

CS1951A - Data Science

Course Syllabus

Spring 2017

Introduction

CS1951A, Introduction to Data Science, is a interdisciplinary survey course that covers the fountains of data science. It takes a holistic look at the foundations of what people today use to store, analyze, and summarize massive amounts of data.

Lectures are held Tuesday and Thursday 9:00-10:30 in 130, 85 Waterman.

Professors

CS1951A is co-taught by

- Dan Potter (dan@cs.brown.edu)
- Carsten Binnig (carsten@cs.brown.edu)
- Eli Upfal (eli@cs.brown.edu)

Head TAs

CS1951A's head TAs are

- Alex Bertsch (asbertsc)
- Karthik Harihar Reddy Battula (kbattula)
- Tiffany Citra (tcitra)

UTAs

CS1951A's undergraduate TAs are

- Andreas Karagounis (akaragou)
- Colby Tresness (ctresnes)
- Eden Weizman (eweizman)
- Jack Kelly (jckelly)

- Jared Cohen ([jtc2](#))
- Joey Tran ([htran6](#))
- Michael Xu ([msxu](#))
- Neel Viridy ([nvirdy](#))
- Vann Ky ([vky](#))

Who should take this class?

Anyone who is interested in data science and its many components, including databases, visualization, and machine learning, should take this engaging survey course to explore an ever growing sub-field of computer science. Data science is becoming integral to all domains from materials science to business analytics.

The formal prerequisites to this course are CSCI 0160, 0180, or 0190. Additional experience in software engineering is recommended, including CSCI 0320 or 1320. This course is taught in Python 3.5, but no prior experience is necessary. We will provide several resources to get students started with Python at the beginning of the course. It is suggested that students also have experience in statistics (APMA 1650 or CSCI 1450) and linear algebra (MATH 0520, MATH 0540, or CSCI 0530) for the statistics and machine learning portion of this course.

Collaboration

Collaboration in this course is governed by a liberal collaboration policy, found on the resources section of the class website. You **must** sign the collaboration policy before any of your work can be graded. Failure to meet the standards set by the collaboration policy can result in failure and possible disciplinary action.

Contacting the Course Staff

If you have a question regarding something personal, e.g. any accommodations that we may provide or circumstances of which we should be made aware, please contact the HTA list or Dan Potter directly. If you have a question regarding logistics, please contact the TA list. When replying to emails sent to a TA list, please always reply all and do not reply directly to a single TA unless requested to do so. In general, do not email a specific TA except for grading complaints.

For all other questions, including those related to course material, please post your question on Piazza. If your question contains any code or pseudocode, we ask that you make your question private to the class. Otherwise, we would prefer if questions are made as public as possible as to be a resource to the class at large. You are free to anonymize your question as to not reveal your identify to either just the class or both the class and the course staff.

Grading

Grading is tentative. The following is our most current estimate:

| Section | Scoring |
|-------------|---------|
| Attendance | 5% |
| Labs | 5% |
| Assignments | 35% |
| Midterms | 15% |
| Project | 40% |

Attendance

Attendance is tracked via iClicker participation with answers being asked in class. Students must attend two-thirds of all classes to receive full attendance credit. Questions in class *do not* have to be answered correctly to get credit (but please don't just mash B). You must answer all questions during a certain class to receive credit for attending that day. iClickers may be set up on the course Canvas page.

Labs

Throughout the course we will have four labs, each worth 1.25% of your overall grade. The labs will cover the practical aspects of various concepts introduced in class, from implementing D3 to using blackbox ML algorithms. These labs will not have associated hours, but are named as such to differentiate themselves from the lengthier assignments. As with the assignments, feel free to visit TA hours or post questions on Piazza.

Assignments

This course will have nine graded assignments, each worth equal points towards your final grade. These assignments will be submitted via the usual course handin script. More specific instructions will be provided on each assignment.

Midterms

There will be two midterms held during our regular meeting period. They will be on the material presented in lectures and will be a comprehensive review of the course so far.

Project

Throughout the course, you will form a group of four students and work on a final project that will utilize all of the skills you have learned during the course. You will submit a pre-proposal early in the course and work closely with a mentor TA who will guide your project.

Project as a Capstone

If you choose to use this course as a capstone, you will extend your project to have a full-fledge web application with an interactive data analysis pipeline. Previous capstones have included web UIs for plotting roadtrips across the United States and restaurant recommendation apps.

Resources

This class does not have a required textbook and there will be little to no required reading outside of lecture. Various resources, including an introduction to Python, may be found on the course website.

Late day and extensions policy

Students are given three late days to use throughout the semester. These may be used in any way possible, e.g. all three on a single assignment or one on three assignments. After a student runs out of late days, late work will not be accepted for a grade. If an extension is needed, please contact the HTAs. For medical extensions, please directly contact Dan Potter.

Acceptable excuses include, but are not limited to:

1. Illness (with a doctor's note)
2. Family emergencies
3. Religious holidays
4. Delayed international or domestic travel

Regardless of circumstance, please contact the HTAs at least 48 hours in advance of the due date to arrange for an extension. HTAs reserve the right to **reject** extensions introduced with less than 48 hours before the due date. Extenuating circumstances will be evaluated on a case-by-case basis.