

DATA IN POLITICS I: AN INTRODUCTION

This draft: October 13, 2020 (Check Sakai for the latest)

Poli281, Fall 2020

Meeting time: Tuesdays and Thursdays, 1:15-2:30 (Section 2) or 3:00 to 4:15 (Section 3)

Professor Timothy Ryan; tjr@email.unc.edu; 919-962-0403.

Office hours: Wednesdays 1-4pm, via Zoom. (See Sakai for the appropriate link.) Sign up for an appointment [here](#), or email me for an appointment outside of the default window.

Office: 307 Hamilton Hall (aka Pauli Murray Hall) but I won't be there during the pandemic.

Classroom: Virtual. See Sakai for the link.

Final Exam: Nov. 18 at Noon (Section 2) or Nov. 19 at noon (Section 3)

The Information Revolution has dramatically expanded the volume of information we have about the world around us. Social scientific analytical skills are transforming many sectors—business, journalism, law, public policy, health care, and finance, to name but a few—and are more valued now than ever. The broad learning objective for this course is to help students develop the tools they need to be informed participants and active leaders in data-driven sectors. More specifically, the learning objectives are:

1. To increase students' comfort and facility managing data in the R statistical language, with an emphasis on versatile tools such as loops, sampling functions, merging datasets, and the GGPlot data visualization software;
2. To teach basic principles of data description, including standard descriptive plots and statistics; and
3. To develop students' ability to use data to answer important social scientific questions.
4. To learn how to acquire new quantitative skills autonomously, such as by reading software documentation.

Students will leave this class with the competencies they need to conduct basic analysis on many different forms of data, as well as the foundation they need to acquire more advanced skills (such as characterizing uncertainty in data and testing formal hypotheses).

The target audience for this course is undergraduate students with interest in the social sciences (not only Political Science), who want to use quantitative approaches to solve important problems and develop marketable analytical skills. This course is a prerequisite for *Poli381: Data in Politics II: Frontiers and Applications*.

This course fulfills the Quantitative Intensive (QI) requirement in UNC's Making Connections curriculum. It also fulfills counts as a course in research methods (required for completing the Political Science Bachelor's degree).

Requirements and Evaluation

Course assessment is broken down as follows.

ASSIGNMENTS (60%) There will be five assignments due as noted in the schedule below. They are weighted equally. Assignments are due at 6pm on the days indicated, unless we decide something different in class.

FINAL PROJECT (25%, broken up as follows) The class has a capstone final project for which students, working in groups, conduct and present an original data analysis on an existing dataset. This project consists of a paper (20%) and a presentation (5%). A separate document specifies final project requirements more completely.

As a default, all final project group members receive the same grade. However, different grades can be assigned when the work was not shared equitably. Additionally, if one group member is significantly disrupting progress (such as by not providing timely responses to communication or by missing scheduled meetings), this person can be removed from the group, and required to complete an alternative final project alone.

PARTICIPATION (15%) Your participation grade has two components. The first component is preparation for (and actual attendance in) class. At each class meeting, you will give yourself 0 to 3 points, depending on how prepared you are for class and how much you intend to participate. (0 = absent from class; 1 = attended class but did not prepare; 2 = attended class but preparation was partial or rushed; 3 = attended class, prepared with care, *and commit to being proactive in asking questions and contributing to discussion.*) The semester-long score generated by these reports is advisory to the instructor-assigned participation grade—I can adjust self-reports that are inconsistent with what I see in class—but I take them seriously. Also, I can administer occasional pop-quizzes to gauge attention to the reading. These quizzes do not figure into your final grade directly. Rather, they help me calibrate my interpretation of your self-assessment above.

The second component is participation in our class's online Piazza discussion forum. (See Sakai for a link.) On Piazza, you can benefit your participation grade **either** by posting your own questions or by providing thoughtful answers to other students' questions. (The website keeps track of your activity.) These statistics, too, are advisory to the summary participation grade.

Because attendance is a course requirement (see the Undergraduate Bulletin) and critical to having a lively conversation, more than three unexcused absences can have a negative effect on your grade—potentially a drastic one—outside the scheme above. (That is, it can affect more than just your 10% participation grade.) **If you have missed more than three classes, it is your responsibility to come and talk to me about what's going on.**

I wish to emphasize that the attendance policy above applies even though this class is being taught remotely, and even though I intend to make most session recordings available after the fact. This is a highly interactive class, and synchronous participation is simply core to this class's structure.

Note also that, the weighting scheme above notwithstanding, all assignments must be submitted. If you have not submitted an assignment, you are at risk of receiving a course grade of I (Incomplete) until it is complete.

Course Texts

This course has one required textbook:

1. Imai, Kosuke. 2017. *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press.

In addition, there are software requirements. Students must download and install R, a free statistical program available at <http://cran.r-project.org/>, as well as RStudio (also free), which is available at www.rstudio.com. They also must register for Data Camp (www.datacamp.com), a resource that provides tutorial videos and interactive training modules to help learn programming skills. Data Camp is free for students enrolled in this class. (I will provide you with login information.) Please do not pay for Data Camp access!

Other Policies

Technology in the classroom. This class is being taught remotely, so obviously you can use technology in class. However, I beg, beseech, exhort, and (most important) enjoin you to use technology only for purposes directly related to class. It hugely impoverishes the synchronous experience to have everyone's attention wandering about to other things. This is why I always turn off my email program (and all other distractions) during class. Please be 100% mentally "present" with us.

I'm going to write that paragraph again as a signal of how strongly I feel about it.

This class is being taught remotely, so obviously you can use technology in class. However, I beg, beseech, exhort, and (most important) enjoin you to use technology only for purposes directly related to class. It hugely impoverishes the synchronous experience to have everyone's attention wandering about to other things. This is why I always turn off my email program (and all other distractions) during class. Please be 100% mentally "present" with us.

Email. I usually respond to emails within 24 hours (weekends excepted). However, please limit your use of email to issues that are private, or at least specific to you. For matters that are not private and where other students might want to see the response, please use Piazza (see above). Note that Piazza permits anonymous posting. (The post will be anonymous to other students, but not to me.)

I typically will not use email to repeat information that was missed because of an absence; I'll direct you to correspond with a classmate or the class recording.

Cooperation and academic integrity. In a class setting, cooperative work has both benefits and pitfalls. Peers learn a lot by explaining things to each other. But it can also be easy to stumble into a passive mindset where you're not really *assimilating* the concepts. To strike a balance, I will designate some activities and assignments (or parts thereof) as being Cooperative, and others as Individual. It is critical that you attend to this distinction, as completing individual work cooperatively would be a breach of academic integrity.

By its nature, this class has an extra matter we need to address. While discussion with other people is permitted and encouraged for work designated as cooperative, there is a distinction between discussing a problem and copying someone else's work. (Writing computer code is an especially tempting activity for which to copy work.) Students can discuss problem-solving strategies, clarify concepts, and point out mistakes—but ultimately each person must generate his or her own path to the solution. *In our class, copying and pasting another person's computer code is potentially tantamount to plagiarism. **Even for work designated as cooperative, you must write your code individually.*** Unless I have given you explicit permission for some special reason, do not do it. Copied code is surprisingly easy to detect (there is

software designed to detect it). Be assured that if I identify a case of cheating or plagiarism, I will handle it 100% “by the book.”

Students with disabilities. If you think you need an accommodation for a disability, please let me know. Some aspects of the course and its assignments may be modified to facilitate your success. I will work with the Office of Accessibility Resources and Services to determine appropriate accommodations. I will treat any information you provide as confidential. Barring unusual circumstances, I require notice of a need for accommodation within the *first two weeks* of the semester.

Grade grievances. Requests for regrades have a time window. They cannot be submitted until at least 48 hours have passed since the assignment was returned (a cool-down period), and then they will only be accepted within three weeks of an assignment being returned (a statute of limitations). To request a regrade, you must submit a written memo (two pages max) explaining what aspect of your original grade you think was in error.

Absences. Requests for an excused absence should come via email (for clear documentation). The request must come as far in advance of the absence as possible if the absence is foreseen, or as soon after as possible if it is not foreseen. Per university policy, only your academic advisor can provide an official final exam excused absence.

This course includes several class periods focused on workshopping final projects. Absences on these days are especially disruptive, since they interfere with your group’s ability to work on the final project together. For this reason, absences on these dates (noted below) will be excused only for reasons of religious observance, illness, or family emergencies. Unexcused absences on these days cause your individual final project grade to receive an automatic 5-point penalty.

Schedule changes: I occasionally modify the schedule to accommodate lesson plans that took more or less time than expected. Assignment due dates also occasionally change, though they are only ever postponed—not advanced.

Course Schedule

August 11: Course Introduction

- Lohr, Steve. 2009. “[For Today’s Graduate, Just One Word: Statistics.](#)”
- Vance, Ashlee. 2009. “[Data Analysts Captivated by R’s Power.](#)”
- Rogers Cook, Lindsey. 2019. “[How 5 Data Dynamos Do Their Jobs.](#)”
- Register for Poll Everywhere and Zoom (via UNC SSO). (Go [here](#) and enter unc.zoom.us as the domain name.) Also, ensure you are using the Zoom ap (not the browser interface).

August 13: The Possibilities and Perils of Data Science

- Register for Data Camp and Piazza.
- Complete the “Getting to Know You” questionnaire and submit it under the “Assignments” tab on Sakai.

Part I: Filling Your Toolkit

August 18: The R Statistical Software: Getting Set Up, Part 1

- Download and install both R and RStudio.

August 20: The R Statistical Software: Getting Set Up, Part 2

- *QSS* Chapter 1. Read this knowing that much here will not make a lot of sense yet. As Imai says, one can learn data analysis only by *doing*, not by reading. In that vein, much of the very same material will be covered interactively in the coming weeks. But, you should be familiar with what is here, so you know what resources and examples you can easily refer back to when questions arise, which they will.

August 25: No Class—cancelled by UNC administration for pandemic reasons.

August 27: The R Statistical Software: Basic Tools

- DataCamp: Introduction to R, Chapters 1-3

September 1: The R Statistical Software: Basic Tools

- DataCamp: Introduction to R, Chapters 4-6, but focus especially hard on ch. 4 and 5. Ch. 6 covers lists, which are useful and you should know what they are. But it is far more urgent to master the content in ch. 4 and 5.

September 3: The R Statistical Software: Intermediate Tools 1

- DataCamp: Intermediate R, Chapter 1 (Conditionals)

September 8: The R Statistical Software: Intermediate Tools 2

- DataCamp: Intermediate R, Chapter 2 (Loops)
- Assignment 1 due

September 10: Political Application #1

- Klar: Partisanship in a Social Setting. (On Sakai)

Part II: Data Analysis as Problem Solving

September 15: Getting to Know Your Data

- Read *QSS*, ch. 1 again.

September 17: Causality

- Bertrand & Mullainathan, pp. 991-997
- *QSS*, pp. 32-48
- Assignment 2 due

September 22: Causality

- Green et al., pp. 33-38
- *QSS*, pp. 48-54

September 24: Causality

- Card & Krueger, pp. 772-778
- *QSS*, pp. 54-69

September 29: Causality

- Mosteller, entire

October 1: Workshop on Causality tools

- Read the article assigned for our in-class activity

October 6: Data Visualization #1

- DataCamp: Introduction to the Tidyverse, Ch. 1-4

October 8: Data Visualization #2

- Assignment 3 due

October 13: Data Visualization #3

October 15: Measurement #1

- *QSS*, pp. 75-111. But note: much of this is covering alternative ways to do data visualization (using R's built-in tools, rather than Tidyverse tools). You can read that part lightly, since you are primarily responsible for learning the Tidyverse tools. BUT you should make sure to assimilate the following important concepts:
 - o `Prop.table()` on p. 77.
 - o `na.omit()` on p. 79.
 - o Survey sampling (section 3.4).
 - o z-scores and correlation coefficients (section 3.5.2)
 - o The `apply()` function (p. 109).

October 20: Prediction #1

- *QSS*, pp. 123-139

October 22: Prediction #2

- *QSS*, pp. 139-148
- Assignment 4 due

October 27: Prediction #3

- *QSS*, pp. 148-161

October 29: Prediction #4

- *QSS*, pp. 161-170
- Assignment 5 due

Part III: Putting Data to Work

November 3: Flex Day—agenda TBD.

November 5: In-class final project workshopping

November 10: In-class final project workshopping

November 12: In-class final project workshopping

November 17: Review and conclusions

November 18 at Noon: Section 2 Final Presentations (a “nontraditional format” final exam)

November 19 at Noon: Section 3 Final Presentations (a “nontraditional format” final exam)