

POSC 499: Data Analysis

Spring 2017

James Lo

Department of Political Science, University of Southern California

This syllabus is tentative and will be updated throughout the semester. New versions of the syllabus will always be announced in lecture, and students are responsible for ensuring they have the latest version. The most up to date version will always be on Blackboard.

In this course, we consider ways to illustrate compelling stories hidden in a blizzard of data. Equal parts art, programming, and statistical reasoning, data visualization is a critical tool for anyone doing analysis. In recent years, data analysis skills have become essential for those pursuing careers in policy advocacy and evaluation, business consulting and management, or academic research in the fields of education, health, medicine, and social science. This course introduces students to the powerful **R** programming language and the basics of creating data-analytic graphics in **R**. From there, we use real datasets to explore topics ranging from network data (like social interactions on Facebook or trade between counties) to geographical data (like county-level election returns in the US or the spatial distribution of insurgent attacks in Afghanistan). No prior background in statistics or programming is required or expected.

Contact Information

Name	James Lo
Office	VKC 348
Email	lojames@usc.edu
Office Hours	TBA and by appointment

Course Requirements

- **Class participation (10%):** Students should actively participate in all aspects of the course. Class participation will be judged based on questions asked/answered during the lectures.
- **Swirl Exercises (10%):** Each week, there is an in-class session where students will work on the assigned portion of the textbook and electronically complete a small set of answers to review questions using the *swirl* platform. The answers to review questions will be graded pass/fail in class. *This is an individual assessment with limited collaboration.*
- **Problem sets (60%):** Six problem sets will be posted during the course via Blackboard. Both hard copies of your problem sets and electronic submission of your computer code via Blackboard must be done. Each problem set will be equally weighted. *This is an individual assessment with no collaboration.*

- **Final project (20%):** This project is due on exam week. I will distribute one possible final project assignment on the last week of lecture. However, those of you with interest in a particular data set are especially encouraged to write a short paper on your analysis of that data, after getting approval from me. This is highly encouraged, and if this is the path you choose, I am open to this project being completed in small groups. Otherwise, this is also an individual assessment with no collaboration.

Submission via Blackboard Folders

For the answers to Problem Sets, students are required to turn in a pdf copy of their answers via assignment folders on Blackboard. In addition, students should submit a paper copy of their solutions. The answers should include the annotated computer code as part of your solutions to the questions. For each assignment, you will submit your code to the appropriate folder as a *single* file named `xxxPSetX.R` where `xxx` is your USC ID and `X` is the handout/problem set number. For example, it might be `lojamesPset3.R` for problem set 3.

Addendum to Plagiarism Policy

Violations of the above collaboration policy will be treated as instances of plagiarism. *Please take this guideline seriously.*

Programming necessitates that you reach your own understanding of the problem and discover a path to its solution. DO NOT, UNDER ANY CIRCUMSTANCES, COPY ANOTHER PERSON'S CODE. Incorporating someone else's code into your program in any form is a violation of academic regulations. Abetting plagiarism or unauthorized collaboration by sharing your code is also prohibited. Sharing code in digital form is an especially egregious violation: do not e-mail your code to anyone.

Novices often have the misconception that copying and mechanically transforming a program (by rearranging independent code, renaming variables, or similar operations) makes it something different. Actually, identifying plagiarized source code is easier than you might think. For example, there exists computer software that can detect plagiarism.

Textbook

This course uses a draft manuscript of the following textbook.

Imai, Kosuke (2015). *A First Course in Quantitative Social Science*. Under contract with Princeton University Press.

The textbook is made freely available to the students on Blackboard. Due to the copyright issues, this file should not be distributed to those who are not taking this class.

Statistical Software

In this course, we use the open-source statistical software **R** (<http://www.r-project.org>). **R** can be more powerful than other statistical software such as SPSS, STATA and SAS, but it can also be more difficult to learn. A variety of resources will be made available for POSC 499 students in order to learn **R** as efficiently as possible. To help make using **R** easier, we'll be using **RStudio** (<http://www.rstudio.com/>)—a user-interface that simplifies many common operations.

Course Outline

Week 1: Introduction

- Topic: Course overview, Introduction to R
- Textbook: Chapter 1 (Section 1.3)
- Swirl lab: Introduction 1 & 2

Week 2: Causality I

- Lecture: Randomized experiments (Sections 2.1-2.4)
- Swirl lab: Causality 1
- Precept: Causality in-class assignment TBD

Week 3: Causality II

- Lecture: Observational Studies (Sections 2.5-2.7)
- Swirl lab: Causality 2
- Precept: Causality in-class assignment TBD
- Problem set #1 distributed

Week 4: Measurement I

- Lecture: Survey Sampling (Sections 3.1-3.4)
- Swirl lab: Measurement 1
- Precept: Measurement in-class assignment TBD
- Problem set #1 due

Week 5: Measurement II

- Lecture: Clustering (Sections 3.5-3.7)
- Swirl lab: Measurement 2
- Precept: Measurement in-class assignment TBD
- Problem set #2 distributed

Week 6: Prediction and Loops I

- Lecture: Prediction and Loop (Section 4.1)
- Swirl lab: Prediction 1
- Precept: Loop in-class assignment TBD
- Problem set #2 due

Week 7: Prediction II

- Lecture: Regression I (Section 4.2)
- Swirl lab: Prediction 2
- Precept: Regression in-class assignment TBD

Week 8: Prediction III

- Lecture: Regression II (Section 4.3)
- Swirl lab: Prediction 3
- Precept: Regression in-class assignment TBD
- Problem set #3 distributed

Week 9: Probability I

- Lectures: Probability and Conditional Distributions (Section 6.1-6.3)
- Swirl lab: Probability I
- Problem set #3 due

Week 10: Probability II

- Lectures: Random variables and distributions (Section 6.4-6.5)
- Swirl lab: Probability II
- Problem set #4 distributed

Week 11: Uncertainty I

- Lectures: Estimation (Section 7.1)
- Swirl lab: Uncertainty I
- Problem set #4 due

Week 12: Uncertainty II

- Lectures: Hypothesis Testing (Section 7.2)
- Swirl lab: Uncertainty II
- Problem set #5 distributed

Week 13: Uncertainty III

- Lectures: Regression with Uncertainty (Section 7.3)
- Swirl lab: Uncertainty III
- Problem set #5 due

Week 14: Networks

- Lectures: Political Networks
- Problem set #6 distributed

Week 15: Data visualization

- Data visualization
- Problem set #6 due
- Final project distributed

Exam week

- Final project due

Statement on Academic Conduct and Support Systems

Plagiarism — presenting someone else's ideas as your own, either verbatim or recast in your own words - is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, Behavior Violating University Standards and Appropriate Sanctions, accessible here: <http://studentaffairs.usc.edu/scampus/>. Other forms of academic dishonesty are equally unacceptable. See the university policies on scientific misconduct: <http://policy.usc.edu/scientific-misconduct>.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity <http://equity.usc.edu/> or to the Department of Public Safety via either of these forms: <http://dps.usc.edu/contact/report/> or "<http://web-app.usc.edu/web/dps/silentWitness/>". The Center for Women and Men <http://engemannshc.usc.edu/cwm/> provides 24/7 confidential support, and the sexual assault resource center webpage <http://sarc.usc.edu/> describes reporting options and other resources.

Help with scholarly writing is provided by a number of USC's schools. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the American Language Institute <http://ali.usc.edu>, which sponsors courses and workshops specifically for international graduate students.

Help arranging accommodation for students with disabilities is provided by the Office of Disability Services and Programs <http://dsp.usc.edu>

Emergency information will be posted at <http://emergency.usc.edu>. If an officially declared emergency makes travel to campus infeasible, this website will provide safety and other updates, including ways in which instruction will be continued by means of Blackboard, teleconferencing, and other technology.