POLI 7450: Computational Methods for Comparative Research

Prof. Frederick Solt

Spring 2019

E-mail: frederick-solt@uiowa.edu
Office Hours: TTh 9:30 to 11:00 am and by appointment
Office: 324 Schaeffer Hall

Web: cmcr-class.github.io
Class Hours: T 11:00 am to 1:50 pm
Class Room: 103 Schaeffer Hall

Overview

In this course, you will learn to identify common methodological issues encountered in quantitative comparative research and to address them using computational techniques. Topics covered include data wrangling, exploration, and analysis (i.e., 'data science'); Bayesian estimation of latent variables such as democracy, judicial independence, and respect for human rights across countries and over time; and computational approaches to the study to the dynamics of public opinion both within and across countries. This course is cross-listed in Comparative Politics and Methodology; it may be used to meet requirements in either subfield (but not both).

Course Requirements

Your grade will be based on class participation (20%), assignments (10%), an exam (25%), a research presentation (20%), and a research paper (25%). The following grades may be assigned at the end of the class: A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F.

Class Participation

By this point in your career this should really go without saying, but here it is anyway: you can't learn in a class if you don't show up. You should attend every meeting having completed the readings and assignments, that is, you should be ready to discuss use cases, ask good questions, and share any difficulties you've encountered.

We will also be using the Issues tab of a dedicated GitHub repository to continue our discussions outside of class. To start a new thread, create a new issue. If you want to be sure to get the attention of someone in particular, use their GitHub handle—mine is `@fsolt`—to tag them in your post: GitHub sends an email to people when they are tagged.

Assignments

To practice your new skills, you will have regular assignments. Some will be completed through GitHub in a private repository for each student (besides you, only I will have access to your homework repo). Others will be DataCamp short courses. Others still will simply be readings that I want you to be sure to step through the embedded code. All will be graded on a pass (that is, substantially completed)/fail basis.

Exam

The exam will be available for your to take anytime after class on March 12 until the start of class on March 26. It will be open-note, open-interwebs—you know, just like IRL. Don't click the GitHub invitation link until you have three hours to spend completing the exam.

Research Paper

The research paper for this course is more straightforward than the usual: the assignment is "simply" to write a fully reproducible version of a quantitative research paper you prepared for another class. Ideally the paper will be for a class you are taking this semester, so that you use best practices as you go; you will need to confirm with the instructor of that course that this is acceptable to her or him. Rewriting a promising paper from past coursework will also be fine, though. If you're stuck for an appropriate paper, talk to me and we will work it out.

For your paper, you will:

- write using RMarkdown (or `knitr`) with interspersed R code and text,
- begin with raw data collection,
- include all steps of data cleaning and analysis,
- present figures rather than tables,
- use BibTEXfor citations,
- automatically generate PDF, html, and ($^{ee}e_{w_{ww_w}}$!) Word versions,
- do your work in a GitHub repository,
- submit a link to that repo.

Papers are due the last day of exam week, Friday, May 10, before midnight.

Research Presentation

In our last class meeting, you will give a presentation based on your research paper to the class as if at an academic conference. You will speak for 12 to 15 minutes. You will have slides; these slides will not be filled with blocks of text. You will use plots, not tables, to present results. You will rock this.

Course Schedule

Week 1, January 15: Introduction to the Course

Assignment:

1. Bryan, Jennifer. 2016. *Happy Git and GitHub for the UseR*. Chapters 1-16. Read this carefully and follow its instructions to get set up with R, RStudio, Git, and GitHub on your laptop before we meet.

Week 2, January 22: The Replication Crisis and Reproducibility

Readings:

1. Aschwanden, Christie, and Maggie Koerth-Baker. 2016. "How Two Grad Students Uncovered an Apparent Fraud—and a Way to Change Opinions on Transgender Rights." *FiveThirtyEight*, April 7, and Dimitrova, Velichka. 2013. "Reinhart-Rogoff Revisited: Coding Errors Happen—Key Problem Was in Not Making the Data Openly Available from the Start." *LSE: The Impact Blog*, April 24.
2. Data Access and Research Transparency (DA-RT): A Joint Statement by Political Science Journal Editors.
3. *AJPS* Replication and Verification Policy and *American Journal of Political Science* Guidelines for Preparing Replication Files.
4. Gelman, Andrew, and Eric Loken. 2013. "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No 'Fishing Expedition' or '*p*-Hacking' and the Research Hypothesis Was Posited Ahead of Time."
5. Leek, Jeffrey T., and Roger D. Peng. 2015. "Opinion: Reproducibile Research Can Still Be Wrong: Adopting a Prevention Approach." *Proceedings of the National Academy of Sciences* 112(6):1645-1646 and Patil, Prasad, Roger D. Peng, and Jeffrey T. Leek. 2016. "A Statistical Definition for Reproducibility and Replicability." *bioRxiv*, July 29.
6. Solt, Frederick, Yue Hu, Kevan Hudson, Jungmin Song, and Dong 'Erico' Yu. 2017. "Economic Inequality and Class Consciousness." *Journal of Politics* 79(3):1079-1083.

Assignments:

1. Bryan, Jennifer. 2016. *Happy Git and GitHub for the UseR*. Chapter 18.

Week 3, January 29: Settling In with R

Readings:

1. Ismay, Chester, and Albert Y. Kim. 2018. *Modern Dive: An Introduction to Statistical and Data Sciences via R*, Chapters 1 and 2.
2. Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*, Preface and Chapter 2. Some overlap with Ismay and Kim, but that will probably be useful to you at this point.
3. Solt, Frederick. 2018. "How to Switch Your Workflow from Stata to R, One Bit at a Time."

Assignments:

1. Use the interactive `swirl` package ("Learn R in R") to learn (or review) some R basics. First, run the following code to get to the `swirl` interface:

```
install.packages("swirl")
library("swirl")
```

   Then follow the prompts to the course `1: R Programming: The basics of programming in R` and complete lessons 1 through 9. (When it asks whether you want Coursera credit, just say no.)

Week 4, February 5: Visualization

Readings:

1. Kastellec, Jonathan P. and Eduardo L. Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics* 5(4):755–771.
2. Ismay, Chester, and Albert Y. Kim. 2018. *Modern Dive: An Introduction to Statistical and Data Sciences via R*, Chapter 3.
3. Solt, Frederick, and Yue Hu. 2015. "`dotwhisker`: Dot-and-Whisker Plots of Regression Results." Available at The Comprehensive R Archive Network (CRAN)., and especially how to remake the examples from Kastellec and Leoni (2007) using `dotwhisker`.

Assignments:

1. Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*, Chapters 1 and 3.

Recommended:

1. Chang, Winston. 2013. *Cookbook for R*, Chapter 8: Graphs. Skim a couple sections to get a feel for this as a reference. I myself refer to it all. the. time.
2. Wilke, Claus O. 2019. "Visualizing Uncertainty," *Fundamentals of Data Visualization*

Week 5, February 12: Wranglin' Data

Readings:

1. Leeper, Thomas J. 2016. "`rio`: A Swiss-Army Knife for Data I/O. Import, Export, and Convert Data Files."
2. Ismay, Chester, and Albert Y. Kim. 2018. *Modern Dive: An Introduction to Statistical and Data Sciences via R*, Chapters 4 and 5.
3. Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*, Chapters 4 and 5.

Assignment:

1. Robinson, David. Datacamp Course: Introduction to the Tidyverse.

Recommended:

1. RStudio's Data Transformation Cheat Sheet. I had the old version open on a dedicated desktop on my computer for at least a year—maybe two!—before I finally stopped looking at it like twenty times a day. Lots of people print it out; I've even heard of people laminating it.

Week 6, February 19: Wranglin' Data II, Working with Text

Assignment:

1. Wickham, Charlotte. Datacamp Course: String Manipulation in R with stringr.

Recommended:

1. RStudio's Work with Strings Cheat Sheet.
2. Silge, Julia, and David Robinson. 2018. *Text Mining with R: A Tidy Approach*.

Week 7, February 26: Scraping (and Managing) Data

Readings:

1. Bryan, Jennifer. "Organization I: File Naming."
2. Janz, Nicole. 2015. "Good Practice in Data Collection and Storing." *Political Science Replication*, March 16.
3. Robinson, David. 2016. "Text Analysis of Trump's Tweets Confirms He Writes Only the (Angrier) Android Half." *Variance Explained*, August 9.
4. Persson, Eric. 2017. `gesis`: R Client for GESIS Data Catalogue (DBK). R package version 0.2.1, Solt, Frederick and Kellen Gracey. 2016. `icpsrdata`: Reproducible Data Retrieval from the ICPSR Archive. R package version 0.3.0., and Frederick Solt (2018). `ropercenter`: Reproducible Data Retrieval from the Roper Center Data Archive. R package version 0.2.0.
5. RStudio. 2014. "`rvest`: Easy Web Scraping with R." *RStudio Blog*, November 24.

Week 8, March 5: Multilevel Models of Multiple Surveys I

Readings:

1. Solt, Frederick. 2008. "Economic Inequality and Democratic Political Engagement." *American Journal of Political Science* 52(1):48-60.
2. Solt, Frederick, Yue Hu, Kevan Hudson, Jungmin Song, and Dong 'Erico' Yu. 2017. "Economic Inequality and Class Consciousness." *Journal of Politics* 79(3):1079-1083. Yes, *again*, but this time, work through the replication code.

Week 9, March 12: Multilevel Models of Multiple Surveys II

Readings:

1. Gabry, Jonah, and Ben Goodrich. 2018. "How to Use the `rstanarm` Package."

Week 10, March 19: Spring Break

Week 11, March 26: Crowdsourcing Cross-National Data

Readings:

1. Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. "Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* 110(2):278-295.

Assignment:

1. Take-home exam: open book, open notes, open internet. Don't click the GitHub invitation link until you have three hours to spend completing the exam.

Week 12, April 2: Estimating Cross-National Latent Variables

Readings:

1. Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426-449.
2. Fariss, Christopher J. 2014. "Respect for Human Rights Has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108(2):297-318.
3. Linzer, Drew A., and Jeffrey K. Staton. 2015. "A Global Measure of Judicial Independence, 1948-2012." *Journal of Law and Courts* 3(2):223-256.
4. Solt, Frederick. 2019. "Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database".

Week 13, April 9: Analysis with Cross-National Latent Variables

Readings:

1. Crabtree, Charles D., and Fariss, Christopher J. 2015. "Uncovering Patterns Among Latent Variables: Human Rights and De Facto Judicial Independence." *Research & Politics* 2(3):1-9.

Assignment:

1. Pemstein, Meserve, and Melton provide an example of how use Stata to employ their Unified Democracy Scores in research. Replicate their example but show instead how to use the UDS in R. You *might* find Crabtree and Fariss's reproducibility materials to be a bit helpful, but you should of course use the `tidyverse` methods we've learned this semester.

Week 14, April 16: Public Opinion in One Country as a Latent Variable

Readings:

1. McGann, Anthony J. 2014. "Estimating the Political Center from Aggregate Data: An Item Response Theory Alternative to the Stimson Dyad Ratios Algorithm." *Political Analysis* 22(1):115-129.

Week 15, April 23: Public Opinion as a Cross-National Latent Variable

Readings:

1. TBA

Week 16, April 30: Presentations

Readings:

1. Huber, Gregory. "How to Present Your Work at Conferences."

Recommended:

1. Secklington, Melinda. 2018. "The Art of Talk Design."
2. Secklington, Melinda. 2017. "The Art of Slide Design."

Further Course Information

Administrative Home

The College of Liberal Arts and Sciences (CLAS) is the administrative home of this course and governs its add/drop deadlines, the second-grade-only option, and other policies. These policies vary by college. See the CLAS *Student Academic Handbook*.

Electronic Communication

Students are responsible for official correspondences sent to their UI email address (uiowa.edu) and must use this address for all communication within UI (*Operations Manual*, III.15.2. Scroll down to k.11.)

Accommodations for Disabilities

UI is committed to an educational experience that is accessible to all students. A student may request academic accommodations for a disability (such as mental health, attention, learning, vision, and physical or health-related condition) by registering with Student Disability Services (SDS). The student should then discuss accommodations with the course instructor. See https://sds.studentlife.uiowa.edu/ for more information.

Nondiscrimination in the Classroom

UI is committed to making the classroom a respectful and inclusive space for all people irrespective of their gender, sexual, racial, religious or other identities. Toward this goal, students are invited to optionally share their preferred names and pronouns with their instructors and classmates. The University of Iowa prohibits discrimination and harassment against individuals on the basis of race, class, gender, sexual orientation, national origin, and other identity categories set forth in the University's Human Rights policy. For more information, contact the Office of Equal Opportunity and Diversity. See <diversity.uiowa.edu> for more information.

Academic Honesty

All CLAS students or students taking classes offered by CLAS have, in essence, agreed to the College's Code of Academic Honesty. Misconduct is reported to the College, resulting in suspension or other sanctions, with sanctions communicated with the student through the UI email address See https://clas.uiowa.edu/students/handbook/academic-fraud-honor-code for more information.

CLAS Final Examination Policies

The final exam schedule for each semester is announced around the fifth week of classes; students are responsible for knowing the date, time, and place of a final exam. Students should not make travel plans until knowing this final exam information. No exams of any kind are allowed the week before finals. See https://clas.uiowa.edu/faculty/teaching-policies-resources-examination-policies for more information.

Making a Complaint

Students with a suggestion or complaint should first visit the instructor, Professor Solt, and then the departmental DEO, Dr. Brian Lai (Office: 343 SH). Students may then bring the concern to CLAS See https://clas.uiowa.edu/students/handbook/student-rights-responsibilities for more information.

Understanding Sexual Harassment

Sexual harassment subverts the mission of the University and threatens the well-being of students, faculty, and staff. All members of the UI community must uphold the UI mission and contribute to a safe environment that enhances learning. Incidents of sexual harassment must be reported immediately. For assistance, definitions, and the full University policy, see https://osmrc.uiowa.edu/.

Reacting Safely to Severe Weather

In severe weather, class members should seek appropriate shelter immediately, leaving the classroom if necessary. The class will continue if possible when the event is over. For more information on Hawk Alert and the siren warning system, visit the Department of Public Safety website.

Student Resources

I encourage interested students to make use of the Writing Center and the Speaking Center.