

This version: November 17, 2016

Political Science 172

Advanced Social Data Analytics

Fall Quarter 2016
SSB 104, Th 3:00-5:50 p.m

Molly Roberts
SSB 399
meroberts@ucsd.edu
Office Hours: Wednesday 4-6PM

Overview

Data analytics are changing the way in which policymakers implement policies, organizations and businesses decide on action plans, and individuals make decisions on how to live their lives. This course will overview methods and tools for empirical data analysis in social science research, methods that can be applied to a large range of questions ranging from policy to business to your everyday life. How do we predict the outcome of presidential elections? Do countries become less democratic when leaders are assassinated? How can data analysts detect election fraud? What is the effect of an increase in the minimum wage? These are a few among the many social science questions we will tackle in this course.

This course will teach you statistical techniques that are useful for the description of, and inference from, observed data for the purpose of testing theories and discovering empirical regularities in social science data. The course will also familiarize you with R, a statistical package for analyzing data. You will leave this course prepared work in any setting that requires a social scientific approach to data analytics, from policy non-profits to government, from Silicon Valley to Wall Street and beyond.

This course does requires PS 30 or a basic statistics class as a prerequisite.

Assessment

There are no written exams in the class, and your grade will be based on a combination of:

- **Homeworks (40%):** Problem sets will be given throughout the quarter, skewed heavily toward the beginning of the quarter. Problem sets will contain analytical, computational, and data analysis questions. Each problem set will be counted equally toward the calculation of the final grade. The following instructions will apply to all problem sets unless otherwise noted.

- Late submission will not be accepted unless you receive special permission from the instructor in advance.
 - One hard-copy of the homework write-up should be turned in in class, and a copy of the homework write-up and accompanying code should be turned in electronically.
 - Working in groups is encouraged for conceptual and sometimes technical discussion, but each student must submit their own writeup of the solutions that shows their independent work on the assignment. In particular, you should not copy someone else's answers or computer code. We also ask you to write down the names of the other students with whom you solved the problems together on the first sheet of your solutions. One problem set will serve as a midterm where you will not be allowed to collaborate with other students.
 - For analytical questions, you should include your intermediate steps, as well as comments on those steps when appropriate. For data analysis questions, include annotated code as part of your answers. All results should be presented so that they can be easily understood.
- **Final project (50%):** The final project will be a short research paper that applies a method learned in this course in an original way to an empirical problem of your substantive interest.
 - Though not required, I encourage you to work with another student on your final project.
 - The most time-consuming part of the research project will likely be locating and collecting the data. Please begin thinking about what data you might like to use as soon as possible. Once you have decided on the data and research question, send an e-mail to me describing the question you are planning on answering at the data you will be using. The deadline for this assignment is October 6, but the sooner you can begin, the better.
- Students are expected to adhere to the following deadlines:
 - October 6: E-mail me with a one paragraph description of the question you are planning to answer and the data you will be using for the project.
 - October 27: Turn in a two page prospectus outlining your proposed project and proposed dataset. This prospectus should:
 - * introduce the topic or topics you are going to study
 - * explain why these topics matter - why are they worthy of study?
 - * explain your theory. That is, how do you expect the dependent variable (the thing you are trying to explain) to be related to the independent variable (the thing that does the explaining)?
 - * precisely state your hypothesis. That is, how will you test your theory? What evidence will you need to see to believe your theory is the correct one? What evidence would suggest your theory was wrong?
 - * explain your research strategy: what data will you use, and where will you get it? What statistical test will you use?
 - * explain what you expect to find. By this date you need to have acquired the data you plan to use, and have a sense of the question you want to answer.

- November 30: Flash talks. Each group will present a five minute presentation about their research.
- December 2: Final project due. Please turn in one printed copy by the end of the day, and email electronic copies to the instructor. The research paper should have the following sections:
 - * Introduction: what is the topic of interest and why is it of interest? Broadly, what is the paper about, and what will it suggest?
 - * Literature: what do we know about this topic so far? Who has worked on this, and how does their work relate to yours?
 - * Theory and Hypotheses: how do you think the dependent and independent variables relate? What precisely will you be testing?
 - * Data and Methods: describe your data and what methods/techniques/approaches you are using to test your theory
 - * Results: what did you find? Use some of the techniques we learned in the course to show your results with tables and graphs.
 - * Discussion/Limitations: summarize the findings and explain why they matter. what are the limitations of your analysis? What further follow-up work needs to be done? how have you contributed to the debate or literature in this part of political science?
 - * The paper should be double-spaced, 12pt font, 1-inch (or so) all round margins, no more than 3000 words (or about 12 pages).
- **Participation (10%):** Students are strongly encouraged to ask questions and actively participate in discussions during lectures.

Academic Honesty and Plagiarism

All of your graded work must be done by you. If you are unfamiliar with the University's policy on academic integrity, please see <http://students.ucsd.edu/academics/academic-integrity/policy.html>.

Syllabus and Plan

The syllabus will be updated periodically throughout the course, so that we can keep with the cadence of the class. I will post to Piazza when such updates are made.

Reading and Textbooks

A First Course in Quantitative Social Science, by Kosuke Imai (Princeton University Press)

The required textbook is forthcoming at Princeton University Press. PUP and the author have kindly given us permission to use the textbook free of charge in advance of its official release. The PDF of the book will be made available on a collaborative reading website called Perusall. Please repay their generosity by not circulating it.

To access the textbook, go to perusall.com and register for an account. You can use Google, Facebook, or Twitter to register, or you can register with your e-mail address. Once you have registered and logged in, you will see a button at the top of the screen “Enter an access code”. To enroll in the class and gain access to the book, enter the access code ROBERTS-7278.

Perusall allows us to collectively read the textbook together. I will highlight and make comments on particular sections of the reading. You can also ask questions by highlighting and commenting on the text. Your participation in perusal will count toward your participation grade.

Piazza

We will be using Piazza for general discussion and questions and answers throughout the class. Piazza allows students to see other students’ questions and learn from them as well as answer them. After the first day of class, I will add you to the Piazza list. Your respectful and thoughtful participation in the discussion forum will count toward your participation grade. Please do not e-mail the instructor with questions (post them on Piazza!) unless they are personal in nature. I will check the Piazza forum daily to provide my own answers and contributions.

Software

We will be using R an open-source statistical package. You can download it from the web here:

<http://cran.r-project.org/>

RStudio is a useful tool for coding in R. You can download it from the web here:

<https://www.rstudio.com/>

COURSE SCHEDULE

1 September 22: Course Introduction and Introduction to R

2 September 29-October 6: Causality and Introduction to R

Reading

Imai, Chapter 1

Imai, 2.1-2.5

3 October 13: Describing a Single Variable and Relating Variables to Each Other

Reading

Imai, 2.6, 3.1-3.3

Imai, 3.4-3.7

4 October 27: Prediction and Regression I

Reading

Imai, 4.1

5 November 3: Regression and Causation

Reading

Imai, 4.2-4.3

6 November 10: Uncertainty in Estimation

Reading

Imai, 7.1

7 November 17: Hypothesis Testing and Regression

Reading

Imai, 7.2

8 November 24: Putting it All Together (Video Lecture)

Reading

Imai, Chapter 5

9 December 1: Data of the Future (Text, Spatial, Network) and Flash Talks

Reading

Imai, 7.3