CSCI 0931: Introduction to Computation for the Humanities and Social Sciences

Hammurabi Mendes
Spring 2015
http://bost.ocks.org/mike/miserables/
Who am I?
Who are you?
About the Course

Everything you need will be on the website:

http://www.cs.brown.edu/courses/cs0931

• Read “About CSCI 0931” and the Course Missive
• Fill out the Collaboration Policy

Three parts:

• Analyzing Voting Patterns using Spreadsheets
• Analyzing Texts using Python
• Spatial Display of Data and Social Media Data

Homework due almost every class

Two Unit Projects, One Final Project

NO Exams
NO Textbook
Here You’ll Learn How To...

• Collect data relevant to problems in your field
• Investigate data computationally
• Formulate problems as *testable hypotheses*
• Design and execute computations that test your hypotheses
• Teach yourself more computational skills and how to communicate findings
You’ll Have Help

Instructor: Hammurabi Mendes

(Call me just Hammurabi)
You’ll Have Even More Help

**Head TA:** Miles Holland

**TAs:**
Jess Tran
Joseph Engelman
Jake Small
Seriously, a Whole Lot of Help :)
Seriously, a Whole Lot of Help :)

Tom Doeppner
Steve Reiss
Shriram Krishnamurthi

Alexandra Papoutsaki
Expectations/Responsibilities

• We won’t waste your time
• We’ll treat you as adults: no coddling

• You’ll keep up
• You’ll read course materials on your own
• You’ll participate
• You’ll submit only your own work
Expectations

• Attend class.
  – If you don’t attend all classes during the first two weeks, you may be \textit{dropped from the course}
    • – lots of others want to take the course.

• Do the homework
  – If you miss one of the first three homeworks, you may be dropped from the course.

• Do \textit{all} the homework.
  – If you don’t hand in \textit{every} assignment, you don’t pass.

• 6-8 hours per week outside of class time
Grading

- Homework: 50%
- Unit Project #1: 10%
- Unit Project #2: 15%
- Final Project: 25%

1) You must do *all* of them to get credit
2) 85% or more: A / 70% or more: B / 50% or more: C
3) Borderline cases adjusted *up* depending on class engagement
Waitlist

• This course is for Humanities and Social Science undergraduates
• We are making new offers as enrolled students decide to leave
• You must be here in the first three classes to get an offer
• If you enter the waitlist but don’t get a spot, you are automatically in the next offering waitlist
PAUSE FOR EXODUS
Liberal Media Bias
“In 1981 ... I was named a national correspondent, which allowed me to cover bigger, more important stories anywhere in the country ... It was in New York that for the first time I started noticing things that made me feel uneasy.

I noticed that we pointedly identified conservatives as conservatives, for example, but for some crazy reason we didn’t bother to identify liberals as liberals ... in the world of the Jenningses and Brokaws and Rathers, conservatives are out of the mainstream and need to be identified. Liberals, on the other hand, are the mainstream and don’t need to be identified.”

Somewhere in here is a testable hypothesis: *What is it?*
Claim: In the media, conservatives are labeled as “conservative” more often than liberals are labeled as “liberal.”
What’s liberal? Who is liberal? How do you know?
The Most Influential US Conservatives & US Liberals. The Telegraph, Jan 2010

Conservatives

1. Dick Cheney (Former VP)
2. Rush Limbaugh (Talk radio)
3. Matt Drudge (Drudge Report)
4. Sarah Palin (Former AK Gov.)
5. Robert Gates (Former Defense Sec.)

Liberals

1. Barack Obama (President)
2. Hillary Clinton (Former Sec. of State)
3. Nancy Pelosi (Minority Leader of US House)
4. Bill Clinton (Former Pres.)
5. Rahm Emanuel (Chicago Mayor)

Claim: In the media, conservatives are labeled as “conservative” more often than liberals are labeled as “liberal.”

We need to figure out a way to test this hypothesis.
Let’s try it
How long would it take you to count for an entire newspaper?

20 Pages
8 columns/page
6 paragraphs/column
40 words/paragraph
= 38,400 words in a newspaper

If you read 250 words per minute, 38,400 / 250 = 153.6 minutes

Word count of NYTimes front-page articles: \textbf{12,496}
(45 double-spaced pages)
PAUSE
How long would it take a computer to do it?

A little less than 0.05 seconds.

Speedup: 200,000 x
Goal: use computers to answer these types of questions

Define Problem

Find Data

Write a set of instructions

Solution

Computer
What Nunberg Did

• Criterion:
  – Name is distant up to 7 words from label

• Source:
  – Major newspapers

Liberals were labeled more
Tonight’s reading: “On the Bias.” Geoffrey Nunberg, on NPR’s Fresh Air

Things to think about:
• What decisions/assumptions did Nunberg make?
• How did he justify his assumptions?
• What are potential criticisms?
How Computation Enabled It

• When it came time to search for other labels (“Pinko”, “left-wing”, ...) it was a matter of minutes
• Searching different newspapers: minutes
• Searching european papers: minutes
• Searching different liberals or conservatives: minutes
What computation enabled

• Careful analysis
• Verification of robustness
• *Easy Repeatability and Comparison*
  – Seismic shift in political debate
  – Far greater risk for any person making a claim!
  – There’s now no excuse for *not* doing it this way
Final result: Both “solution” and “instructions”!

Define Problem

Find Data

Write a set of instructions

Computer

Solution
Odds & Ends

• Backup everything.
  – Data integrity is your responsibility
• Do all homework and projects on any Windows or Mac laptop.
• Be here! Take a break from email/Facebook
• Keep interesting questions and data sources in mind – you might use them in projects!
Wrapping Up

Due next class:

• Print and sign the Collaboration Policy

• Do HW 0-1 (includes reading/understanding the collaboration policy, and Nunberg’s “On the Bias”)

• Do HW 1-1 (an intro to spreadsheets)
Fin