Final Project!

April 16, 2015
Final Project

- Define Problem
- Find Data
- Write a set of instructions
- Solution
Final Project

1. Define Problem
2. Write a set of instructions
3. Find Data
   - ANYTHING
     - Numerical Data
     - Textual Data
     - Twitter Data
     - Geographical Data
     - ...
4. Solution
Final Project

Define Problem

ANYTHING COMPUTATIONAL
Find Trends/Pat terns
Predict Outliers/Groups
Describe Anomalies
Assess Someone’s Claims...

Write a set of instructions

Find Data

ANYTHING
Numerical Data
Textual Data
Twitter Data
Geographical Data...

Solution
Define Problem

**ANYTHING COMPUTATIONAL**
- Find Trends/Patterns
- Predict Outliers/Groups
- Describe Anomalies
- Assess Someone’s Claims
...

Find Data

**ANYTHING**
- Numerical Data
- Textual Data
- Twitter Data
- Geographical Data
...

Write a set of instructions

**ANY TOOL**
- Excel, Python,
  (Probably Both)

Solution
Final Project

Define Problem

ANYTHING COMPUTATIONAL
Find Trends/Patters
Predict Outliers/Groups
Describe Anomalies
Assess Someone’s Claims
...

Write a set of instructions

Find Data

ANYTHING
Numerical Data
Textual Data
Twitter Data
Geographical Data
...

ANY TOOL
Excel, Python,
(Probably Both)

Solution

ANY ANALYSIS
Plots, Tables, KML Files, ...

CSCI 0931 - Intro. to Comp. for the Humanities and Social Sciences
Final Project

*Only Constant:* you will use computers to **process & visualize** data automatically

- **Define Problem**
  - ANYTHING COMPUTATIONAL
    - Find Trends/Patterns
    - Predict Outliers/Groups
    - Describe Anomalies
    - Assess Someone’s Claims
    - ...

- **Find Data**
  - ANYTHING Numerical Data
  - Textual Data
  - Twitter Data
  - Geographical Data
  - ...

- **Write a set of instructions**
  - ANY TOOL
    - Excel, Python, (Probably Both)

- **Solution**
  - ANY ANALYSIS
    - Plots, Tables, KML Files, ...

CSCI 0931 - Intro. to Comp. for the Humanities and Social Sciences
# Grading Breakdown

<table>
<thead>
<tr>
<th>Component</th>
<th>% of Overall Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Assignments</td>
<td>50</td>
</tr>
<tr>
<td>Unit Project #1</td>
<td>10</td>
</tr>
<tr>
<td>Unit Project #2</td>
<td>15</td>
</tr>
<tr>
<td>Final Project</td>
<td>25</td>
</tr>
</tbody>
</table>
Grading Breakdown

<table>
<thead>
<tr>
<th>Component</th>
<th>% of Overall Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Assignments</td>
<td>50</td>
</tr>
<tr>
<td>Unit Project #1</td>
<td>10</td>
</tr>
<tr>
<td>Unit Project #2</td>
<td>15</td>
</tr>
<tr>
<td>Final Project</td>
<td>25</td>
</tr>
</tbody>
</table>

This is important!
Don’t drop the ball!
Today

• Example Final Projects
• General Project Description
  – Scope
  – Timeline
Example 1: Where should I Live?

I want to buy a house somewhere near my hometown. Where should I buy?

Kasey Hass's Project
https://sites.google.com/a/brown.edu/house-hunting-mecklenburg-county/home

In Each Zip Code:
Rent price: Red dollar sign
Vacancies: For sale sign
Income: Green dollar sign
Marriage rate: Wedding rings
Example 1: Where should I Live?

Kasey Hass’s Project
https://sites.google.com/a/brown.edu/house-hunting-mecklenburg-county/home

In Each Zip Code:
Rent price: Red dollar sign
Vacancies: For sale sign
Income: Green dollar sign
Marriage rate: Wedding rings
Example 1: Where should I Live?

Kasey Hass’s Project
https://sites.google.com/a/brown.edu/house-hunting-mecklenburg-county/home

In Each Zip Code:
Rent price: Red dollar sign
Vacancies: For sale sign
Income: Green dollar sign
Marriage rate: Wedding rings
Example 2: Athletes & Geography

Malcolm Gladwell’s *Outliers*

- Professional Canadian hockey players
Example 2: Athletes & Geography

Malcolm Gladwell’s *Outliers*

- Professional Canadian hockey players
  - Mostly born in January.
  - Few are born in December.
  - Junior league’s cutoff is Dec. 31st
Example 2: Athletes & Geography

Malcolm Gladwell’s *Outliers*

- Professional Canadian hockey players
  - Mostly born in January.
  - Few are born in December.
  - Junior league’s cutoff is Dec. 31st

- Any association between professional players and *where* they were born?
Example 2: Athletes & Geography

Daniel Newmark’s Project
https://sites.google.com/a/brown.edu/daniel-s-data/project-4--distribution-of-athletes

Football (NFL)  
Hockey (NHL)
Example 3: Growth of Top Digital Media Companies

How does the Growth Rate Compare in the Past 10 Years?
Example 3: Growth of Top Digital Media Companies

Nicholas Talbott’s Project
https://sites.google.com/site/growthofdigitalmediacompanies/home

One KML File for Each Year
Logo Size shows Growth

YouTube Video:
http://www.youtube.com/watch?v=QDtDF6_mrkc
Today

• Example Final Projects

• General Project Description
  – Scope
  – Timeline
Define Problem

Find Data

Write a set of instructions

Solution
Scope – Think Big!
(& Have Backup Plans)

Define Problem

Find Data

Write a set of instructions

Utilize MANY data sources to test your claim.

Solution
Scope – Think Big! (& Have Backup Plans)

Define Problem

Find Data

Write a set of instructions

Utilize MANY data sources to test your claim.

Analyze the same dataset in DIFFERENT ways:
Make a number of related claims/hypotheses

Solution
Scope – Think Big!  
(& Have Backup Plans)

Define Problem

Find Data

Write a set of instructions

Utilize MANY data sources to test your claim.

Analyze the same dataset in DIFFERENT ways: Make a number of related claims/hypotheses

GENERALIZE the Program Try to Implement something we haven’t discussed

Solution

CSCI 0931 - Intro. to Comp. for the Humanities and Social Sciences
Scope – Think Big! (& Have Backup Plans)

Define Problem

Analyze the same dataset in DIFFERENT ways:
Make a number of related claims/hypotheses

Write a set of instructions

Find Data

Utilize MANY data sources to test your claim.

GENERALIZE the Program
Try to Implement something we haven’t discussed

Display the Results in Different Ways

Solution
Today

• Example Final Projects

• General Project Description
  – Scope
  – Timeline
## Timeline

<table>
<thead>
<tr>
<th>Sun</th>
<th>Mon</th>
<th>Tues</th>
<th>Wed</th>
<th>Thurs</th>
<th>Fri</th>
<th>Sat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HW 3-4 <strong>Out</strong></td>
<td></td>
<td></td>
<td>Project <strong>Out</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Meet With Staff At Least Once</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>In-Class Work Time (HW &amp; proposal due)</strong></td>
<td><strong>Python Odds &amp; Ends</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/27</td>
<td>4/28</td>
<td>4/29</td>
<td>4/30</td>
<td>5/1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Meet With Staff At Least Once</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>In-Class Work Time (RP)</strong></td>
<td><strong>Computing Applications (RP)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5/4</td>
<td>5/5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Project <strong>Due</strong></td>
<td><strong>Last Class:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>2min Presentations and reflection (and food)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Before You Begin...

Say you have a vague idea of what you want to do....

Define Problem

Find Data

Write a set of instructions

Solution
Before You Begin...

Say you have a vague idea of what you want to do....

Define Problem

Find Data

Where is the data? What does it look like? What format will be easiest for you to analyze?

Write a set of instructions

Solution
Before You Begin...

Say you have a vague idea of what you want to do....

Define Problem

What is your claim?
Are there any sub-claims?
What are your assumptions?

Find Data

Where is the data?
What does it look like?
What format will be easiest for you to analyze?

Write a set of instructions

Solution
Before You Begin...

Say you have a vague idea of what you want to do....

Define Problem

What is your claim?
Are there any sub-claims?
What are your assumptions?

Find Data

Where is the data?
What does it look like?
What format will be easiest for you to analyze?

Write a set of instructions

Solution

What do you want the output to be?
What analysis will best test your claim?
Before You Begin...

Say you have a vague idea of what you want to do....

Define Problem

What is your claim?
Are there any sub-claims?
What are your assumptions?

Find Data

Where is the data?
What does it look like?
What format will be easiest for you to analyze?

Write a set of instructions

What are the steps?
What tools are best for each step?
How can you GENERALIZE for a broader analysis?

What do you want the output to be?
What analysis will best test your claim?

Solution
Things you know

• How to obtain data from the web (*urllib*)
• How to parse XML files in Python (*lxml*)
• How to look for specific patterns in a text (*re*)
  – Advanced pattern matching & transformations
• How to obtain tweets (*tweepy*)
  – From a user’s timeline
  – Based on a search term (+ restricted geographically)
• How to use Google’s geocoding API (*using JSON*)
  – Place name to latitude/longitude
  – Latitude/longitude to place name
• Generate KMLs! (*HW 3-4*)
Things you know

• How to obtain data from the web (*urllib*)
• How to parse XML files in Python (*lxml*)
• How to look for specific patterns in a text (*re*)
  – Advanced pattern matching & transformations
• How to obtain tweets (*tweepy*)
  – From a user’s timeline
  – Based on a search term (+ restricted geographically)
• How to use Google’s geocoding API (*using JSON*)
  – Place name to latitude/longitude
  – Latitude/longitude to place name
• Generate KMLs! (*HW 3-4*)

Use them!
Power to you!
Things to consider

• Don’t reinvent the wheel
  – Use code from activities, homeworks
  – Use your own code

• Start working *early*
  – *Don’t wait* to start after handing in the proposal
  – The proposal is “early feedback” not “pre-feedback”

• Ask for help during *office hours* and the *in-class Project Workshops* (4/21 & 4/28)
Last curves in the roller coaster ride...