PSCI T280: Data Analysis for Activists

Spring 2019 – Face-to-face Edition
Drexel University

Monday and Wednesday, 2:00-3:50 PM
Randell Hall, Room 329

Instructor: Jack Santucci, Ph.D. (jas948@drexel.edu)
Office: MacAlister Hall 3021
Appointments: https://jacksantucci.acuityscheduling.com

1   Overview & goals

One goal of this course is to get you more comfortable with computer programming. Most of the jobs I ever got were because I could program – first to build web sites, then to create facts from numerical data. The ability to program is powerful. It might become essential to individual prosperity.

Another goal of this course is to teach you how to think scientifically. We are going to cover four ways to analyze and depict data: linear regression (which may be review), ideal point estimation, social network analysis, and machine learning. Each of these tools lets you draw a picture of politics and make statements about power. The ability to do this comes with responsibility. One way to exercise responsibility is to follow the rules of science.

Finally, we will practice writing. Good writing is brief (or it won't be read), clear (or it confuses), and honest (otherwise it is propaganda). Writing briefly, clearly, and honestly is one good way to think more scientifically.

When this course is over, you will:

- Feel more comfortable using R and consulting the help files. You won't become an expert programmer, but you will know what to do next.

- Know some R basics, which we will learn as we practice the above techniques: indexing, data merges, custom functions, cleaning data, processing text, loading in-development packages, making the computer do the same thing repeatedly (e.g., with loops and the `apply()` family), etc.

- Have exposure to some new-ish, non-regression tools of data analysis and visualization.

- Be a sharper thinker-writer.

2   Things you need in this course

1. You need a **laptop** when we're in class. We are going to code together.

2. We will install the R statistical package in class on the first day: `https://www.r-project.org/`.

3 Requirements

Your grade will be based on:

- Attendance (20%) – Missing class will hurt you. Each day's lesson builds on what came before.

- Two short papers (20% each) – These will test your ability to be critical of some data-driven result. The second of these also will test your ability to produce such a result.

- Final project (40%) – You will find or build a data set on a topic of your choice, analyze it with some tool of choice, visualize the result, and explain its key point crisply in a **paper** (65% of the grade). Part of this grade will be based on a **presentation** you give in class (35% of the grade). The purpose of this presentation is to get feedback on your results, so that you can write a stronger final paper.

4 Skim most of the readings; revisit them as needed

Some of our readings are long and technical, which makes them difficult. I have assigned those readings in order to make you aware that they exist, then get you familiar with ideas about human behavior that underpin the tools we'll use in class. Skim these readings to get a sense of what they say. Then, refer back to them as needed.

5 How to get what you want efficiently by sending me e-mail

1. Use a subject line that indicates your need. Please don't reply to a blanket email sent through Blackboard, nor to a thread about some other issue.

2. Include the full course number (i.e., PSCI-T280-001).

3. Be as specific as possible, and use complete sentences.

4. Keep all correspondence with respect to the same issue in the same thread.

5. Use your Drexel email so that I can search my inbox for your username.

6 Other policies

You agree to:

1. Treat guest speakers and classmates with respect.

2. Check Drexel email daily through the date of the final exam.

3. Attend class and meet course deadlines. Out-of-class written work loses five points for each 24-hour period past the time and date it is due.

4. Use Internet search or a dictionary to look up words you do not understand.

5. Abide by all other Drexel policies, found at the following websites. Note: I reserve the right to use plagiarism-detection software.

  Academic integrity: http://www.drexel.edu/provost/policies/academic_dishonesty.asp

  Disability accommodation: http://drexel.edu/oed/disabilityResources/students/

  Add/Drop: http://www.drexel.edu/provost/policies/course-add-drop

  Course withdrawal: http://drexel.edu/provost/policies/course-withdrawal

I agree to:

1. Abide by any grading guidelines in this syllabus and related instructions for assignments. Grading is on the 0-100 scale: 97 and up (A+), 93-96 (A), 90-92 (A-), 87-89 (B+), 83-87 (B), 80-82 (B-), and so on, down to 50 (F).

2. Answer student e-mail within 48 business hours. Business hours are 9 AM to 5:30 PM, Monday through Friday, non-holiday.

3. Except for the required books, post links to readings on Blackboard, if not the readings themselves.

4. Grade written work within 10 business days of submission.

5. Notify you of changes to this syllabus at least one week in advance of affected due dates.

6. Give office hours on a by-appointment basis. (See link above.) Students who are not in Philadelphia can book these times for phone conversations.

7 Schedule of work

7.1 Rules of evidence: hypotheses and theories

April 1 Naessan, Petter A. 2005. "*On Bullsh*t* by Harry Frankfurt," *Philosophy Now*, `https://bit.ly/2b6Zix1`. **Coding goals:** installing R.

April 3 Cameron, Charles. 2009. "What is Political Science?" In *A Quantitative Tour of the Social Sciences*, Andrew Gelman and Jeronimo Cortina (eds.), pp. 207-222, New York: Cambridge University Press. **Coding goals:** setting up a working directory, scripting, and documenting code.

7.2 Simple models of relationships between (among?) variables

April 8 Sykes, Alan O. 1993. "An Introduction to Regression Analysis." Working Paper No. 20, Coase-Sandor Institute for Law & Economics. **Coding goals:** generate fake data, and play with the linear model.

April 10 Taagepera, Rein. 2018. "Science Walks on Two Legs, but Social Sciences Try to Hop on One." *International Political Science Review* 39 (1): 145-159. **Coding goals:** replicate the plot from my last lecture (random draws, line of best fit, adding line segments for residuals).

April 15 Back to Sykes (1993). **Coding goals:** work with real data from the American National Election Study.

April 17 Back to Sykes (1993). **Coding goals:** plotting bivariate scatter, add prediction and confidence interval, and plot coefficient estimates.

7.3 Models of who agrees with whom on what issues

April 22 Poole, Keith and Howard Rosenthal. 1997. "Introduction: The Liberal/Conservative Structure," 3-10. In *Congress: A Political-Economic History of Roll Call Voting*, New York: Oxford University Press. **Coding goals:** play with data gathered from our survey. **First short paper is due.**

April 24 Poole and Rosenthal, "The Spatial Model and Congressional Voting" and "The Spatial Model: Accuracy and Dimensionality," 11-57. **Coding goals:** estimate and plot ideal points via optimal classification.

April 29 Kruschke, John K. 2015. "Bayesian Item Response Theory in JAGS: A Hierarchical Two Parameter Logistic Model." Online: `https://bit.ly/2V84Gt5`. **Coding goals:** introduction to Bayesian statistics and the 2PL ideal-point model.

May 1 Armstrong, David A., Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal. 2014. "Unfolding Analysis of Binary Choice Data," 183-275. In *Analyzing Spatial Models of Choice and Judgment with R*, Boca Raton: Taylor & Francis. **Coding goals:** move from our survey data to real data from Congress. Compare W-NOMINATE and OC.

7.4 Models of who does and does not talk to whom

May 6 Barberá, Pablo. 2017. "Introduction to social network analysis with R."
Online: `https://bit.ly/2LjEBCS`.

May 8 Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360-1380. **Coding goals:** making an edge list and doing social network analysis.

May 13 Victor, Jennifer Nicoll and Nils Ringe. 2009. "The Social Utility of Informal Institutions: Caucuses as Networks in the 110th U.S. House of Representatives." *American Politics Research* 37 (5): 742-766.

7.5   Machine learning

May 15 Hegelich, Simon. 2016. "Decision Trees and Random Forests: Machine Learning Techniques to Classify Rare Events." *European Policy Analysis* 2 (1): 98-120. **Coding goals:** setting up a decision tree.

May 20 No reading. **Coding goals:** training and testing data sets, pruning, and visualization.

7.6   "Scraping" data from the Internet

May 22 Wickham, Hadley. 2014. "rvest: Easy Web Scraping with R." *RStudio Blog*, `https://blog.rstudio.com/2014/11/24/rvest-easy-web-scraping-with-r/`.

7.7   Improving data visualizations

May 29 Two items:

- Grogan-Kaylor, Andy. 2019. "Quick Introduction to ggplot2," `https://bit.ly/2GXlIijP`.
- Piwek, Lukasz. 2017. "Tufte in R," `http://motioninsocial.com/tufte/`.

7.8   Project presentations and feedback

June 3 and 5 Please use the feedback you get here to improve your final paper.