Political Science 270
Understanding Political Numbers

Instructor: Michael DeCrescenzo

University of Wisconsin–Madison
Spring 2019

(Updated February 04, 2019)

Credits: 3
Lecture: MW 11:00–11:50 a.m., Ingraham 222
Website: Canvas and https://mikedecr.github.io/ps-270/

E-mail: decrescenzo@wisc.edu
Office Hours: M 2:30–4:30 p.m.
Office: North Hall 101A

TA: Michael Masterson (masterson2@wisc.edu)
Office Hours: WR 9:45–10:45 a.m. (121 North Hall)

Course Description

During the 2016 presidential election campaign, an election forecast developed by Nate Silver (editor in chief of FiveThirtyEight) predicted that Hillary Clinton had roughly a 70% chance of winning the presidency.[1] Other forecasts predicted a Clinton victory with about 90% probability.[2] Donald Trump, the underdog in these forecasts, was the eventual winner. These forecasts raised many questions about data, probability, and political phenomena. Voters wondered, where do these forecasts come from, what data do they use, and should they be trusted? Academics and journalists asked a different set of questions: were these forecasts "wrong," did they do a good job communicating their *uncertainty* to the public, and are these forecasts ethically responsible in the first place?

These are all questions related to *political numbers*—the use of data and statistics to analyze the events of the world, promote public policies, or build social theories. Data are increasingly prevalent in politics—and in journalism, industry ("data science"), and academics—but data are complex and challenging. What can we learn about politics (and the social world more broadly) when we take a scientific, empirical view of it? When should we trust data, and when will data lead us astray? In this class, we will learn to consume, criticize, and produce our own political numbers. On the consumption side, we will discuss how data are used in journalism, policy advocacy, and academic research. On the production side, we will learn the fundamentals of computational data analysis and data visualization.

[1] https://projects.fivethirtyeight.com/2016-election-forecast/#plus

[2] Nate Silver's explanation of the differences between models: https://fivethirtyeight.com/features/why-fivethirtyeight-gave-trump-a-better-chance-than-almost-anyone-else/

Learning Outcomes

Here are a few different ways to express our objectives for the semester:

1. Think critically, skeptically, and *ethically* about the use of data in politics and policy debate.
2. Evaluate whether ideas about the social world are supported by evidence.
3. Ask relevant, specific, and *falsifiable* research questions.
4. Collect, analyze, and visualize real-world data.
5. Build computational, statistical, and communicative skills.

The skills you build in this course will transfer to other areas of school and life. If you learn to ask the right questions, develop evidence-based answers, and communicate those answers clearly to an audience, then you will get an "A" on every term paper and impress future employers.

This course fulfills the Quantitative Reasoning *B* (QR-B) requirement. Students must complete their QR-A before enrolling. The instructional mode is all face-to-face and follows the "Traditional/Carnegie" credit model.[3] Other designations: social science breadth, elementary level, L&S credit.

Computing and Technical Requirements

We will learn data analysis by *doing* data analysis. We will use the open-source (and free!) statistical programming language called **R**. R is very popular in data science and statistics, it makes beautiful graphics, and it is great fun. A few things to flag:

- **This course requires programming.** R is a programming language; it works by writing commands to manipulate data. *This will require practice*, but rest assured that students excel at learning R every semester.
- **This course has no programming prerequisites,** but you should have some familiarity with your computer (e.g. installing applications, creating folders). Every programing language has its quirks, but R is a good language for beginners, and we approach it carefully.
- **This course has no statistical prerequisites.** We will discuss and employ statistical concepts, but we will introduce and practice these concepts as we go.
- **We *must* use math.** There's no avoiding it. We will not go *crazy*, we will not use calculus, but we will work with some equations and algebra concepts that you have seen before.

In order to fulfill the computing requirements for the course, you must…

- **Have a computer for you can use.**[4]
- **Install R.** Download the appropriate version for your computer at `https://cloud.r-project.org/` and then install.
- **Install RStudio.** RStudio is an app that makes it easier to write and run R code. Imagine that R is the engine, and RStudio is the dashboard. Download the free *desktop version* at

[3] At least two hours of outside work for every one hour spent in class, per the university's credit policy: `https://teachlearn.provost.wisc.edu/course-syllabi/`

[4] The university has laptop-lending programs for students who need one. If this applies to you, come see me so we can set this up. Sections are held in a computer lab.

https://www.rstudio.com/products/rstudio/download/. Be sure to install it *fully*.[5]

Seeking Help

Computer programming can be frustrating because computers are stupid, and error messages are not always helpful. Fortunately, help is out there. StackOverflow is a Q&A site dedicated to helping people with their programming problems. RStudio Community is a similar concept, but tailored for people using RStudio and "Tidyverse" packages (which applies to us). Google is helpful, but be sure to search for "rstats" instead of "r."

Your teaching assistant and I can provide some help with R by email and in office hours, but please refrain from reaching out to us *until you have exhausted other sources of help*. Why? Because navigating online resources is an essential skill for learning any programming language. You will learn R better by learning to help yourself!

Class Setup

Lectures will discuss broad concepts in political numbers, social science research, "data science," and statistics. We will cover some R material, but the majority of R instruction will take place in lab.

I will post lecture slides online after class, but you should be aware: I do not design my slides as outlines for your personal notes. They are not substitutes for attendance and individual note-taking.

Discussion sections (sometimes referred to as "lab") will be a space for practicing the hairy details: programming with R, reviewing useful math concepts, and practicing statistical concepts.

Readings and other Materials

We will refer to material from several helpful texts. Luckily, they are all online and free, so bookmark them *now*. Primarily we will rely on…

- *Modern Dive* (Ismay and Kim, https://moderndive.com)
- *R for Data Science* (Grolemund and Wickham, https://r4ds.had.co.nz/)
- *Data Visualization: A practical introduction* (Healy, http://socviz.co/)

I have also acquired an educational license for DataCamp, an online video lecture platform for learning to program. I strongly recommend that you practice the DataCamp lessons in the course schedule as we advance. I will send an email with information about setting this up soon.

Other readings are listed in the schedule below. Some other interesting sites for learning about political numbers include Flowing Data, FiveThirtyEight, Dear Data, and Political Violence at a Glance.

[5] On some computers, it is possible to *run* Rstudio without having fully installed it. This mostly applies to Mac users, who must be sure to drag the RStudio app icon into their "applications" folder and then delete the installer from their computer.

Deliverables

Grades are intended to assess your learning about (1) data and code, (2) scientific and critical reasoning, and (3) statistical reasoning. Your grade will be computed using…

- Three short exercises (20%)
- Two short essays (30%)
- Research project (50%), based on the following components:
  - Proposal (5%)
  - Data (10%)
  - Presentation (5%)
  - Final paper (30%)

There are *no exams*, and aside from the presentation "tournament" (see below), there will be no extra credit either. All assignments and code shall be submitted on Canvas.[6] The following grading scale will be used (note the closed and open intervals):

- A: [93, 100]
- AB: [88, 93)
- B: [83, 88)
- BC: [78, 82)
- C: [70, 78)
- D: [60, 70)
- F: [0, 60)

Exercises are designed to practice essential skills of R programming and analysis. These will be brief, oriented around data manipulation, data visualization, and data analysis.

Short essays are designed to blend critical thinking with technical skills. These assignments require you to make an argument and support it with evidence.

The research project is an original analysis of a topic of your choosing. Ask an original research question, gather relevant data, and conduct an analysis (using statistical tools and graphics) to address the question. This assignment is completed in stages; we separately ask for a research proposal, the data you are using, a brief presentation in your discussion section, and the final paper itself (due on the final exam date).

Every student will deliver a brief (5–7 minutes) presentation of their research project in their discussion section. Each discussion section will nominate the best project from their section to advance to "the finals." Finalists will present their project to the entire class in the final lecture of the semester. All finalists will be guaranteed at least an *AB* grade in the course. The winning presentation, as determined by a vote of the class and the instructors, will earn a guaranteed *A* for the course.[7]

[6] Code should run without errors, and written work should be uploaded in PDF format (which have a `.pdf` file extension; for instructions on converting Word documents to PDF, visit https://www.bettercloud.com/monitor/the-academy/save-word-doc-pdf/).

[7] Yes, you can still earn an *A* or *AB* without placing in the tournament.

Grading and Assignment Policies

Collaboration: Collaboration and consultation with your classmates will help you learn. I encourage good-natured collaboration on assignments, but you must turn in your own work. Plagiarism is easy to spot—I know because I have spotted it in the past—and will not be tolerated (see "Academic Integrity" below). Assignment sheets may contain additional guidance on acceptable forms of collaboration, and I am *more than happy* to answer your questions on acceptable forms of collaboration.

Attendance: Your attendance is your own responsibility, but skipping class is sure to harm your grade. Do you already know how statistics and R? If "No," you need to come to class. Consult your classmates for notes from days you miss. If difficult life circumstances are regularly affecting your ability to attend class, please speak with me ASAP so we can create accommodations.

Late work: We will accept late work for certain assignments. We penalize late work by 10 percentage points, plus an additional 10 points for each 24-hour period that passes after the assignment's due date. We can't accept late work for exercises because we review the solutions the day they are due.

Grade disputes: Mistakes happen. If you want to dispute a grade, give us a principled argument about how your work was better than the grade it received. Grade discussions must be initiated (1) in person, i.e. during office hours or before/after class, and (2) at least 24 hours after a grade is returned.

Academic Integrity

We encourage collaboration but take academic integrity and honesty very seriously. If you cheat, fabricate, plagiarize, collaborate outside acceptable limits,or help others commit these acts, you can face disciplinary action including but not limited to failing an assignment, failing the course, or facing more serious disciplinary action dispensed by higher university authorities. Substantial or repeated cases of misconduct will be referred to the Office of Student Conduct & Community Standards for additional review. For more, see https://conduct.students.wisc.edu/academic-integrity/. I am happy to answer questions about responsible citation practices and other concerns!

Diversity and Inclusion

Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals. The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background—people who as students, faculty, and staff serve Wisconsin and the world. Visit https://diversity.wisc.edu/ for more.

Accommodations

The University of Wisconsin–Madison supports the right of all enrolled students to a full and equal educational opportunity.

The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy

(Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA. For more, visit http://mcburney.wisc.edu/facstaffother/faculty/syllabus.php.

Other extraordinary life circumstances can interfere with class performance. There are various resources available to help students through difficult times, including the Dean of Students Office (https://doso.students.wisc.edu/) and the Division of Student Life (https://students.wisc.edu/).

Please discuss any potential accommodations with me ASAP.

Email Policies

It is virtually certain that I will send important emails to update you on assignments and course materials. You will be expected to read and act on the information contained in emails.

I am usually quick to respond to student emails, but some warnings:

- I may not answer questions directly and instead refer you to the syllabus, a textbook chapter, or other materials I distribute for class.
- We give at least a week and a half to work on every assignment. That's more than enough time to identify issues and seek help. Eleventh-hour emails about an assignment are not guaranteed an immediate response.
- Grade disputes should be raised in person, not over email (see above).

Computers in Class

This class requires programming, so computers are allowed in class *for work purposes*. I won't stop you if you choose to waste your time in class, but I care very much about whether your computer misuse is distracting to your classmates. I have authorized your TA to snitch on irresponsible computer use in class. Consider installing an app on your computer that blocks distracting websites during class.

Class Schedule

Review each lecture's associated readings before coming to class. Please consult lectures and email for updates to the schedule (depending on the progression of the course). Other copies of this schedule can be found on Canvas and at `https://mikedecr.github.io/ps-270/`.

Foundations: Science and evidence

Wednesday, January 23: Introduction

Monday, January 28: Evidence and empiricism

- Olugbenga Ajilore, "Why Social Science? Because It Can Challenge Conventional Wisdom"
- Margherita Belgioioso, "Is Radical Islam the Main Terrorist Threat to the EU?"

Wednesday, January 30: Theorizing, hypothesizing, falsification

- Kieran Healy, "F**k Nuance."

Social Science Data: What is it and why does it suck?

Monday, February 4: Measurement and causality

- Lynn Vavreck, "A Campaign Dollar's Power Is More Valuable to a Challenger"
- (Skim) Jenny Bryan, "Naming Things"

Wednesday, February 6: Data

- (Practice) *R for Data Science*, "Workflow: basics"
- (Practice) Datacamp's *Free Introduction to R*, Ch. 1

Graphics: The "fun" part of data analysis

Monday, February 11: Good, bad, and ugly graphics

- *Data Visualization: A Practical Introduction*, "Look at Data"
- (Browse) Claus Wilke, *Fundamentals of Data Visualization*, "Directory of Visualizations"

Wednesday, February 13: Graphics in R

- **DUE:** Essay 1 (Brainstorming a Study)
- (Practice) *Data Visualization*, "Make a Plot"

Signal and Noise

Monday, February 18: Averages and aggregation

- *R for Data Science*, "Piping"
- (Practice) *R for Data Science*, "Data transformation"

- (Practice) Datacamp's *Introduction to the Tidyverse*, "Data Wrangling" (up through the 'Filtering and arranging' lesson)

Wednesday, February 20: Variation and randomness

- (Watch the first 5.5 mins) The Galton Board
- (Practice) *Modern Dive*, "Data Visualization via `ggplot2`"

Monday, February 25: "Piping" data in R

X and Y: Relationships and regression

Wednesday, February 27: Relationships as "conditional expectations"

- **DUE:** Exercise 1 (Polls)

Monday, March 4: Linear regression

Wednesday, March 6: Statistical significance and hypothesis testing

- **DUE:** Project proposal/research question

Fancy Regression

Monday, March 11: Multiple regression

Wednesday, March 13: Measurement and finding data

- **DUE:** Exercise 2 (WI Recount)
- Nick Barrowman, "Why Data Is Never Raw"

Monday, March 18: *Spring Break* (Collect your project data!)

Wednesday, March 20: *Spring Break* (Collect your project data!)

Monday, March 25: Indicators and interactions

- Enos, Fowler, and Vavreck, "Increasing Inequality: The Effect of GOTV Mobilization on the Composition of the Electorate." *The Journal of Politics*

Wednesday, March 27: (Spooky voice) Nonlinear relationships

- **DUE:** Project data
- Watch: The Zipf Mystery
- Watch: "How I feel about logarithms"
- (Practice) "Logarithms… How?"
- (Practice) *R for Data Science*, "Model Building"

Political Numbers "In the Wild"

Monday, April 1: Evaluating public policy

- Watch: "Regression to the mean"

Wednesday, April 3: Experiments and interventions

- Derek Willis, "Professors' Research Project Stirs Political Outrage in Montana"

Monday, April 8: Big data, ethics, and power

- Claudia Brugman, "Why Social Science? Because Social Science Research and Education Are Critical for National Security"
- Listen: *99% Invisible*, "The Age of the Algorithm"

Wednesday, April 10: Elections and campaigns

- **DUE:** Essay 2 (World Health Data)
- Rothschild and Goeld, "When You Hear the Margin of Error Is Plus or Minus 3 Percent, Think 7 Instead"
- (Practice) *Data Visualization*, "Work with Models"

Monday, April 15: Writing a research paper

- Tiffany Barnes and Stephanie Burchard, "'Engendering' Politics: The Impact of Descriptive Representation on Women's Political Engagement in Sub-Saharan Africa"
- Steven V. Miller, "Assorted Tips for Students on Writing Research Papers"

Wednesday, April 17: Data and the Judiciary

- **DUE:** Presentations (week 1)
- Kenneth R. Mayer, "Is Political Science Relevant? Ask an Expert Witness." *The Forum*

Monday, April 22: Election Forecasting

- (Listen) "Opinion Polling for Presidential Elections", *Data Skeptic* podcast

Wrapping Up/Looking Ahead

Wednesday, April 24: Big problems in political and social research

- **DUE:** Presentations (week 2)
- Watch: "Is Most Published Research Wrong?"

Monday, April 29: How people have tried to solve big problems

- (Skim & Bookmark) Kieran Healy, *The Plain Person's Guide to Plain Text Social Science*
- (Skim & Bookmark) Jenny Bryan, "Excuse me, do you have a moment to talk about version control?"
- (Skim & Bookmark) Jenny Bryan et al., *Happy Git and GitHub for the useR*

Wednesday, May 1: Finalists present

Research Paper due May 6, 11:59 p.m.