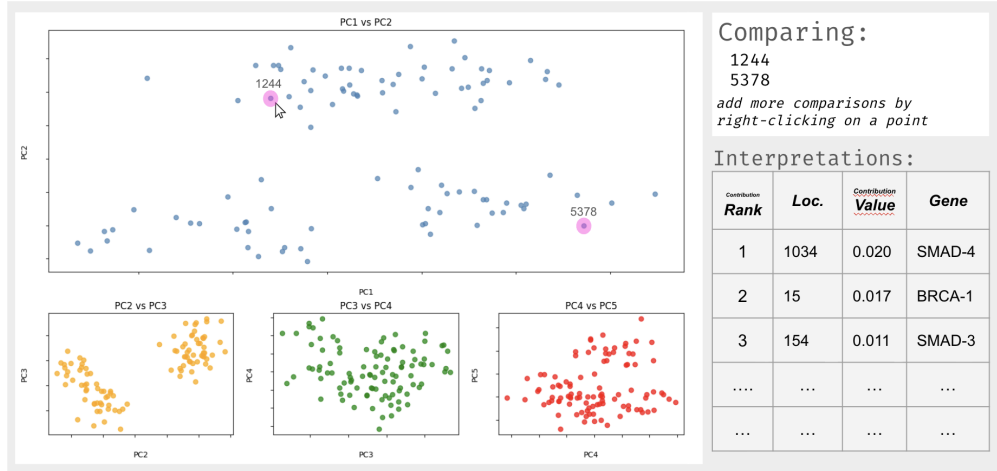


Figure 1: Mock Dashboard

## 1 Aims

We aim to extend the standard PCA/UMAP visualizations used in population genetics to incorporate greater scientific context and emphasize the variability in a single cluster plot. In assessing these advances, we seek to create a list of guidelines describing how to create socially-responsible human population genomics visualizations. Therefore, our aims include:

1. Integrating hoverable interpretations explaining differentiating features between points, a novel visualization framework,
2. Adding alternative views below main plot,
3. Surveying a group of non-scientists to determine the robustness to misinterpretation in standard UMAP/PCA vs. proposed method, and
4. Creating a novel set of guidelines based on results of survey.

### 1.1 Hoverable Interpretations and Interpretation Panel

To create hoverable interpretations in PCA and UMAP plots, we will calculate feature contributions for each selected data point using SHAP values for UMAP and component loadings for PCA. When a user selects a series of points, the corresponding side panel will add the selected point to the interpretation table. This table displays the most contributing loci to the locations of each point. By linking these annotations to known population and functional genomic features, users can gain insight into genetic structure, therefore enhancing interpretability, highlighting biological drivers, and reducing the risk of misinterpretation.

Calculating SHAP values directly from UMAP projections is not possible. However, we can create approximations of the UMAP projections with neural network embeddings [8]. These embeddings enable us to run the UMAP embeddings through the standard Shapley value mechanism, then yield an interpreted result.

In a preliminary study, we created a PCA of the Y chromosome data from the 1000 Genomes Pilot data [3] (Fig. 2). Here, we were able to use loading data (left) to understand the most contributing values

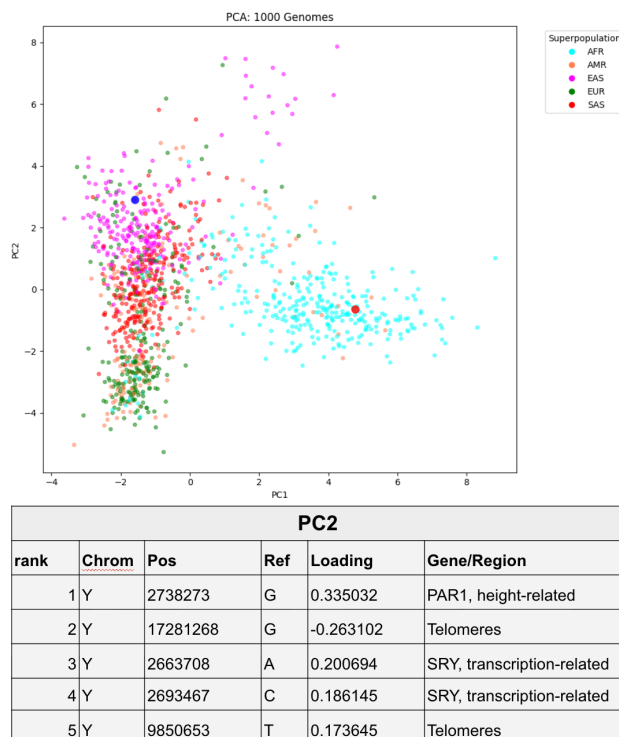


Figure 2: Preliminary PCA interpretation results for PC2

to family differentiation. These scores and mutant positions are annotated with their corresponding genes, providing an example of new biological information that can be found through this method. Additionally, we were able to gain information on the most contributing PCs between the selected points, which were PCs 2, 3, and 19. Incorporating these findings into the results will create a better understanding of what each PC could represent as well as factors explaining differentiation.

## 1.2 Multiview Design

The multi-view design provides additional context by presenting supplementary views beneath the main plot. These views contain alternative embeddings from subsampled datasets or alternative PCs, capturing the variability in points across samplings. This layered approach reinforces the concept of dimensionality reduction and clustering being a "fuzzy" source of information, further preventing misuse.

## 1.3 Evaluation

To evaluate these enhancements, we will conduct a survey of scientific and non-scientific participants. Three domain experts will review the tool from a genomics perspective, providing insight into how built-in interpretability impacts the exploration process as compared to standard UMAP/PCA. Planned survey questions include:

1. Preliminary question: How do you use UMAP and PCA to visualize human genomic data? What methods do you use to interpret these results, if at all?



2. How does this tool compare to standard UMAP/PCA in ability to interpret results? Does it provide any additional or reduced granularity in ability to understand contributing genes or genomic features?
3. How could you deliberately misinterpret the results shown in these visualizations?

Meanwhile, non-scientific participants will be given both standard plots and enriched plots and asked to state 1) their analysis of the image and 2) attempt bad-faith misinterpretations. This will therefore give us feedback from both an exploratory lens from the researchers and an explanatory lens from the participants. Participants in this survey come from Brown’s BIOL1430 course, a class on mathematical theory of population genetics. These students represent a number of academic backgrounds and experiences, bearing in mind that the quality of a survey is dependent on audience diversity.

All results in this study are qualitative. Thus, these qualitative observations about preferred visualization strategies and elements will also form preliminary guidelines on creating best practices for visualization of human genetic variation. This consequently provides a novel, practical resource for researchers and reinforce a socially-responsible scientific environment.

## 2 Significance

This work addresses two critical gaps in population genetics research: deliberate misinterpretation and clustering interpretability. Traditional methods like PCA, t-SNE, and UMAP often exaggerate population differences, which can lead to misinterpretation and the reinforcement of harmful narratives about biological race [1]. By integrating multiple views of a visualization, we support that these images are highly variable, reinforcing the idea that cluster boundaries and population separations are not fixed biological truths but are sensitive to analytic choices, data sampling, and methodological parameters. This approach helps viewers recognize the inherent uncertainty and complexity in genetic data, discouraging oversimplified or deterministic interpretations of human population structure.

This problem is somewhat fed by the problem of interpretability: standard visualizations will present clusters as discrete, biologically-significant groupings, yet the “reasoning” behind assignment of individuals to clusters is left out of the visualization. Likewise, methods like UMAP exaggerate distinctiveness and produce clusters whose global relationships do not always correspond to true genetic similarity, leading to risks of misinterpretation for both scientists and the public [9]. In directly integrating interpretability into the visualization, we provide direct and actionable insight into cluster decision-making, enabling viewers to understand what genetic features drive form. Critically, by making these feature contribution explicit, we also promote the exploration of areas of understanding human genetic structure, therefore contributing to biological discovery in the process.

This work therefore has three sources of novelty. First, the visualization is novel in its integration of hoverable interpretation and multiple views to display human genomic data. Next, it presents novel biological knowledge in the outputs of the interpretability features. Finally, the suggestion framework for responsible visualization is a novel scientific and sociological guideline.

We anticipate that this interpretable dashboard framework can be adapted to fields outside of population genetics, acting as a broader technical contribution.

## 3 Related Work

Clustering and dimensionality reduction are critical tools in visualizing human population data. These are usually computed over genotype arrays, which show any SNPs (single nucleotide polymorphisms, or

mutations) over whole populations. Conducting a dimensionality reduction over this matrix can display relationships in shared mutations and alleles, reflecting lineage and different human genotypes [6].

Principal component analysis (PCA) and UMAP are widely-used methods for this task. PCA projects high-dimensional data onto orthogonal axes (or principal components) that capture the greatest variance, enabling the visualization of variably similar elements. Since PCA is a linear algorithm, it often fails to capture nonlinear relationships. Depending on the principal components being used, the outputted visualizations may have strong differences in view [5].

More recently, UMAP has emerged as a dimensionality reduction alternative. UMAP aims to preserve local and global structures in a population, but the quality and usefulness of the outputted visualization is highly dependent on hyperparameter choices. This element of hyperparameter-based variability can make UMAP harder to interpret directly [4].

This method builds on the preexisting use of UMAP and PCA and seeks to fill in the gaps in their abilities. Critically, neither have explicit interpretations built into their visualization systems, proliferating the problems of lack of scientific context. Furthermore, both of these models can have strong visuals change depending on hyperparameters or the PCs in use. Acknowledging this limitation of the models via multi-view integration enables a greater understanding of this shortcoming.

Past work in improving these visualizations has incorporated integrated interpretation. GAUDI (Group Aggregation via UMAP Data Integration) is a recent unsupervised method for interpretability of multi-omics data visualizations [2]. GAUDI applies UMAP independently to each inputted omics dataset, then concatenates the embeddings and applies a second UMAP is applied to the dataset, allowing for the identification of underlying biological patterns that may not be evident in higher dimensions. For interpretation, they employ SHAP values [7], which correspond to contribution in the output location. Our work differs in that we 1) extend interpretability to also include PCA and 2) apply our method to population genetics.

## 4 Research Plan

Week 1:

- Set up Python environment, Jupyter notebooks with numpy, pandas, scikit-learn, seaborn, UMAP, etc
- Download and preprocess dummy 1000 Genomes Project dataset CITE
- Generate baseline PCA, UMAP visualizations
- **Deliverables:** baseline PCA, UMAP visualizations

Week 2:

- Generate alternative views of PCA/UMAP visualizations by resampling data
- Automate pipeline for generating the multiview visualizations
- **Deliverable:** pipeline for multiview generation

Week 3:

- Create interpretability portal: PCA loading scores
- Time-dependent, start looking at integrating SHAP values for UMAP embeddings

- **Deliverable:** demonstrate that interpretability portal works on PCA

Week 4:

- Create interpretability portal: UMAP-based SHAP values. First, we create parametric UMAP embeddings of the initial data. Then, calculate SHAP values from these neural network-based results, in effect creating a contribution score-based system for interpretation.
- Start of week: contact possible interviewees for project (**Scientific audience confirmed, BIOL1430 class confirmed**) **Scientific sources:** Professors in the DSI (Dan Weinreich, Emilia Huerta-Sanchez, associates) **Non-scientific sources:** Students in BIOL1430 (Population Genetics), students in BIOL0200 (general biology-aware audience).

Week 5:

- Survey interviewees for project
- See Section 1.3 for question list
- **Deliverable:** Guidelines for human population genomic visualization based on interview responses

Week 6:

- Draft final paper/presentation/demonstration
- **Deliverable:** Finished paper/presentation

Facilities:

- Computational facilities: Standard laptop, OSCAR if computational power needed
- Data: 1000 Genomes Project
- Survey Search - Domain experts: email collaborators for contacts, reach out to professors who have taught population genetics courses
- Survey Search - Non-scientific: email to Today@Brown

## References

- [1] J. Carlson, B. M. Henn, D. R. Al-Hindi, and S. Ramachandran. Counter the weaponization of genetics research by extremists. *Nature Genetics*, 2022.
- [2] P. Castellano-Escuder, D. K. Zachman, K. Han, and M. D. Hirschey. Interpretable multi-omics integration with umap embeddings and density-based clustering. *bioRxiv (Preprint)*, 2024.
- [3] T. . G. P. Consortium. A map of human genome variation from population-scale sequencing. 2010.
- [4] A. Diaz-Papkovich, L. Anderson-Trocme, and S. Gravel. A review of umap in population genetics. *Journal of Human Genetics*.
- [5] E. Elhaik. Principal component analyses (pca)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, 2022.
- [6] S. Hendricks, E. C. Anderson, T. Antao, L. Bernatchez, B. R. Forester, B. Garner, B. K. Hand, P. A. Hohenlohe, M. Kardos, B. Koop, A. Sethuraman, R. S. Waples, and G. Luikart. Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, 2018.
- [7] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. 2017.
- [8] T. Sainburg, L. McInnes, and T. Q. Gentner. Parametric umap embeddings for representation and semi-supervised learning. 2020.
- [9] F. Thor and C. Nettelblad. Dimensionality reduction of genetic data using contrastive learning. *Genetics*, 2025.

# Skylar Walters

69 Brown St, Box 3979 | Providence, RI 02912 | Phone: (973) 610-8724 | E-Mail: skylar\_walters@brown.edu

## EDUCATION

**Brown University, B.S. Computational Biology, Computer Science Track** Providence, RI | **Grad. May 2026**

- **Coursework:** Deep Learning, Computational Biology, Statistics and Inference in Genomics, Genetics, Data Structures & Algorithms, Linear Algebra, Statistical Inference I, Biotechnology in Medicine, Organic Chemistry

## EMPLOYMENT

**Humans in Data Integration, Visualization, and Exploration (HIDIVE) Lab, Intern** Boston, MA | June-Oct. 2025

- Developing a system to collect and analyze Gosling visualization specifications for genomics data, enabling task-driven visualization design and recommendation.
- Leveraging a formal taxonomy of genomics data, tasks, and visual encodings to support systematic analysis and potential machine learning applications.
- Associated with Harvard Department of Biomedical Informatics Summer Institute in Biomedical Informatics

**Surveillance of Infectious Disease Lab (SIDL), Research Assistant** Providence, RI | Aug. 2023-Present

- Develops computational techniques to detect defective viral genomes (DVGs) from large metagenomic datasets (1 billion+ data points) as a means of understanding mechanism by which DVGs trigger antiviral host responses
- Applies recombination mapping techniques to Ebola and SARS-CoV-2 datasets via shell scripts
- Granted \$2,500 Brown University Undergraduate and Teaching Research Award to continue research
- Present at weekly journal club meetings to gain proficiency in science communications and group collaboration

**Brown University, Deep Learning and Viral Metagenomics Researcher** Providence, RI | June-Aug. 2024

- Awarded NIH R25 grant to design a deep learning-based approach to identify novel and highly divergent viral genomes free of reference sequences, providing an alternative diagnostic tool for unidentified illnesses
- Adapts NLP techniques, variational autoencoders, and clustering algorithms to process 15,000+ viral and bacterial whole-genome sequences using Keras, h5py, TensorFlow, Scikit-Learn, and more
- Constructed latent-layer PCA models, weighting schema to reduce bias in data, and plots of loss and accuracy per epoch

**Brown University Department of Computer Science, Meta Teaching Assistant** Providence, RI | Oct. 2024-Present

- Collaborates with Director of Undergraduate CS Studies to organize the daily activities of the Computer Science Department
- Oversees hiring of over 300 TAs and Head TAs and organizes department-wide TA trainings
- Scripts algorithms to streamline hiring, evaluations, and room booking

**Brown University Department of Computer Science, TA & Head Teaching Assistant,** Providence, RI | Aug. 2023-Present

- Facilitates dynamic learning environment for 100 students in Computer Science: an Integrated Introduction
- Liaisons between students, Professor Philip Klein, and team of 11 Teaching Assistants to ensure course runs smoothly
- Leads development and launch of weekly homework assignments, while hosting office hours, labs, and grading sessions

## PEER-REVIEWED PUBLICATIONS

**Sargent Walters, S., Valderrama, A., Smits, T., Kouřil, D., Nguyen, H., L'Yi, S., Lange, D., Gehlenborg, N., "GQVis: A Dataset of Genomics Data Questions and Visualizations for Generative AI," In Proc, IEEE VIS, 2025**

## CAMPUS/COMMUNITY INVOLVEMENTS

**Appalachian String Ensemble, Banjo & Mandolin** Providence, RI | Sept. 2022-Present

- Studies traditional Appalachian folk music in small and large ensembles, including Brown's Old Time String Band
- Trained to learn and teach tunes by ear, continuing centuries-old tradition of Old Time Music

## AWARDS

|                                                                                                    |           |
|----------------------------------------------------------------------------------------------------|-----------|
| <b>Barry Goldwater Scholarship</b>                                                                 | Mar. 2025 |
| <b>Annual Biomedical Research Conference for Minoritized Students, Computational Biology Award</b> | Nov. 2024 |
| <b>Research Grant via NIH R25 Student Grant</b>                                                    | June 2024 |
| <b>Brown University Undergraduate Teaching and Research Award</b>                                  | Nov. 2023 |
| <b>United States Presidential Scholar</b>                                                          | May 2022  |
| <b>Coca-Cola Scholar</b>                                                                           | May 2022  |

# Kyle Wisialowski

☎ (860) 800-4466 | ✉ [kyle\\_wisialowski@brown.edu](mailto:kyle_wisialowski@brown.edu) | 🌐 <https://github.com/kylewisialowski> | 💼 <https://www.linkedin.com/in/kyle-wisialowski-0551121a9/>

## EDUCATION

- Brown University, Sc.M. Computer Science** Providence, RI | **Expected Graduation May 2026**
- Current Coursework: Deep Learning, Human-AI Interaction, Interdisciplinary Scientific Visualization
- Brown University, Sc.B. Applied Math - Biology, 3.88/4.0 GPA** Providence, RI | Class of 2025
- Relevant Courses: Software Engineering, Foundations of Artificial Intelligence, Program Design: Data Structures and Algorithms, Statistical Inference I, Ordinary Differential Equations, Multivariable Calculus, Linear Algebra, Logic for Systems
- Old Saybrook High School, 103.8/100 GPA** Old Saybrook, CT | Class of 2021
- Salutatorian, National Merit Commended Scholar, Bausch + Lomb Science Award, Varsity Cross Country, Swim, & Crew

## PROFESSIONAL EXPERIENCE

- Center for Computation and Visualization, Software Development Intern** Providence, RI | June 2025 – August 2025
- Developed a parsing algorithm using ANTLR to extract parameter information from commonly used biological workflows
  - Integrated a self-hosted Ollama Large Language Model into the above parsing algorithm using Unsloth and Hugging Face to autonomously annotate workflow parameters
  - Trained a Llama 3 LLM to understand prompting with a niche biological workflow language (Workflow Development Language)
  - Built a Docker Container to host the ANTLR + LLM software, deployed through GitHub Actions
  - Restructured the above Docker workflow into a script compatible with Brown's High Performance Computing cluster
  - Automated cross-referencing between 2 genomic databases with dynamically-updated React dropdown components
- Center for Computation and Visualization, Research Assistant** Providence, RI | June 2024 – May 2025
- Student researcher under Dr. Katherine Siddle's SARS-CoV-2 genome study with Brown's Computational Biology Core
  - Hosted and edited Nextstrain phylogeny software on Brown's High Performance Computing cluster to produce phylogenetic analyses for a dataset of 9 million SARS-CoV-2 genomes
  - Merged a subsampling pipeline and a phylogenetic pipeline into a combined script with a Terminal User Interface — reducing script-running workload for Dr. Siddle and a doctoral student by 50%
  - Developed R-analyses for phylogenetic data (temporal distribution plots, frequency distribution plots, variance calculations)
  - Synthesized results in semesterly presentations for a group of researchers, doctoral students, and undergraduates
- Brown University Department of Biology, Teaching Assistant** Providence, RI | January 2024 – May 2024
- Graded weekly lab write-ups and provided feedback for students in BIOL 0285 (Inquiry in Biochemistry)
  - Assisted students during lab exercises, including gel electrophoresis, PCR reactions, protein induction, and kinetics assays
  - Helped students produce Michaelis-Menten and Lineweaver-Burke plots to interpret their results
- Pods Swimming, Head Instructor** Providence, RI | March 2023 – August 2024
- Taught weekly lessons in water safety and basic swimming techniques for kids aged 3-9, teens, and adults
  - Cultivated a fun learning environment where students mastered lifesaving skills like floating/kicking comfortably on their back
- College Matters, Summer Intern** Madison, CT | June 2022 – August 2022
- Used Excel, Unito, Trello, and Zapier to automate the client onboarding process for a college application consulting firm
  - Re-developed an existing protocol to efficiently store and organize consultant notes from college visits

## EXTRACURRICULAR/VOLUNTEER EXPERIENCE

- Club Swim Team, Co-Captain** Providence, RI | September 2021 – Present
- Help lead a team of 11 board members and 150+ swimmers
  - Manage team finances, travel logistics, practice schedules, and social events
  - Attended CCS National Championship in 2024/2025 and team record holder in 100 fly, 200 fly, 100 IM, and 200 IM
- Hermes Free Telehealth, Medical Outreach Team Member** Virtual | June 2022 – May 2024
- Assisted with outreach for a student-led telehealth nonprofit working to increase access to healthcare for underserved populations
  - Attended weekly meetings to brainstorm methods for recruiting volunteer health professions
- Applied Math Directed Reading Program, Participant** Providence, RI | September 2023 – December 2023
- Selected to participate in Brown's independent reading program lead by the Division of Applied Mathematics
  - Studied mathematical models of disease under the guidance of Dr. Juniper Cocomello
  - Worked with another student to develop a presentation about the theory, application, and modification of the SIR model
- Go Baby Go, Constructor/Materials Coordinator** Old Saybrook, CT | August 2020 – August 2021
- Partnered with the Go Baby Go organization to provide adapted ride-on to cars for children with mobility challenges
  - Gathered building materials and re-wired cars, then met with children and families to make modifications for specific needs

## SKILLS & INTERESTS

**Technical Skills:** Proficient in Python, Java, Linux/Unix environments, Javascript/Typescript, React, Matlab, R, data visualization  
**Language:** High School Seal of Biliteracy in Spanish (elementary proficiency)  
**Interests:** Rollercoasters, trying new food, pickleball, skiing, sunsets, anything involving a body of water

**Alex Diaz-Papkovich, Ph.D.**

Alex Diaz-Papkovich is a post-doctoral researcher at Brown University's Data Science Institute. He has broad interests in statistics and data science, particularly in the world of genetics and biology. His doctoral research was with Simon Gravel's lab studying structure in population genetics and how to use dimensionality reduction to understand biobank data and improve inference. He is also interested in using data from single-cell RNA sequencing, neuroinformatics, and biological data.

**Sohini Ramachandran, Ph.D.**

Sohini Ramachandran's research focuses on understanding human population genetics and evolutionary history through computational and statistical approaches. Her lab studies how migration, adaptation, and demographic events have shaped genetic diversity across global populations. She also develops models and algorithms to infer population structure and demographic parameters from genomic data. Beyond humans, her work extends to pathogens and other species to explore how evolutionary forces operate across systems.

She joined the faculty of Brown University in July 2010, and was named the Hermon C. Bumpus Professor of Biology in January 2023; she also holds a courtesy faculty appointment in Computer Science.



Diaz-Papkovich, Alex

to me ▾

Tue, Sep 23, 9:36 AM



Hi Skylar,

This looks really interesting and like a worthwhile direction, thanks for sending me the slides. Prof. Laidlaw has a PhD student that's working on a similar idea (Jasmine Liu, who is looking at subsets of SNPs and seeing how that impacts visualization), though her project is larger in scope and thus incorporates many other measures and approaches. Measuring which SNPs give weight to points in embeddings is also something I've been curious about, as well as interactive visualizations and subclustering. With respect to the first point, I think it's an underexplored area ([here](#) is a paper that tries to find how much influence alleles have within UMAP embeddings; there are a smattering of other papers but I think this is the most relevant for you). If you'd like some code and data, I have a repo that has some interactive plots using plotly that you may find useful: [https://github.com/diazale/1KGP\\_dimred](https://github.com/diazale/1KGP_dimred). I also have a version of the 1KGP data on [dropbox](#) that could save you some pre-processing time in the notebook.

Also for what it's worth, Otto (one of your classmates) has spoken to me about doing a project related to my presentation as well---his is not as thought out as yours at this point, but again it could be worth discussing in the case the two of you find some common ground.

Thanks again for the slides, and let me know if you have any more questions.

Alex.



# Immersive 3D Modeling of Spatiotemporal Transmission Grid Dynamics over Network Graphs

Yonas Amha  
PI

yonas-amha@brown.edu

Korey Sam  
Co-PI

korey-sam@brown.edu

Kenny Gruchalla  
Collaborator

kenny\_gruchalla@nrel.gov

Rahul Kadavil  
Collaborator

rahul\_kadavil@nypa.gov

October 16, 2025

## Abstract

We present a novel and immersive 3D method for visualizing the dynamics of the transmission grid and the spatio-temporal dynamics of the transmission grid data, with a specific emphasis on illustrating how network disturbances affect voltage in the wider electrical grid. Our visualization harnesses data from the Western Electricity Coordinating Council (WECC) and depicts transmission grid variables (frequency, voltage, fluctuations, etc.) in an intuitive manner for users with a non-technical electrical engineering background.

Link to the manuscript pdf:

[https://docs.google.com/document/d/1QmM\\_OQI\\_EuJDu-ZKafb\\_RBOeFDFLxIAKVxMdoIYGP1g/edit?tab=t.0](https://docs.google.com/document/d/1QmM_OQI_EuJDu-ZKafb_RBOeFDFLxIAKVxMdoIYGP1g/edit?tab=t.0)

## Response Letter

Dear Editor and Reviewers,

We appreciate the constructive feedback on the manuscript, and we have prepared a revision accordingly. Please see below for an itemized list of your feedback and a brief description of how each is addressed in the revised manuscript.

[R1.1] The researchers offered a very clear explanation about the purpose of mobilizing the VR system: why it is necessary to make use of it, and what advantage it will have over the traditional methods. However, the authors can certainly benefit from explaining more about how their approach will be different from previous works - how would they benefit from previous works, and how will they improve them? I think this part is not clearly stated in the related works section.

**Response:** This is a very valid point that I will factor into the next revision of the proposal. The main way in which I am trying to improve upon previous works is to automate 2D designs with timestamps and immersive VR modeling of spatiotemporal interactions and variables.

[R1.2] Most of the envisioned scientific components of the project are solid, especially in simulating WECC dynamics and electricity flow in various regions. The researchers connected these components effectively. However, I found the methodology and evaluation design is slightly undermining the scientific merit - the definition of "intuitiveness" (as a part of aims) is not defined clearly enough. I would love to learn more about the methodology, expected difficulties, and the metrics for subjective evaluations a bit more.

**Response:** The intuitiveness metric is going to be determined in week 1-2 of the proposal. It is a little hard to ascertain what exactly we are looking for in terms of intuitiveness because the immersive VR design hasn't been fleshed out yet.

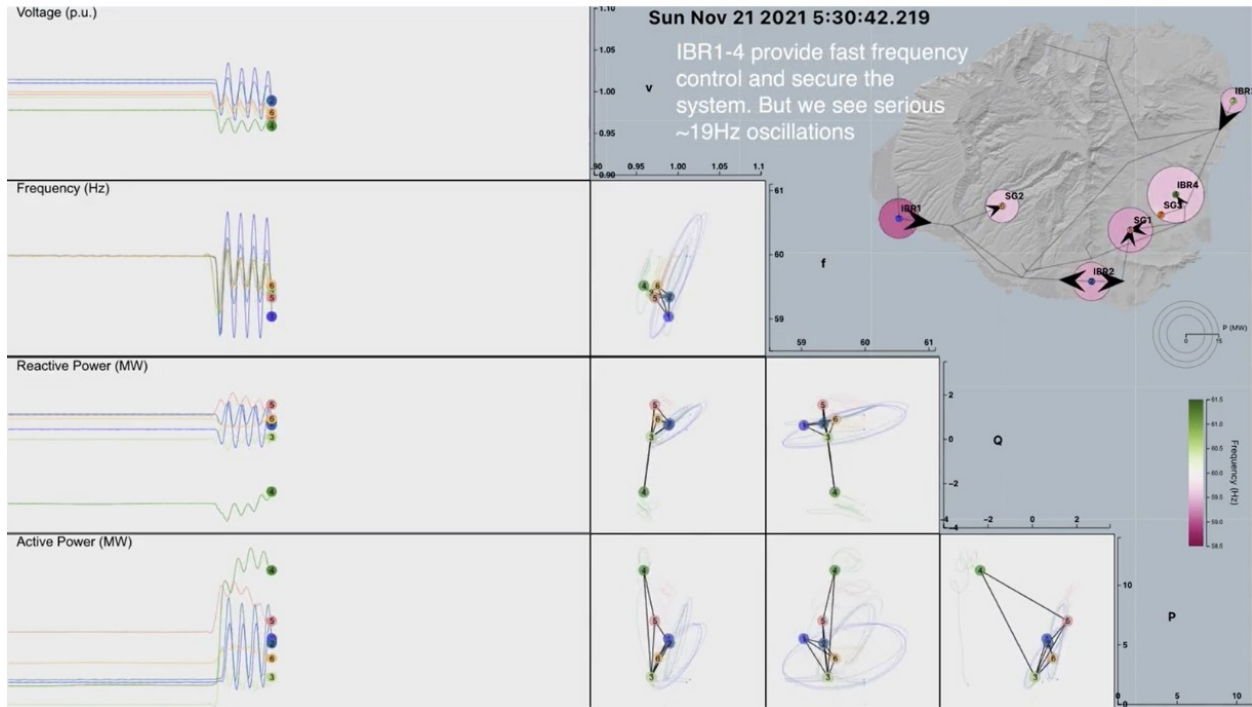
[R2.1] The research plan is not just ambitious; it's impossible. Planning to recruit, test, and analyze feedback from 50-100 end users within a single week is not feasible for a project of this scale. In my opinion, this suggests a major misunderstanding of the time required for human-subject studies.

**Response:** The number 50-100 is a little unreasonable. I will amend the amount of users to a more suitable number, but for the purposes of this proposal, the more individuals the better. The amount of end users is meant to be more of goal rather than a necessary benchmark.

[R3.1] The proposal is novel because, based on the text, immersive VR in this context is uncommon. The proposal could be strengthened with a description of the "non-extensive prior work" in the field beyond CSLens, even if it is not as developed. This additional detail could help reinforce novelty and provide helpful background.

**Response:** I have added additional information to the novelty section to address this shortcoming. It was well spotted.

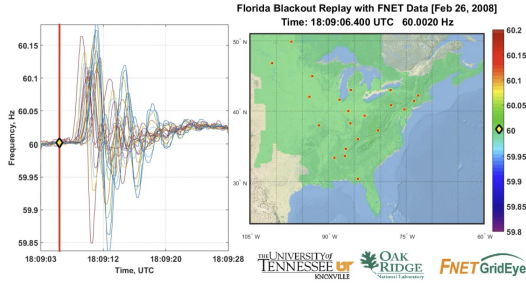
# 1 Background



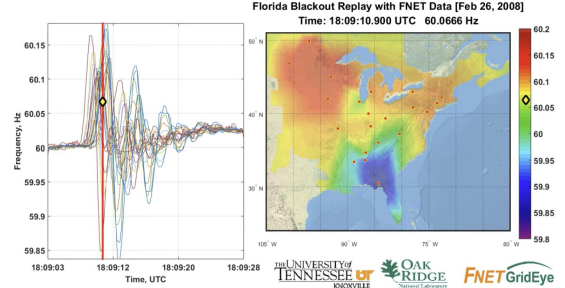
**Figure 1:** The image above depicts a current two-dimensional prototype for visualizing spatio-temporal transmission grid dynamics on island of the Kauai in the Hawaiian archipelago. The developers of this visualization are Kenny Gruchalla and Rahul Kadavil, both of whom are collaborators on this project. This visualization will serve as the inspiration for this project as the variables encoded are almost identical those that we plan to implement in our immersive 3D visualization.

In the past, transmission grid dynamics has traditionally been depicted through 2D visualizations. However, these visualizations have a few major shortcomings. The first is that 2D visualizations tend to model solely a single variable, almost always frequency, rather than modeling multiple variables at once. Modeling multiple variables allows us to gain understanding how their interactions influence widespread grid behavior. The second is that 2D visualizations are extremely difficult for users to interpret, especially stakeholders from a nontechnical background, which can be a major issue when stakeholders rely on these 2D visualizations to make real-time decisions regarding power distribution and blackouts. Traditional 2D visualizations often lead to miscommunications between stakeholders from a technical background and users from a non-technical background. These miscommunications can be extremely costly for companies which could end up in delayed decision making and losses of millions of dollars.

Additionally, 2D visualizations do a poor job of overlaying grid variables over network topography in a dynamic manner. Almost all visualizations that overlay grid variables over network topography use clusters or heatmaps to indicate the location and size of those variables.



**Figure 2:** 2D heatmap visualization of the 2008 Florida Blackout prior to the power outage



**Figure 3:** 2D heatmap visualization of the 2008 Florida Blackout during the power outage

Shown above is a traditional 2D visualization that illustrates the 2008 Florida blackout. The major shortcomings are traditional 2D visualizations highlighted earlier are clearly shown in this visualization. The visualization only models frequency rather than modeling multiple variables and interplay those dynamic grid variables. Additionally, although the heatmap visualization shows the cascading domino effect of the Florida outage on the major U.S. grid, the 2-dimensional nature of the visualization does really capture the dynamic nature of the grid variables on the network topography. Our immersive visualization will build upon these shortcomings and create an immersive and intuitive experience for the user.

## 2 Aims

The proposal's objectives are divided into two categories, primary goals and stretch goals, that are clearly delineated in the outline provided.

### Primary Goals

1. *Accurately use the dynamic variables present within the WECC metadata (frequency, voltage, fluctuations, active power, reactive power, and topological data) to depict multivariate spatiotemporal interactions in transmission grid dynamics through immersive three-dimensional visualizations.*
2. *Create an intuitive visualization that improves on the shortcomings of two-dimensional visualizations.*
3. *Visualize the effect of local disturbances on nationwide grids, with an emphasis on how said disturbances influence the power flow and voltage fluctuations in various regions.*
4. *Develop the immersive visualization through a desktop/web application that has an intuitive design and is easily exportable to end users*
5. *Develop a visualization prototype that encodes multiple variables (frequency, voltage, power, reactive power, active power) simultaneously across the network topography*
6. *Utilize the grid data to achieve a better understanding of behavior, stability, and vulnerabilities with power grids*
7. *Limit the effect of background noise (topological disturbances, frequency oscillations) on the data projections from the visualizations.*

8. *Conduct field testing on 5-10 end users (most of whom hail from a non-technical background in electrical engineering) to ascertain the intuitiveness of the visualization as well a quantitative and qualitative analysis. End users are a culmination of electrical engineers and other specialists that have ties to the New York Power Authority*

### **Stretch Goals**

1. *Expand upon the three dimensional immersive visualization by implementing a large-scale visualization that will be designed for a Cave Automatic Virtual Environment (CAVE) that accurately utilizes the WECC metadata.*
2. *Conduct field testing on a further 5-10 users, most of whom hail from a non-technical background in electrical engineering*

## **3 Significance**

The significance of the work put forth by this proposal is noteworthy on several fronts. Immersive 3D visualization would improve upon existing two-dimensional models by encoding multiple different variables to dynamically model multivariate spatiotemporal interactions. Three dimensional immersive visualization allows the end users to focus in on distinct variables within the visualization and interact with those variables and the network topography to develop a deeper understanding of how the topography interacts with power fluctuations and larger grid disturbances. In other words, this method of visualization can build a better understanding how voltage, frequency power, active/reactive power, current flow, and network topography are interconnected, and how distinct behavior in one variable can influence the remaining variables.

From the perspective of the end users, one of the current major complaints about existing two-dimensional visualization is that it's difficult for those with a non technical background to interpret the data being shown from the modeling. This leaves a disconnect between the end users, many of whom come from a nontechnical background, and the electrical engineering specialists who come from a well verse technical background. Transferring the visualization into an immersive 3D environment would make it easier for end users to develop intuition surrounding spatiotemporal interactions and their effect on grid dynamics and be better in tune with electrical engineers and other stakeholders hailing from a technical background.

Additionally, an immersive 3D visualization may be more effective in helping researchers understand the cascading effect of small disturbances and how they ultimately influence the large transmission grid chain. According to Kenny Gruchalla, the Principal Scientist at the National Renewable Energy Laboratory, and Rahul Kadavil, the AGILe Manager at the New York Power Authority, one of the major areas that two-dimensional visualization struggle to visualize is the cause and effect relationship between local disturbances and statewide power dips and oscillations. Our visualization will seek to improve upon providing immersive modeling on local disturbance events so that researchers can zoom in and isolate the connection between those local disturbances and larger transmission grid dynamics.

## **4 Related Work**

While the prior work in this field of study has not been extensive, there are have been a few notable projects that utilized a similar type of visualization of that found in this project. One of these notable projects

was called CSLens, which seeks to implement “a condensed three-dimensional visual representation of fluctuations in traffic hotspots and charging demand” with specific emphasis on overlaying the visualizations analytical variables over the topography of a selected region [1]. The goal of CSLens was to create a dynamic 3-dimensional visual model, fully equipped with analytics that model the affect of traffic on the fluctuation in usage of different charging stations over a certain region. The model contained a “Map View” which overlaid the variables of the model over the topography of a given region, which is extremely similar to the implementation that we are trying to do with our immersive 3-dimensional visualization. While this proposal will look to CSLens for notes of inspiration, there are several ways in which our immersive visualization will differ from that of the former. Our immersive visualization will contain dynamic modeling, rather than capturing the visualization at select points during the day. Additionally, the CSLens visualization does not do a good job encapsulate the cascading affect that local disturbances (a power outage, accident, or abnormal amount of traffic for instance) can have on the behavior of certain variables and affect the larger charging grid network

## 5 Research Plan

The workflow for the proposal will be as follows. There will be milestones inserted throughout the duration of the proposal to make sure that all aspects of the project are moving forward as planned. The duration of the project will be 6 weeks.

- **Week 0:** Final proposal approved and ready for implementation.
- **Week 1:** The drafting an immersive 3D Prototype has been completed. Collaborators have met and discussed the desktop/web application that will house the 3D visualization. The coding repository has been setup and is reading for implementation.
- **Week 2:** Coding for the immersive 3-Dimensional visualization is nearly 25% implemented. Visual projections have been implemented but fine-tuning is needed to accurately portray spatiotemporal grid dynamics.
- **Week 3:** Coding for the immersive 3-Dimensional visualization is nearly 75% implemented. PI’s and Co-PI’s meet with Kenny and Rahul to discuss progress. A feedback survey for the end users testing the immersive visualization has been completed and is ready to be sent to the end users.
- **Week 4.5:** The implementation of the immersive 3-Dimensional prototype application has been completed. The prototype is ready for qualitative and quantitative field testing among end users. The variable data (all of which is contained in CSV and text files) has been fully incorporated into the final design. Live demo with collaborators Kenny and Rahul, followed by active feedback. Field testing to promptly begin as ~5-10 end users will be asked to demo the prototype design and submit feedback.
- **Week 5:** Field testing with end users has been concluded. The feedback has been analyzed by the PI and Co-PIs. The final abstract regarding the findings of field testing is in the drafting process.
- **Week 6:** Final presentation and demonstration ready. The final abstract has been fully completed. Final comments regarding the project are retrieved from Kenny and Rahul.

## 6 Collaborators

# Yonas Amha

Providence, RI | Phone: (703) 772-1887 | E-mail: [yonas\\_amha@brown.edu](mailto:yonas_amha@brown.edu) |  
[linkedin.com/in/yonas-amha-a85a5525b/](https://www.linkedin.com/in/yonas-amha-a85a5525b/) | [github.com/yonasA2025](https://github.com/yonasA2025)

## EDUCATION

**Brown University, M.S. Computer Science** Providence, RI | **Expected Graduation May 2026**  
Relevant Courses: *Machine Learning, Reintegrating AI, Computer Security, Computer Architecture*  
**Brown University, B.S. Computer Science, 3.6/4.0 GPA** Providence, RI | **September 2021 - May 2025**  
Relevant Courses: *Computer Systems, Program Design: Data Structure & Algorithm, Logic of Computer Systems, Cryptographic Systems, Software Security Exploitation, Computational Molecular Bio, Applied Cryptography*

## EXPERIENCE

**HP Inc., DevOps Engineer** Houston, TX | June 2024 – September 2024

- Assigned to the Data Science Scrum Team for the launch of HP AI Studio. Implemented database management, telemetry, and API connection using MongoDB, Azure DevOps, Kubernetes, and GoLand

**Communities Together, Web App Development** New York, NY | July 2023 – Present

- Built a website that prepares Amharic speakers to take the U.S. Civics Test in Amharic. The website is equipped with full text-to-speech capabilities. Website link: <https://usacivictest.web.app/>
- Website uses Azure Text-to-Speech

**Ringle, Tutor** Fairfax, VA | June 2023 – September 2023

- Tutored 18 high school students in weekly English lessons, advising them on grammar, vocabulary and sentence structures
- Assigned a reading for each lesson and sent detailed feedback analyzing student's performance during the lesson along with exercises and worksheets to target areas of weakness

**Second Brain Foods, Product Marketing Intern** Alexandria, VA | June 2022 – September 2022

- Assisted with packaging, presentation and pricing of the health snacks product-line for the startup company
- Conducted market analysis on six barley products within the company's main line focusing on performance and visual appeal

## PROJECTS

**Database | C** [github.com/yonasA2025/database](https://github.com/yonasA2025/database)

- Created a fully functioning database in which the user can store, pop and search for given values
- Implemented thread safety with coarse and fine grain locking

**Amharic Civics Test | TypeScript | JavaScript | HTML | CSS** [github.com/yonasA2025/amhariccivictest-v3](https://github.com/yonasA2025/amhariccivictest-v3)

- Built a website that prepares Amharic speakers to take the U.S. Civics Test in their native language

**Ratty Reviews | JAVA | TypeScript** [github.com/yonasA2025/rattyReviews](https://github.com/yonasA2025/rattyReviews)

- Created a website that allows users to rate the meals that they have at the Ratty (popular Brown dining hall) on any given day
- The website uses web scraping to acquire the meals from [brown.dining.edu](https://brown.dining.edu) and uses caching to store user ratings
- Google authentication was implemented so the user could sign into their gmail to submit a rating

**Maze | C** <https://github.com/yonasA2025/maze>

- Built a 2D maze generator that could generate up to 25x25 size mazes
- Generated solutions to the maze using Breadth First Search (BFS), Depth First Search (DFS), and Dijkstra Algorithms and then encoded the solution in hexadecimal

**Search | Python** [github.com/yonasA2025/search](https://github.com/yonasA2025/search)

- Created a search engine that queries all wikipedia pages for any given search
- Search engine implements PageRank and handles WikiLinks

## LEADERSHIP & PROFESSIONAL DEVELOPMENT

**Brown Ethiopian Eritrean Student Association, Senior Executive Board Member** Providence, RI | October 2022 – Present

- Design Instagram posts, merch, and coordinate with other clubs around promoting events and general board meetings
- Design posts have yielded an 70% increase in engagement for the Instagram page

**Merrone Teklu for Alexandria City Council, Election Campaign Volunteer** Alexandria, VA | May 2021 – July 2021

- Increased public awareness and advertising for Meronne Teklu, an Alexandria City Council candidate by actively engaging in phone banking, election-day campaigning, and distributing campaign flyers
- Participated in bi-weekly phone banking in the lead up to the election and called 100 voters and encouraged them to get to the polls and provided closest voting locations

**Georgia Canvassing Project, Election Outreach Volunteer** Atlanta, GA | October 2020 – November 2020

- Collaborated on a team of 50 volunteers to help canvas voters for Georgia's runoff election
- Separated a list of 5,000 names into various nationalities so that they could be canvassed in their native language.

## SKILLS

**Front-End Development:** React, TypeScript, JavaScript, Java, Python, C  
**Languages:** English (fluent), Spanish (fluent), Amharic (speaking: fluent)

# Korey Sam

(617)-967-0381 • korey\_sam@brown.edu • [LinkedIn](#)

## EDUCATION

### Brown University

Master of Science Candidate in Computer Science

Providence, RI

Expected May 2026

Bachelor of Science in Computer Science, GPA: 3.84

May 2025

Relevant Coursework: Intro to Software Engineering, Fundamentals of Computer Systems, Program Design with Data Structures and Algorithms, Introduction to Object-Oriented Programming, Logic for Systems

## TECHNICAL SKILLS

**Programming Languages:** Java, React/TypeScript, C, C++, C#, HTML, CSS, and Python

**Software:** Adobe Creative Suite, SQL Server, ASP.Net

**Certifications:** AWS Cloud Practitioner

## PROJECTS

### Maps | Java / React / TypeScript

- Developed a full-stack web application mapping historical redlining data from the U.S. Census using Mapbox for interactive and customizable map views.
- Implemented API endpoints to overlay redlining data with advanced filtering and search capabilities, enhancing user experience and data accessibility.
- Integrated user-specific map pins linked to Google Firestore, allowing users to securely add, store, and delete pins, persisting data across sessions.

### Eco Equity | Java / React / TypeScript

- Created a full-stack application with Java for backend API connections and React/TypeScript for a dynamic, responsive front-end user experience.
- Integrated Google's Moshi Adapter for efficient JSON parsing from API requests, converting data into usable objects.
- Enabled sustainably conscious investors to access ESG and financial data, allowing them to favorite, search, and filter companies and relevant financial articles through seamless Google Firestore and Microsoft Bing integration.

## WORK EXPERIENCE

### Accenture, Technology Architecture Analyst

Boston, MA | Summer 2024

- Supported a \$45M Higher Education Workday implementation to bridge international campus financial systems
- Led a successful integration of Grants and Customer Accounts, achieving 35% completion of client data migration
- Built a dynamic excel tracking system to optimize project management and milestone tracking for 53 deliverables
- Developed a customer confirmation deck to ensure project alignment with client needs and facilitate decision-making

### Novanta, Cambridge Technology, Software Engineering Intern

Bedford, MA | Summer 2023

- Developed a full-stack web application using C# and HTML, enhancing UI/UX and ensuring smooth data handling with an asynchronously updated SQL server
- Gained proficiency in frameworks like Avalonia UI, Microsoft Blazor, and ASP.Net for web development
- Presented web application to a group of 10 engineers and business executives to support a seamless deployment
- Ensured project alignment through bi-weekly meetings with engineering team and senior leadership

### Brown University, Community Coordinator

Providence, RI | Fall 2022 - Current

- Developed a vibrant residential community for 110 residents, assuming the role of a peer leader
- Addressed resident concerns and mediated conflicts in weekly community hours, enhancing satisfaction
- Execute 3 large-scale educational and wellbeing programs each semester, boosting community participation

## LEADERSHIP AND PROFESSIONAL DEVELOPMENT

### Beta Omega Chi, Chapter President

Providence RI | Spring 2024 – Current

- Directed a 10-member board, organizing bi-weekly meetings to ensure task completion and team collaboration
- Raised \$15,000 through fundraising initiatives, funding community outreach and chapter activities
- Implemented diversity, equity, and inclusion initiatives into recruitment, ensuring a culture of safety for all members
- Launched an alumni mentorship program, pairing members with experienced alums for professional development

### National Society of Black Engineers, Executive Board Member

Providence, RI | Fall 2022 - Current

- Boosted member engagement and attendance by 25% through sponsor-backed events and corporate raffles
- Organized community service projects with local nonprofits, increasing chapter visibility and impact
- Led career development workshops, increasing participation by 30% and securing sponsorships from top tech firms
- Led collaborations with Fortune 500 companies, creating unique recruitment opportunities for chapter members

## INTERESTS

Basketball, photography, weight training, leisure autobiographical and fictional reading, travel





## Kenny Gruchalla

Principal Scientist, Computational Science  
Computational Science Research Topic

<https://orcid.org/0000-0002-8155-1175>

Email  
[Kenny.Gruchalla@nrel.gov](mailto:Kenny.Gruchalla@nrel.gov)



2010 2025  
Research Activity per Year

Overview

Fingerprint

Network

Research Output (66)

Awards & Honors (1)

Similar Profiles (6)

## Overview

### Personal Profile

Kenny Gruchalla leads the visualization efforts in the Computational Science Center at NREL. His current research is focused on developing interactive scientific visualization techniques that provide tools for finding meaning in increasingly large and complex data. Kenny has more than 25 years of professional experience in scientific programming and scientific visualization, spanning several scientific disciplines, including aerospace, geophysics, molecular biology, and environmental engineering. Gruchalla's dissertation work focused on the development of a general framework to define multi-scale, multivariate, non-Cartesian regions of interest in large-scale turbulent flow data. His master's thesis research evaluated scientific workflows in immersive virtual environments. His current research interests include scientific visualization, immersive visualization, GPU computing, human-computer interaction, and energy-systems analysis.

### Education/Academic Qualification

Bachelor, Computer Science, New Mexico Institute of Mining and Technology

PhD, Computer Science, University of Colorado Boulder

Master, Computer Science, University of Colorado Boulder

[Follow](#)[Manage Profile](#)

## Rahul Kadavil [?](#)

Also published under: [R. G. Kadavil](#)

### Affiliation

Advanced Grid Innovation Laboratory for Energy (AGILE)  
New York Power Authority  
White Plains, NY, USA

### Publication Topics

[Power System](#), [Real-time Simulation](#), [Power Grid](#), [Transmission Line](#), [Power System Model](#), [Data Exchange](#), [Distributed Energy Resources](#), [Distribution Lines](#), [Large-scale Power Grids](#), [New York State](#), [Time Synchronization](#), [Transmission System](#)

[Show More](#)

### Biography

Rahul Kadavil (Senior Member, IEEE) received the B.E. degree in electrical engineering from Fr. C. Rodrigues Institute of Technology, University of Mumbai, Mumbai, India, in 2009, and the M.S. degree in electrical engineering from Colorado State University, Fort Collins, CO, USA, in 2017. He was engaged as a Research Engineer with the Power and Energy Real-Time Laboratory (PERL), Idaho National Laboratory for modeling and integration testing of Hybrid Energy Storage Systems (HESS) using real-time simulation tools. He is currently a Manager with the Advanced Grid Innovation Lab for Energy (AGILE), New York Power Authority, White Plains, NY, USA. He designs and tests hardware in the loop experiments to analyze the diverse smart technologies and understand how they complement each other. His research interests includes power systems digital simulation, digital twins, protection, cyber-physical testbeds, and integration of renewable energy resources. He is the second inventor of Energy Storage Ramping Optimization for aggregating ramping capabilities from multiple HESS. (Based on [document published on 10 September 2024](#)). [Show Less](#)

## References

- [1] Yutian Zhang, Liwen Xu, Shaocong Tao, Quanxue Guan, Quan Li, Haipeng Zeng, "CSLens: Towards Better Deploying Charging Stations via Visual Analytics — a Coupled Networks Perspective", *IEEE Transactions on Visualization and Computer Graphics*, vol.31, no.1, pp.251–261, 2025.
- [2] Graham Johnson, Sam Molnar, Nicholas Brunhart-Lupo, Kenny Gruchalla, "Architecture for Web-Based Visualization of Large-Scale Energy Domains", *2024 IEEE Workshop on Energy Data Visualization (EnergyVis)*, pp.46–51, 2024.
- [3] Lixiang Zhao, Tobias Isenberg, Fuqi Xie, Hai-Ning Liang, Lingyun Yu, "MeTACAST: Target- and Context-Aware Spatial Selection in VR", *IEEE Transactions on Visualization and Computer Graphics*, vol.30, no.1, pp.480–494, 2024.
- [4] Nila Bagheri, Emanuel Teixeira, Fernando J. Velez, Jon M. Peha, "Multi-Band Resonant Photonic Crystal Antenna for 5G Applications", *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*, pp.526–531, 2024.
- [5] David Moreno-Lumbreras, "Enhancing HTML Structure Comprehension: Real-Time 3D/XR Visualization of the DOM", *2024 IEEE Working Conference on Software Visualization (VISSOFT)*, pp.127–132, 2024.
- [6] Musaab Osamah Anwer, Ezzeddine Zagrouba, "The Role of Virtual Reality Technology in Knowledge Acquisition: A Systematic Review", *2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP)*, vol.1, pp.255–260, 2024.
- [7] G. Bruder, F. Steinicke, and K. H. Hinrichs, "Arch-explore: A natural user interface for immersive architectural walkthroughs", *IEEE Symposium on 3D User Interfaces*, pp.75–82, 2009.
- [8] P. Moyer-Packenham, "How design features in digital math games support learning and mathematics connections", *Computers in Human Behavior*, vol.91, pp.316–332, 2019.

**Reviewer akanna15:**

**Visualization:6**

**The visualization component of the project could be explained in more detail. It wasn't clear from the proposal alone**

Response: many reviewers have addressed this issue. The core of this project is not to develop a grand new technological approach for LiDAR datasets, but offer a more accurate algorithm for element classification within the datasets and semantically label each classified cluster that allow resolving the problem through training or further postprocessing.

**Novel:4**

**how this algorithm surpasses existing ones could be explained in detail, to highlight the novelty.**

Reviewer's comment on this regard is highly appreciated. This part has been reinforced in the editing of second version. Generally speaking, the novelty of this project does not lie in visualization techniques, but rather the approach to process the LiDAR datasets by labeling clusters with reasonable semantic labels. It is quite difficult to evaluate the algorithm's general performance at this level, but I believe introducing semantic labeling contributes to the novelty of the project and will allow more efficient postprocessing.

**Likelihood of Success:7**

**Since visualization components aren't defined, it is difficult to say**

I believe previous response covered this concern.

**Reviewer alabella:**

**Overall: 4. Goals are well stated, and the proposed methods seem sensible given the problem. What remains unclear is the second aim, specifically how deep learning can help and what "semantic labeling" means.**

This has been further clarified in the corresponding section. Current application of deep learning in shape/scene segmentation works mostly concentrate in applying a classification algorithm in the feature space of a dataset. While we aim to explore if this technique can be developed or improved, we will also seek for other viable solutions to improve the accuracy of classification. "Semantic labeling" means to give each classified cluster a descriptive label that can precisely describe the content of the cluster (e.g. vegetation, trunk, ground, ceiling, etc.)

**Interdisciplinary: 4. Only one collaborator is listed. I think having another collaborator would bolster this proposal.**

Another collaborator is now added into the proposal.

**Scientific: 5. The author lists what they intend to do but how they will empirically validate their findings is a bit vague.**

This is further addressed in the revised version. I expect our clustering algorithms to produce both a more accurate classification outcome and a semantic label that can precisely describe the content of each cluster. This should contribute to both the evaluation of classification accuracy and more efficient postprocessing of classified datasets.

**Novel: 4. 3D point clouds have been around for a while, so they themselves are not inherently novel. What is novel is how the authors plan on visualizing these clouds.**

This will be further acknowledged and reemphasized in the revised proposal.

**Reviewer ccasti18:**

**Overall: 5 (Good)**

**This proposal addresses an important challenge in archaeological LiDAR classification and has potential for meaningful contributions. However, weaknesses in evaluation methods and project structure should be addressed to strengthen feasibility and impact.**

The proposal is now revised to better explain how semantic labeling will also contribute to the evaluation process. In short, we will mobilize semantic labels to identify whether the label provides a description of the cluster's content. Manual inspection will also take place to ensure the semantic labels are sufficiently accurate. Project structure wise, the order of deep learning method exploration and semantic labeling to ensure the labeling process benefits maximally from the deep learning yields.

#### **Visualization: 6 (Satisfactory)**

**The visualization component, using semantic labels and possibly mesh reconstructions, is interesting and could be valuable for archaeologists. At the same time, the plan feels underdeveloped and leaves several important questions unanswered, such as how archaeologists will interact with the outputs. A helpful next step may be to include evaluation with domain experts in addition to novice users. Adding a relevant image or a mock-up of the intended visualization could also make the goals clearer and strengthen this section.**

The primary purpose of the project is not to develop a new visualization method. Rather, we intend to better facilitate the current visualization techniques through providing better classified datasets. Therefore, I do not anticipate that visualization will be a primary target for the project. The reviewer also mentioned about archaeologists' interaction formats are not defined clearly – this is further clarified in the revision, indicating that archaeologists will be able to inspect the clusters more straightforwardly with the dataset; original datasets contain only the point cloud but and clusters without descriptive information. I wish that effective semantic labeling would enable certain parts of the dataset to be removed temporarily to avoid distraction, allowing the archaeologists to interact with the site structure with less noisy information. I also appreciate reviewer's comments on how inviting domain experts participating in the research can contribute to the evaluation process – qualitative feedback from experts will be collected in addition to novices' evaluation. The feedback will primarily be inquired in regard to the interaction method – the experts will offer evaluation on whether semantic labeling based selective visualization contributed to their work.

#### **Goals clearly stated: 4 (Very Good)**

**The goals are clearly stated and organized into phases that follow a logical sequence. The timeline is slightly vague though, with four weeks described in detail while the rest relies on loosely defined extra time.**

The reviewer pointed out a significant point. However, our current methodology involves large uncertainty – semantic labeling, at the moment, is still at large a big challenge even for the frontier researchers. We are rather suspicious that whether the time will be sufficient for us to explore the topic in-depth, so we make up tentative plans to allow flexibility in case of encountering difficulties.

# An Exploration of Semantic Labeling and Classification in Ancient Sites

Yuqiao Guan

PI

yuqiao\_guan@brown.edu

Yiyang Zhang

Co-PI

yiyang\_zhang3@brown.edu

Parker VanValkenburgh

Collaborator

parker\_vanvalkenburgh@brown.edu

October 14, 2025

## Abstract

We propose to explore a more effective classification algorithm for ancient site with standing structures - compared to mesh, point clouds contains less structural information and is difficult to reason structures. This research presents an improved algorithm that focus on identifying structures and provide a corresponding semantic label to facilitate structural understanding.

# 1 Aims

There exists two primary aims for this project. The first one is improving existing classification algorithm. Current algorithms have difficulty identifying standing structural components of a human constructed site and tree trunks - we anticipate that this error can be addressed by mobilizing deep learning networks instead of using purely traditional clustering algorithms like DBScan or K-Means.

The other aim is to explore how semantic labeling of the sampled dataset can contribute to a better deep learning-based classification algorithm. With the advancement in deep learning and VLM, it is worth exploring that whether current visual understanding can effectively tackle the task of generating a semantic label that provide accurate description about the content of the classified clusters.

# 2 Significance

This project fulfill the demand of a more effective classification algorithm for drone-based LiDAR data collected from heavily vegetated sites. Given that existing classification algorithms struggle to identify standing structures like mound walls and tree trunks, our most prominent significance is to push beyond the current reliance on mound-centric algorithms and advance approaches that are able to distinguish structural components with relatively more patterns than vegetation. Such breakthrough is, presumably, particularly helpful for archaeological excavations in tropical areas like South America and Southeast Asia where heavy vegetation is common.

Theory-wise, we expect our research findings contribute to the production of high-resolution elevation models and hill-shades by modeling more reliable ground and wall information. This will turns out to be highly contributive in the second phase of the research project which aims at understanding the spatial distribution patterns of a highly Balkanized site clusters through intervisibility. A more concrete modeling of the internal site structures offers a solid ground to evaluate the extent of visibility other sites are.

In terms of visualization, we expect to implement a selective visualization method. The classified point cloud, along with the semantic labels, will allow us to selectively visualize points with certain properties. For instance, users can choose to visualize only the human-constructed sites to see how they fit the terrain, or visualize only the vegetation to see how they embed into the site structure. We also intend to give different colors for the points representing different elements - this will facilitate archaeologists' immediate grasp on the structure of the point cloud dataset.

Moving beyond site-understanding field, we expect this classification algorithm to also contribute in differentiating human constructions over vegetation when broadly searching for sites rather than analyzing a specific one. It will also contribute to modeling techniques of ancient sites for preservation and reconstruction purposes. Its semantic labeling technique may also enable procedural reconstruction and generative restoration of damaged sites. Our ultimate goal is to explore and explain the rationale of such technical approach in archaeology.

# 3 Related Work

There are numerous existing literature that seek to resolve the problem from different approaches. Early works by Brodu and Lague introduced the concept of multi-scale dimensionality analysis, quantifying local point cloud geometry (e.g. linear, planar and volumetric) to distinguish vegetation, ground and terrain objects [1]. This "dimensionality" is an emphasis of how structural clues across all scales can distinguish objects that traditional algorithms may compromise. More recent research focus on mobilizing machine

learning and neural structures - Ciou et al (2024) propose a learning-based technological framework that combines multiple classifiers to improve robustness in heterogeneous landscapes [3]. Similarly, Liu et al (2024) described a strategy to apply transformer-based models for automated classification of national LiDAR dataset during a US Geographical Survey, demonstrating its flexibility and reliability in such task [4]. Zhang et al. (2023), though earlier in time, demonstrated a systematic survey of deep learning-based 3D point cloud classification, comparing voxel, point and graph-based architectures and identified ongoing challenges in each of these methods [7]. These papers effectively broadened the interdisciplinary horizon and exposed deep learning’s tremendous potential in archaeology, and now it is our turn to bridge the gap between the deep learning-based technological framework and specific tasks that not only requires classification but also effective illustration of structures.

Several studies have also examined how LiDAR classification can be applied on archaeology. Štular et al. (2021) present a pipeline and QGIS toolbox for archaeology, providing an easy-to-use and deployable workflow to convert unclassified point clouds to visualization products, including confidence maps and hybrid interpolation [8]. Wang et al. (2024) introduced CNN-based models into the technological framework and successfully reveal trivial cultural heritage features covered by dense vegetation [6]. Poux et al. (2017) made a bolder attempt, integrating point cloud datasets into archaeological knowledge base systems, proposing semantic enrichment and ontology-based classification of quasi-planar objects. Their research collectively shown that vegetation density could be the primary issue in visualizing the classification results, and interpretability of a LiDAR dataset is extremely important - these two features seem irrelevant, but inspired by the inspiration of successes achieved in computer vision fields, we realize that there exists a knowledge gap: connecting automatic removal of noisy information and semantic-based data understanding will tremendously contribute to the researchers.

Classification tasks, in many scenarios, are intended for a more effective visualization and scene understanding; numerous research also highlighted this and mobilize these visionary observations to explore a spatial setting. Risetto (2017) combines 3D models with GIS viewshed analysis to examine how visibility shaped ancient Maya social and political institution [5], and Cassidy (2019) introduced VR platform that allow archaeologists to immersively explore LiDAR data, further facilitating feature exploration [2]. These studies have collected highlighted a way of how a well-classified dataset can contribute to a more in-depth exploration of site features, and it further provoked our thinking about how our method can be maximally exploited for visualization and pattern examination.

## 4 Research Plan

The first phase of this research will focus on reading the point cloud and identify the problem. We will observe the dataset to understand the difficulty of classification task: to what extent can algorithms be confused with standing structures and trunks, and how difficult it is for humans to identify such structures? This step will enable us, as researchers, to understand the nature of the problem that we intend to solve and provide a more accurate intuition on the proper strategies that should be employed in the process. It is expected that we should attempt the existing classification algorithms like K-Means and DBScan and how they fail in meeting the expectations. I anticipate this to require approximately 7 to 10 days to complete.

The second phase will aim at granting each segmented cluster an understandable semantic label. The purpose of this step is to further illustrate the structural information of the collected point cloud data. Point clouds, by nature, is not very effective in reflecting structures, so assigning semantic labels helps audiences to better evaluate the accuracy and effectiveness of classification. It will also provide more data-wise guidance in training a model on classifying elements within similar settings. We also consider reconstructing a



mesh from the point cloud to better highlight the structural information. Estimated time of completion will be approximately 1 week.

The third phase marks the start of exploration of Deep Learning-based algorithm attempts. Problem needs to be addressed in this research is highly specialized: to distinguish standing structures and vegetation. This specialized task might benefit enormously from training a model, though it involves intensive data collection and training process. We estimate completion of this task to take 2 weeks, adding time required for dataset structuring and training.

The last phase will evaluate the effectiveness of our classification model by evaluating the quality of clusters. Such evaluation will be first conducted by experts with profound understanding in the field - they will evaluate whether our new visualization approach facilitates dataset exploration. The second stage will include novices who have no strong background in archaeology and could only visually reason the reliability of a clustering strategy applied on a given point cloud dataset. It is expected that their intuition in visual reasoning will contribute most effectively.

## References

- [1] N. Brodu and D. Lague. 3d terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology, 2012.
- [2] B. Cassidy, G. Sim, D. W. Robinson, and D. Gandy. A virtual reality platform for analyzing remote archaeological sites. *Interacting with Computers*, 31(2):167–176, 2019.
- [3] T.-S. Ciou, C.-H. Lin, and C.-K. Wang. Airborne lidar point cloud classification using ensemble learning for dem generation. *Sensors*, 24(21):6858, 2024.
- [4] J. K. Liu, R. Qin, and S. Song. Automated deep learning-based point cloud classification on usgs 3dep lidar data using a transformer. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 8518–8521, 2024.
- [5] H. Richards-Rissetto. An iterative 3d gis analysis of the role of visibility in ancient maya landscapes: A case study from copan, honduras. *Digital Scholarship in the Humanities*, 32(suppl<sub>2</sub>) : ii195 – –ii212, 032017.
- [6] S. Wang, Q. Hu, S. Wang, M. Ai, and P. Zhao. Archaeological site segmentation of ancient city walls based on deep learning and lidar remote sensing. *Journal of Cultural Heritage*, 66:117–131, 2024.
- [7] H. Zhang, C. Wang, S. Tian, B. Lu, L. Zhang, X. Ning, and X. Bai. Deep learning-based 3d point cloud classification: A systematic survey and outlook. *Displays*, 79:102456, 2023.
- [8] B. Štular, S. Eichert, and E. Lozić. Airborne lidar point cloud processing for archaeology. pipeline and qgis toolbox. *Remote Sensing*, 13(16):3225, 2021.

Link to the manuscript pdf:

<https://drive.google.com/file/d/1mwBwQnnDoYqZuBtT2U08eFx1AbdfkPh0/view?usp=sharing>

## Response Letter

Dear Editor and Reviewers,

We appreciate the constructive feedback on the manuscript, and we have prepared a revision accordingly. Please see below for an itemized list of your feedback and a brief description of how each is addressed in the revised manuscript.

R1: [https://drive.google.com/file/d/1-KT2BEWcOI-m\\_0S\\_Le3XqASsk-8lj1Ka/view?usp=drive\\_link](https://drive.google.com/file/d/1-KT2BEWcOI-m_0S_Le3XqASsk-8lj1Ka/view?usp=drive_link)

[R1.6] Project is specialized towards Chachapoya culture

**Response:** For novelty, although the project is centered on Chachapoya sites, its methods are broadly applicable to other forested archaeological regions. The Chachapoya landscape simply provides an ideal test case for developing a generalizable LiDAR classification and visibility workflow useful worldwide.

[R1.8] Aims of the project are a little bit difficult to conceptualize, for those not familiar with immersive VR.

**Response:** While the immersive overlaying visualization component may seem abstract to those unfamiliar with virtual reality, its purpose is straightforward: to provide a clear, experiential way to understand spatial relationships and visibility that traditional maps cannot show. The VR element is a tool for allowing both experts and non-specialists to intuitively explore how Chachapoya sites interconnect across the landscape.

R2: <https://drive.google.com/file/d/1DqMimmpNEjamL2vWb8CoZwjS7syi1Wf5/view?usp=sharing>

[R2.2] Only one collaborator is listed and they are from Brown. Would it help to add collaborators that are from another institution?

**Response:** The project already integrates computer science, archaeology, and geospatial visualization—three distinct disciplines. If needed we could explore informal collaboration with external archaeologists and remote-sensing specialists to broaden institutional scope, but the interdisciplinarity lies mainly in the fusion of technical and humanistic approaches, not institutional affiliation.

[R2.3] Methods for classification are listed, but no real means for evaluation are given. What is the ground truth? What metrics will you use to benchmark your results?

**Response:** We will evaluate classification accuracy using manually labeled ground-truth data from selected training patches and expert-verified wall segments. Model performance will be evaluated with IoU-score, and overall accuracy for each class (architecture, vegetation, ground). We will also include visual validation to ensure the reconstructed surfaces preserve architectural integrity.

[R2.6] 3D point clouds have been around for a while, so they themselves are not inherently novel. What is novel is how the authors plan on visualizing these clouds.

**Response:** We agree that 3D point clouds themselves are established, but our novelty lies in the pipeline that combines semantic labeling, Digital Feature Model generation, and immersive intervisibility visualization.

R3: <https://drive.google.com/file/d/1vJdb7W11BOjnpimeDIRjJfljqlr9D4kK/view?usp=sharing>

[R3.6] The gaps filled in by this proposal in existing works can be explained in detail, to highlight novelty.

**Response:** See [R2.6]

R4: [https://drive.google.com/file/d/1O7q0KMjMD40rzi1AVeaOT44lv\\_fpYTfj/view?usp=sharing](https://drive.google.com/file/d/1O7q0KMjMD40rzi1AVeaOT44lv_fpYTfj/view?usp=sharing)

[R4.8] The methods are clearly laid out and mostly feasible. The VR portion is more ambitious but likely achievable in prototype form. The biggest worry is the plan to freeze model design by Week 3, which may not allow enough iteration if initial performance is weak. It may be better to state this as a target goal rather than a fixed requirement, or to include a fallback plan if time becomes constrained.

**Response:** We do agree that flexibility is important. The Week 3 model freeze is meant as a target milestone, not a rigid cutoff. If initial classification results are weaker than expected, we will extend tuning in parallel with downstream visualization work, using a simplified or interim model to keep the project on schedule. The improvements to the classifier can be integrated later without delaying the visibility or visualization stages.

# LiDAR Classification and Intervisibility Visualization for Chachapoya Archaeology

Yiyang Zhang  
PI

yyiyang-zhang3@brown.edu

Yuqiao Guan  
Co-PI

yuqiao-guan@brown.edu

Parker VanValkenburgh  
Collaborator

parker.vanvalkenburgh@brown.edu

October 14, 2025

## Abstract

We propose a novel pipeline to improve the analysis of densely vegetated archaeological sites by combining advanced LiDAR point-cloud classification with immersive visibility network visualization. Our method addresses the limitation of standard filtering techniques by introducing a machine learning classifier tailored to LiDAR data of standing architecture. Classified 'clean' point clouds are used to generate high-resolution elevation models for multiple settlements on the Chachapoya hilltops. We then construct intervisibility networks (graphs of which sites can see which others) and explore them through both quantitative network analysis and an interactive VR system. This approach overcomes obstacles in current techniques for modeling sightlines in complex terrain. We anticipate that the refined site models and network visualizations will yield new insights into the spatial organization, communication, and social dynamics of the Chachapoya landscape.

# 1 Aims

For enhancing LiDAR point-cloud classification, we will develop an improved method for classifying LiDAR point clouds of archaeological sites with dense forest cover, specifically to separate and identify standing architectural structures like stone walls, buildings, terraces versus vegetation.

Unlike standard ground filtering algorithms [5], that adopts simulation filter which drapes a virtual cloth over an inverted landscape but cannot fully distinguish trees from man-made walls, our approach will leverage machine learning and multi-scale geometric analysis. We plan to train a classifier, potentially a convolutional neural network or a multi-scale shape descriptor method, on labeled point cloud data where architectural points and vegetation points are annotated. The classifier will exploit features such as local geometric texture (planarity vs. linearity) and radiometric intensity differences between stone surfaces and organic matter. By building on methods like the CANUPO [1] multi-scale dimensionality algorithm, we aim to reliably filter out canopy and tree trunks while retaining fine details of ruins. The result will be clean, classified point clouds and derived Digital Terrain Models where archaeological structures are clearly delineated from natural features.

Using the enhanced site models from Aim 1, we will analyze and visualize the intervisibility between multiple Chachapoya settlements. First, we will compute viewsheds (the areas visible from a given viewpoint) and line-of-sight connections among sites using the classified elevation models. This will produce a network graph of sites as nodes and intervisible links as edges. We aim to evaluate whether these hilltop communities form clustered communication networks—as hypothesized, given historical suggestions of fire or smoke signaling in the Andes [3]—or if their arrangement is random. To do so, we will apply network science metrics (connectivity, clustering coefficients, network density) as advocated by Brughmans and Brandes [frontiersin.org](http://frontiersin.org) [2] to quantify the patterns of visual connectivity.

Secondly, we will create an immersive 3D visualization with VR support to let researchers virtually stand at one site and see which other sites would have been visible in the ancient landscape. This goes beyond traditional 2D GIS maps by providing a first-person perspective and real-time exploration of sightlines in a simulated environment. The VR component will serve both as an analytical tool and as a means to communicate findings to a broader audience. By the end of the project, we plan to demonstrate a prototype where users can toggle the presence/absence of modern vegetation and experience the network of hilltop settlements through interactive visualization.

## **2 Significance**

This project is significant across multiple dimensions, including its archaeological impact, methodological contributions, and implications for heritage management and public engagement. Together, these outcomes demonstrate the interdisciplinary value of combining archaeology, computer vision, and virtual reality in a single framework.

### **2.1 Archaeological and Anthropological Impact**

The project investigates the Chachapoya, an Indigenous Andean culture known for dispersed hilltop settlements traditionally interpreted as isolated due to rugged terrain and political fragmentation. We propose instead that many sites were strategically positioned within line-of-sight to enable communication and integration across the landscape. Demonstrating such intervisibility networks would challenge prevailing narratives and suggest a higher degree of regional organization. Because colonial accounts are limited and disrupted Indigenous systems, spatial analysis offers a novel way to reconstruct pre-colonial connectivity beyond the historical record.

### **2.2 Technological and Methodological Contributions**

Our LiDAR classification approach addresses a major challenge in archaeological remote sensing: distinguishing standing architecture from vegetation under dense forest canopy. Unlike studies that focus on large earthworks or terrain features, we target the task of separating vertical stone walls from tree trunks in point cloud data. By combining multi-scale geometric and radiometric features with machine learning [4], we aim to deliver a broadly applicable method for detecting overgrown ruins worldwide. In parallel, we will create one of the first VR-based visibility analysis tools in archaeology, allowing researchers to experience spatial relationships in 3D rather than through abstract maps. This integration of machine learning and VR represents a significant methodological innovation for heritage science.

### **2.3 Heritage Management and Public Engagement**

Chachapoya sites such as Kuélap are gaining importance as heritage tourism destinations and alternatives to the crowded Machu Picchu circuit. Our vegetation-free models and VR experiences will support virtual tours and educational outreach, enhancing public appreciation of Chachapoya culture. The project also sheds light on long-term environmental change by comparing modern forest cover with ancient, likely cleared landscapes. These insights have direct value for both ecological research and the management of fragile cloud-forest regions.

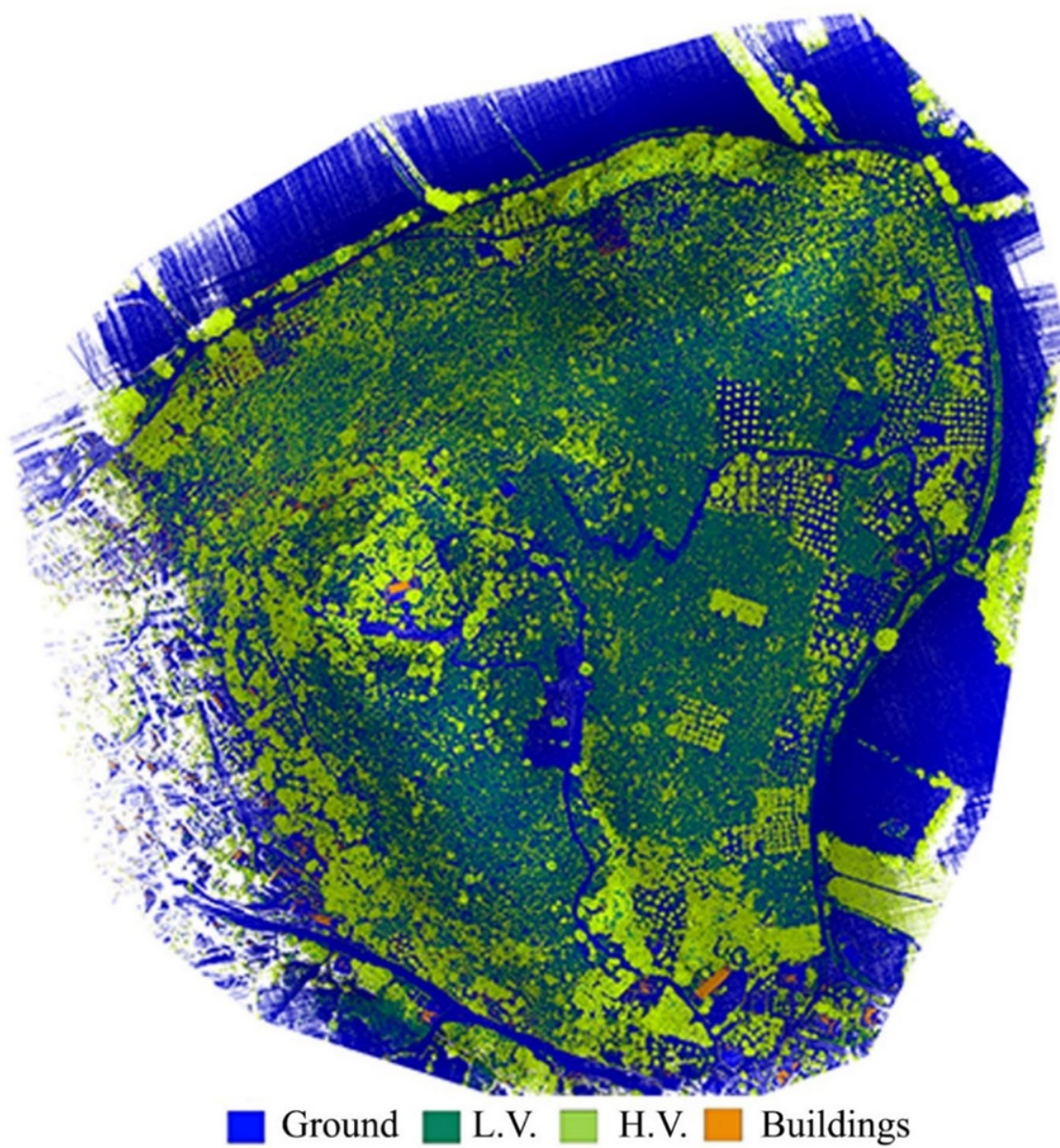
### 3 Related Work

**LiDAR Classification in Archaeology** Early archaeological applications of LiDAR focused on ground and vegetation filtering with manual feature recognition on Digital Terrain. While methods such as the Cloth Simulation Filter (CSF) [7] can generate bare-earth models, they often misclassify ruins as ground or retain tree trunks, limiting accuracy in dense forests. Multi-scale algorithms like CANUPO [1] improved separation of vegetation and terrain by computing geometric features across scales. Building on this, our approach specifically tailors classification to distinguish man-made linear or planar structures from organic forms in complex forested contexts.

**Machine Learning for Feature Detection** Recent studies have applied machine and deep learning to LiDAR for detecting features such as burial mounds, terraces, and building remains [6]. CNN-based methods have successfully classified Maya structures and even historic dry-stone walls, though many operate on rasterized elevation models rather than 3D point clouds. Our project will combine both approaches: leveraging detailed 3D point-cloud geometry while using CNN techniques on derived models for pattern recognition. This hybrid strategy is novel in focusing on point-level architectural versus vegetation classification, informed by domain knowledge of Chachapoya architecture.

**Viewshed Analysis and Intervisibility** Viewshed analysis has long been used in archaeology to study site visibility, defensive placement, and communication networks [2]. Network-based approaches treat intervisibility as a graph, allowing analysis of hubs, clusters, and connectivity against null models. More recently, immersive 3D and VR environments have allowed archaeologists to directly simulate and perceive visibility from ancient sites. Our project extends this trend by combining VR with machine-learning-enhanced reconstructions, creating a uniquely accurate and interactive framework for evaluating Chachapoya intervisibility networks.





**Fig. 7** 3DMASC classification results in CloudCompare (top view)

Figure 1: Example of the result visualization of semantically labeled point clouds

## 4 Research Plan

- **Week 1–2:** Prepare LiDAR data and orthophotos; provide LiDAR data + orthos, run a quick baseline model with standard ground filters, and test baseline tools (e.g., CSF, CANUPO) to establish a benchmark classification. Decide on model architecture and set up the training environment.
- **Week 2–3:** Train a supervised point-level classifier (with CNN masking for object-level prior) using labeled samples; fuse outputs to generate semantic labels. Experiment with features (geometry, intensity, returns) and evaluate accuracy against the benchmark. Freeze model design by end of Week 3.
- **Week 4:** Batch process all site point clouds with the trained classifier; generate de-vegetated surfaces (architecture-only and Digital Feature Model meshes per patch/site), DTMs, canopy height models, and visualizations. Validate outputs with archaeologists and refine if needed.
- **Week 5:** Conduct large-scale viewshed and visibility network analysis using cleaned DTMs; overlay per-point classification confidence on point clouds with hillshade/orthophoto for walls versus vegetation. Compute intervisibility, compare scenarios (modern vs. ancient vegetation), and derive network metrics.
- **Week 5–6:** Zoom in to reconstruct de-vegetated, architecture-preserving meshes and verify that low-confidence zones coincide with vegetation-covered areas. Develop a VR/interactive 3D environment for site visualization; enable exploration, visibility toggles, and communication simulations. Prepare a working demo by end of Week 6.
- **Week 6:** Synthesize results into the final report and presentation; include classifier performance, confidence overlays, visibility network visualizations, VR demo feedback, and interdisciplinary conclusions.

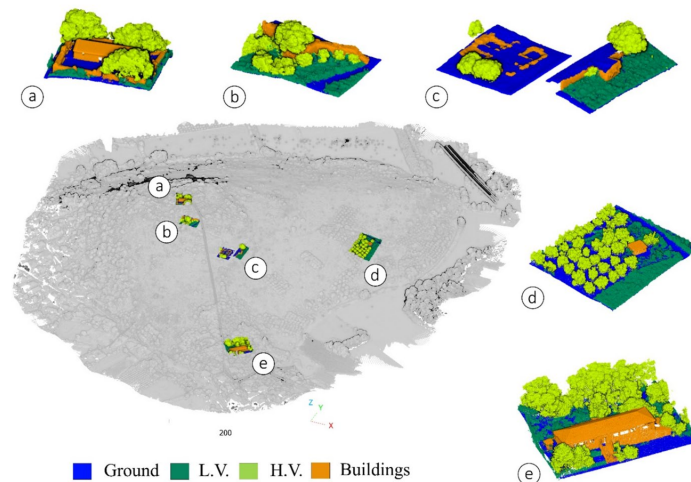


Fig. 4 Sample of areas used as training data with classes: ground, low vegetation, high vegetation and buildings

Figure 2: Example of result rasterization of mesh-reconstructed structures

## References

- [1] N. Brodu and D. Lague. 3d terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS journal of photogrammetry and remote sensing*, 68:121–134, 2012.
- [2] T. Brughmans and U. Brandes. Visibility network patterns and methods for studying visual relational phenomena in archeology. *Frontiers in Digital Humanities*, 4:17, 2017.
- [3] T. Earley-Spadoni. Landscapes of warfare: Intervisibility analysis of early iron and urartian fire beacon stations (armenia). *Journal of Archaeological Science: Reports*, 3:22–30, 2015.
- [4] T. Hackel, J. D. Wegner, and K. Schindler. Fast semantic segmentation of 3d point clouds with strongly varying density. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, 3:177–184, 2016.
- [5] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105:286–304, 2015.
- [6] H. Yang. Segmenting ancient cemeteries under forests using synthesized lidar-derived data and deep convolutional neural network. *npj Heritage Science*, 13(1):232, 2025.
- [7] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan. An easy-to-use airborne lidar data filtering method based on cloth simulation. *Remote sensing*, 8(6):501, 2016.