FALL 2018
IR 312 Introduction to Data Analysis
School of International Relations
University of Southern California
Schedule: Monday, Wednesday 3:30 to 4.50pm, VKC 152

Syllabus version: August 3, 2018

Instructor: Prof. Roberts email: tysonrob@usc.edu
Office Hours: Tuesday, Wednesday, Thursday, 11am-12noon Office: VKC 313

Course Description and Objectives:

Has income inequality increased across and within countries over the past few decades? Do foreign countries tend to agree more with the United States or Russia in the United Nations? Does winning political office enrich politicians, or are wealthy candidates merely more likely to win office? Does assassinating a country's leader change the probability of war? Do cash transfer programs improve poverty outcomes in developing countries? Do female legislators provide more public goods than male legislators?

Academic researchers and policy-makers increasingly rely on quantitative methods to answer these questions. As the sheer volume of data available grows, the ability to analyze data, interpret the results, and effectively communicate key findings has become an essential skill to conduct empirical research in the social sciences. The ability to extract valuable insights from quantitative data – often referred to as "data science" – is also a common demand by employers in the private sector.

This course aims at filling this gap by providing students with the fundamental statistical and computing skills necessary to apply data analysis methods to the study of International Relations. After the course, students will be able to:

1. Read and process data in multiple formats and conduct basic descriptive analysis
2. Create effective visual depictions of statistical patterns in data using visualization techniques
3. Formulate hypotheses about relationships between variables and test them with multivariate regression analysis in order to answer research questions
4. Measure uncertainty using confidence intervals and interpret these results building upon basic concepts in probability theory
5. Identify potential challenges to causal inference and design empirical strategies to address them

The course will follow a "learning-by-doing" approach and will place emphasis on gaining experience in analyzing data. Students are expected to do the required readings for each week, run the code from the textbook, and complete the *swirl* exercises corresponding to the reading.

The lectures will reinforce the readings and provide an opportunity to collaboratively work on assigned problem sets. Most of the applications will be related to International Relations questions, while others address other social sciences such as economics and sociology.

Students will learn how to use the open-source software R, a popular statistical programming environment, described by *The New York Times* as "the lingua franca" of data analysis in corporations and academia for its flexibility, scalability, and ease of use. R is freely available for download and runs on Macintosh, Windows, and Linux computers. It is more powerful than other statistical software, such as Excel, SPSS, and STATA, but it is a bit more difficult to learn. No previous programming experience is required, and a variety of resources will be made available to the students in order to learn R as efficiently as possible.

Additional R resources:

- Seminars, modules, etc. available at **https://stats.idre.ucla.edu/r/**
- Online manuals available at http://www.cran.r-project.org/other-docs.html
- Search answers to questions at https://stackoverflow.com/questions/tagged/r

Course Requirements and Grading:

- Class participation: 15%
- Problem sets: 60%
- Midterm exam: 10%
- Final exam: 15%

Class participation

Students are expected to attend every session and do the assigned readings before each session. Doing the assigned readings implies not only reading the required pages of the book (and required Application materials), but also running the code on their own (use the code available online or, if you have time, type in the code by hand), and completing the additional swirl exercises. Come to class with questions and ready to engage in a discussion about that week's topic. If students can demonstrate that they understand the reading and code assigned for the week, we will have time to begin working through the problem sets in class. The class participation grade will be based on attendance, demonstrating mastery of the assigned material, and contributing to collaborative work on the problem sets.

Problem sets

There will be a total of seven problem sets, each focusing on applying different statistical and programming concepts covered in the course to new datasets in R. Students are encouraged to collaborate but in the submission they must write up the code and answers on their own, and report the name of other students that they worked with in a footnote on the first page. Each problem set will be graded on a 0 to 10 scale. If all problem sets are submitted complete and on time, the lowest grade will be dropped. Two points per day will be deducted for late problem sets.

Midterm and final exams

The midterm and final exam will be in-class exams. Both tests will be open-book, and will assess how well students understand the key statistical and programming concepts in the class up to that point in the semester. They will include both short conceptual questions and longer questions where students will be asked to interpret the output of a statistical analysis.

Online Forum:

We will have a dedicated Discussion Board on the course Blackboard site for discussion and student questions. All course-related questions (especially those related to coding in R) should be asked here. This way all students will be able to benefit from seeing the questions and answers.

Required Readings:

Imai, Kosuke. (2016 or 2018). A First Course in Quantitative Social Science. Princeton University Press. (QSS in the course outline below). There are corrections to errors in the textbook at http://qss.princeton.press/

All other required readings are available online.

Optional Readings:

Monogan, Jamie (2015) Political Analysis Using R. Springer. (PAUR in the course outline below) An e-Book is available through the USC Library website.

COURSE OUTLINE AND READINGS

Week 1: August 20 & 22: Introduction

- Required: QSS Chapter 1.
  - (For Monday, read Chapter 1.0-1.3.0. For Wednesday, read 1.3.1-1.4.)
- Optional: PAUR Ch. 1
- Download R: https://cran.r-project.org/
- Download RStudio: https://www.rstudio.com/
- Download the QSS package and data, swirl exercises, Errata ,and the extdata Data Files. (Don't worry about tidyverse or the Federalist papers data.)
- Swirl exercises INTRO1 and INTRO2
- Application: Understanding world population dynamics

Weeks 2: August 27 & 29. Data Manipulation and Causality

- Required: QSS 2-2.3
- Optional: PAUR Ch. 2-2.4
- Application reading: https://www.poverty-action.org/study/discrimination-job-market-united-states
- Optional Application reading: Bertrand and Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *(Skim this and/or read the Alternative Application 1 material)* http://cos.gatech.edu/facultyres/Diversity_Studies/Bertrand_LakishaJamal.pdf
- ***Problem Set 1:*** QSS Exercise 1.5.2 (Due August 24)

Week 3: September 5. Randomized Controlled Trials

- Required: QSS 2.4
- Swirl exercises CAUSALITY1
- Application reading: Gerber, Green, and Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Largescale Field Experiment" https://isps.yale.edu/sites/default/files/publication/2012/12/ISPS08-001.pdf
- Alternative Application reading: https://www.povertyactionlab.org/evaluation/social-pressure-and-voter-turnout-united-states

Week 4: September 12. Causality and Observational Studies

- Required: QSS 2.5-2.7
- Optional: PAUR Ch. 4
- Swirl exercises CAUSALITY2
- Application reading: Jones and Olken. 2015. "Do assassins really change history?" https://www.nytimes.com/2015/04/12/opinion/sunday/do-assassins-really-change-history.html
- Optional Application materials:
    - Jones and Olken. 2009. "Hit or Miss? The Effect of Assassinations on Institutions and War." https://www.aeaweb.org/articles?id=10.1257/mac.1.2.55
    - Alternative application materials: Planet Money podcast. 2014. "Does raising the minimum wage kill jobs?" https://www.npr.org/sections/money/2014/03/06/286861541/does-raising-the-minimum-wage-kill-jobs
- Bonus reading: Quealy and Katz. 2018. "Nike Says Its $250 Running Shoes Will Make You Run Much Faster. What if That's Actually True?" https://www.nytimes.com/interactive/2018/07/18/upshot/nike-vaporfly-shoe-strava.html
- ***Problem Set 2:*** QSS Exercise 2.8.3. (Due September 14.)

Week 5: September 17. Measurement and Univariate Distribution

- Required: QSS 3-3.4
- Optional: PAUR Ch. 3-3.1
- Swirl exercises: MEASUREMENT1
- Application reading: Lyall. 2014. "How hard is it to win hearts and minds in Afghanistan? Very hard." *(Click through to read academic articles, if you like)*

https://www.washingtonpost.com/news/monkey-cage/wp/2014/01/06/measuring-hearts-and-minds-in-afghanistan/?utm_term=.28cde9d5e2ac

Week 6: September 24 & 26. Measurement and Bivariate Relationships

- Required: QSS 3.5-3.8
- Optional: PAUR Ch. 3.2
- Swirl exercises: MEASUREMENT2
- Application reading: "The U.S. is still lonely at the United Nations." https://www.washingtonpost.com/news/monkey-cage/wp/2013/09/25/the-u-s-is-still-lonely-at-the-united-nations/?utm_term=.8adcef0c59c1
- Optional Application reading: Bailey, Strezhnev, and Voeten. 2015. "Estimating Dynamic State Preferences from United Nations Voting Data." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2330913
- ***Problem set 3:*** QSS Exercise 3.9.3. (Due September 28.)

Week 7: October 1 & 3. Review and Midterm

Week 8: October 8 & 11. Prediction

- Required: QSS 4.1
- Swirl exercises: PREDICTION1
- Application reading: http://electoral-vote.com/evp2018/Info/welcome.html

Week 9: October 15 & 17. Regression

- Required: QSS 4.2
- Optional: PAUR 2.5 and Ch. 6
- Swirl exercises: PREDICTION2
- Application reading: Todorov, Mandisodza, Goren, and Hall. 2005. "Inferences of Competence from Faces Predict Election Outcomes." http://faculty.missouri.edu/segerti/capstone/Todorov3.pdf
- ***Problem Set 4***: Exercise 4.5.1, Questions 1-5. (Due October 19)

Week 10: October 22 & 24. Regression and Causation

- QSS 4.3
- Swirl exercises: PREDICTION3
- Application reading: Imai, King, and Velasco Rivera. 2018. "Do nonpartisan programmatic policies have partisan effects?" Abstract. https://gking.harvard.edu/publications/do-nonpartisan-programmatic-policies-have-partisan-electoral-effects-evidence-two
- Optional Application reading: De La O, Ana. 2013. "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico." *(Skim, or read Alternative reading)* https://leitner.yale.edu/sites/default/files/files/resources/PMF-papers/delao_progresa_finalb.pdf
- ***Problem Set 5***: Exercise 4.5.2 (Due October 26)

- Note: In Question 6, uses official turnout rate and PRI vote share (as in Question 5), and use average precinct population when plotting average effects.

Week 11: October 29 & 31. Probability & Bayes Rule

- QSS 6.1-6.2.3. (6.2.4 is *OPTIONAL*)
- http://www.rossmanchance.com/applets/MontyHall/Monty04.html
- Swirl exercises: PROBABILITY1

Week 12: November 5 & 7. Random Variables & Probability Distributions

- QSS 6.3-6.5.
- http://www.rossmanchance.com/applets/OneProp/OneProp.htm?candy=1
- https://fivethirtyeight.com/features/the-media-has-a-probability-problem/
- Swirl exercises: PROBABILITY2
- ***Problem Set 6***: Exercise 6.6.2 (Due November 9).

Week 13: November 12 & 14. Estimation Uncertainty

- QSS 7.1
- http://rocknpoll.graphics/
- https://fivethirtyeight.com/features/why-fivethirtyeight-gave-trump-a-better-chance-than-almost-anyone-else/
- Swirl exercises: UNCERTAINTY1

Week 14: November 19. Hypothesis Testing

- QSS 7.2
- John Bohannon. 2015. "I Fooled Millions Into Thinking Chocolate Helps Weight Loss. Here's How." https://io9.gizmodo.com/i-fooled-millions-into-thinking-chocolate-helps-weight-1707251800
- Kevin Drum. 2013. "What We Know—And What We Don't—About the Oregon Medicaid Study." https://www.motherjones.com/kevin-drum/2013/05/what-we-know-oregon-medicaid-study/
- Austin Frakt. 2013. "Power calculations for the Oregon Medicaid Study." https://theincidentaleconomist.com/wordpress/power-calculations-for-the-oregon-medicaid-study/
- Swirl exercises: UNCERTAINTY2

Week 15: November 26-28. Linear Regression with Uncertainty

- QSS 7.3
- PAUR 7-7.1. *(Don't worry about doing the R code for this chapter – the goal for this reading is to expose students to non-continuous dependent variables)*
- https://zeligproject.org/ (Skim – these are some more advanced R tools available to explore as you become a more sophisticated statistician.)
- Swirl exercises: UNCERTAINTY3
- ***Problem Set 7***: Exercise 7.5.3. Due Nov. 30.

Monday, December 10, 2pm: Final Exam