

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 even become essential to economic survival.

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.
- 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 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.

5 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.

6 Schedule of work

6.1 Rules of evidence: hypotheses and theories

April 1 Naessan, Petter A. 2005. “*On Bullshit* 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.

April 8 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:** generate fake data, and play with the linear model.

6.2 Simple models of relationships between variables

April 10 Sykes, Alan O. 1993. “An Introduction to Regression Analysis.” Working Paper No. 20, Coase-Sandor Institute for Law & Economics. **Coding goals:** replicate the plot from my last lecture (random draws, line of best fit, adding line segments for residuals).

April 15 Continue Sykes (1993). **Coding goals:** work with real data.

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

6.3 Getting and cleaning your own data

April 22 First short paper is due. Scraping the Web and PDF files, writing custom functions.

6.4 Models of who agrees with whom on what issues

April 24 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:** installing and loading packages, getting data from the Web.

April 29 Poole and Rosenthal, "The Spatial Model and Congressional Voting," 11-26. **Coding goals:** estimate and plot ideal points via optimal classification.

May 1 Poole and Rosenthal, "The Spatial Model: Accuracy and Dimensionality," 27-57. **Coding goals:** interrogate model fit and bill parameters.

May 6 Hare, Christopher, Tzu-Ping Liu, and Robert N. Lupton. 2018. "What Ordered Optimal Classification Reveals about Ideological Structure, Cleavages, and Polarization in the American Mass Public." *Public Choice* 176 (1-2): 57-78. **Coding goals:** subsetting data, recoding variables, and loading in-development packages.

6.5 More on getting data, the final project

May 8 Second short paper is due. Read: "Writing Effective Op-Eds" at <https://bit.ly/2TBr78A>.

6.6 Models of who talks to whom

May 13 Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360-1380.

May 15 Continue Granovetter (1973). **Coding goals:** making an edge list and doing social network analysis.

May 20 Continue Granovetter (1973). **Coding goals:** measures of centrality and node influence.

6.7 Machine learning

May 22 Hegelich, Simon. 2016. "Decision Trees and Random Forests: Machine Learning Techniques to Classify Rare Events." *European Policy Analysis* 2 (1): 98-120.

May 29 No reading. **Coding goals:** training and testing data sets.

June 3 No reading. **Coding goals:** pruning the tree and visualizing results.

6.8 Group feedback on your projects

June 5, 10, and 12 In-class presentations.