Your Classmates

- Pair up and introduce yourselves
- Find out:
  - name
  - origin
  - interests
  - fun fact?
- In a few minutes, you will be asked to introduce your partner to the rest of the class

A New Paradigm

- Social science is experiencing a revolution:
  - more ‘big’ data available from governments, non-profits, citizens (social media data)
  - better computing power to analyze it
  - more sophisticated statistical techniques
- It’s changing the way we understand and engage in politics, how we implement policy, and how organizations across the world make decisions

Why Should You Take This Course ( Seriously)?

This course will teach you
1. how to interpret and evaluate social science studies
2. how to answer important political questions using data
3. practical skills, which may lead to a well-paid job conducting data analyses in the private sector, non-profit, academia, ...

Jobs for Data Scientists

“For the smaller category of data scientists, there are more than 1,200 job openings. . . . Nationally, the average base salary for software engineers is $100,000, and $112,000 for data scientists.”

“For Today’s Graduate, Just One Word: Statistics”
(New York Times, August 5, 2009)
Course Goals

In this course, you will
- learn the fundamental principles of statistical inference
- develop programming skills
- combine political data with statistical concepts to answer a wide range of political and policy oriented questions
  - Who is most likely to win the upcoming elections?
  - Is there racial discrimination in the labor market?
  - Do countries become less democratic when leaders are assassinated?

Specific Skills

In this course, you will learn how to
- use and understand political data
- link political questions to statistical solutions
- write R code
- import and manipulate data
- calculate statistical quantities of interest
- assess our confidence in these quantities
- visualize data
- be an informed consumer of statistical evidence

Why R?

- It’s what employers are looking for
- It’s FREE and open-source
- It’s flexible
- It’s more powerful than STATA, SPSS, and SAS
- It’s a full programming language; once you understand how to use it, you can learn other languages too

Selected Employers Using R

- Bank of America
- Bing
- ANZ Bank
- Facebook
- FDA
- Ford
- Human Rights Data Analysis Group
- Kickstarter
- Microsoft
- Firefox
- The New York Times
- OkCupid
- Orbitz
- Thomas Cook
- Trulia
- Twitter
- Uber

How to Learn R: Practice, Practice, Practice!

- Traditional lecture approach is bad for learning code
- You can only learn by doing
- You will code each concept multiple times
  - in lecture
  - in problem sets
  - in exams

No Coding Pre-Req

- You come from a wide range of backgrounds
- The course is setup so anyone can take it
- Knowing how to code is NOT a pre-req for the class
- We will start from scratch!
In It Together

- During lectures, you will be allowed to consult your peers
- In problem sets, collaboration with your own peers or with CLAS tutors is encouraged BUT you are responsible for your own code and your own answers
  - do not, under any circumstances, copy another person’s code
  - collaborators should be identified in each pset
- In exams, NO collaboration will be permitted
- Questions about pset or other course related materials should be posted on the discussion forum
  - everybody benefits from answers
  - students can answer each other’s questions

The Textbook


- It has not been published yet
- Available on the course website
- Please repay the generosity of the author and publisher by not circulating it!
- Important that you do the readings before coming to class
  - We will start each class with a multiple choice question about readings (counts towards participation grade)
  - See syllabus for tentative schedule

The Course Website

- Course website on blackboard
- Where textbook and other useful materials will be available
- Where lecture slides and psets will be posted
- Where announcements will be made
- Where discussion forum is hosted

Evaluation

- Final grade:
  - problem sets (40%)
  - midterm (20%)
  - final exam (30%)
  - class participation (10%)
- Exams
  - NO collaboration will be permitted
  - BUT you will be allowed a single letter-sized sheet of notes (two-sided)
- I recommend you take notes as you do the readings/problem sets and attend lecture
  - it will make writing the cheat sheet easier

Cumulative Material

- The course is cumulative
  - materials later in the course assume that you know what was covered earlier in the course
- Attendance is mandatory
- Important to keep up with assigned readings
  - not all concepts and commands will be covered in class
- Important to ask questions if you are confused

Administrative Details

- Take a few minutes to look over syllabus
  - notice that attendance is mandatory; any absences must be justified by a letter from the Dean of Students
  - students are expected to arrive to class on time
  - psets are due on Tuesdays
  - I will not answer emails during weekends or after 5pm on weekdays
- Questions?
- How many of you have laptops?
  - Bring them on Thursday
- For Thursday, do Homework 0 and first set of readings (1.3)
Details on blackboard under homework
Due by beginning of class on Thursday
Estimated completion time: 1-2 hours
If you have issues:
  ▶ First, try Google and StackOverflow
  ▶ If that fails, post on discussion forum

1. Register and certify your phone number with www.polleverywhere.com (instructions on blackboard)
2. Complete the course survey (link on blackboard)
3. Get your computer ready (instructions on blackboard)
   ▶ If you are going to use your own home computer or laptop to complete your homework:
     ▶ Install R AND RStudio
     ▶ Create a folder in your Desktop called ‘GVT201’
   ▶ OR, if you are going to use a Suffolk or other public computer to complete your homework:
     ▶ Familiarize yourself with Suffolk Virtual Desktop (it has R and RStudio already installed)
     ▶ Create a folder in your Desktop called ‘GVT201’