

DATA IN POLITICS I: AN INTRODUCTION

This draft: August 21, 2024 (Check Canvas for the latest)

POLI281, Fall 2024

Meeting time: Mondays and Wednesdays, 3:35pm-4:50pm

Professor Ye Wang; yewang@unc.edu.

Office hours: Tuesdays 12:30-3:30pm. Sign up for an appointment on Canvas before 9am on that day or email me for an appointment outside of the default window.

Office: 322 Hamilton Hall (aka Pauli Murray Hall).

Classroom: Woollen Gym, Room 0303.

Final project presentation: Dec 7, 4-7pm, in class.

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. However, data can also be used to mislead, and without strong data skills it's hard to be an informed data consumer. The broad learning objective for this course is to help you develop the tools you need to be informed participants and active leaders in data-driven sectors. More specifically, the learning objectives are:

1. To increase your 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 method.
2. To teach basic principles of data description, including standard descriptive plots and statistics.
3. To develop your ability to use data to answer important social science questions.

You 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). It also counts towards the Data Science minor.

Requirements and Evaluation

Course assessment is broken down as follows.

PARTICIPATION (20%) Your participation grade has two components.

1. **Attending and participating in class.** We will have many activities throughout the semester, some individual and some in groups. Full participation in class means being active during lecture, discussion, and group work.

2. **Participation in our class's online Piazza discussion forum.** (See Canvas 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 are advisory to the summary participation grade.

You can miss up to six classes without impacting your participation score. However, if you miss more than three classes, you need to submit a two-page essay for each extra class you miss (up to three classes). To save your time and mine, **please do not email me about your excuses of absence.**

ASSIGNMENTS (50%) There will be five assignments due as noted in the schedule below. They are weighted equally. Assignments are due at 11:59pm on the days indicated. A late penalty will be automatically imposed by Canvas if you miss the deadline. Submission a week later than the deadline will not be accepted.

FINAL PROJECT (30%, broken up as follows) The class has a capstone final project for which students will perform and present original data analysis on an existing dataset. This project consists of a paper (25%) and a presentation (5%).

Note 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 IN (Incomplete) until it is complete. **Please double check your documents before submission.**

Your final grade will be determined by the [grading scheme](#) on Canvas.

Course Texts

This course has no textbook. However, 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.

Class Policies

Technology in the classroom. Due to the nature of the course, you will typically be using your computers in each class. However, If I see you using your computer for extraneous purposes—e.g., chat, social media, ESPN, or checking the news— this will negatively impact your participation grade. This policy is motivated by [evidence](#) that extraneous technology use decreases learning not only for the user, but for the people around them.

Email. I usually respond to emails within 2 business days. 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. **Please send a follow-up email if you have not heard from me in 72 hours (not including weekends.)**

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 or written analysis is plagiarism. This is true if you copy is from any source, including a classmate or former student, an online source, or an interface like ChatGPT. Even for work designated as cooperative, you must write your code and analysis individually.** Unless I have given you explicit permission for some special reason, do not do it. Copied code in particular 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 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.

Absence on critical days. Generally, absences on exam days will only be excused for reasons of religious observance, illness, or family emergencies. All requests for an excused exam absence must be submitted and approved **by email** (for 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.

Schedule changes: I occasionally modify the schedule to accommodate lesson plans that took more or less time than expected. In particular, assignment due dates occasionally change.

We all hope this will be a normal semester. However, we all need to be flexible about technology challenges and other unanticipated crises. This means two things. First, pieces of the course may be adjusted as needed if it will help us all complete the semester successfully. Second, please let me know if you face any special challenges this semester. I'm generally happy to work with students who need extra help or accommodations.

Learning Outcomes

Writing. Students are required to submit a final paper with at least 10 pages in length.

Presenting. Students need to give a 5-minute oral presentation on their final project at the end of the semester.

Collaborating. Students will work in groups on in-class exercises through the semester.

Course Schedule

Week 1

Monday, 8/19: Course Introduction

Part I: Using Data to Describe the World

Wednesday, 8/21: The R Statistical Software: Getting Set Up

- Register for Piazza.
- Download and install both R and RStudio
- After class: review LearnR Tutorial 1: Intro to R

Week 2

Monday, 8/26: Data Structures Part 1: Vectors

- LearnR Tutorial 2, Topics 1-3: Vectors

Wednesday, 8/28: Data Structures Part 2: Data Types & Operations

- LearnR Tutorial 2, Topics 4-5: Vectors

Week 3

Monday, 9/2: Labor Day; No class

Wednesday, 9/4: Data Structures Part 3: Data frames and Matrices

- LearnR Tutorial 3, Topics 1-3: Introduction to Dataframes

Week 4

Monday, 9/9: Summarizing Data, Part 1

- LearnR Tutorial 3, Topic 4: Importing data

Wednesday, 9/11: Summarizing Data, Part 2

- LearnR Tutorial 4, Topics 1-3: Basics of data cleaning

Week 5

Monday, 9/16: Summarizing Data, Part 3

- LearnR Tutorial 4, Topics 4-6: Creating new variables

Wednesday, 9/18: Putting Data in Context: Sampling

- No reading

Friday, 9/20: Assignment 1 due

Week 6

Monday, 9/23: Well-being day, no class

Wednesday, 9/25: Advanced Data Tools: The Tidyverse

- LearnR Tutorial 5, all topics: Introduction to the Tidyverse

Week 7

Monday, 9/30: Data Visualization Part 1: Types of Visualizations

- LearnR Tutorial 6, Topics 1-3: Types of visualizations and variables

Wednesday, 10/2: Data Visualization Part 2: Making Graphs and Figures

- LearnR Tutorial 6, Topics 4-5: Introduction to ggplot

Friday, 10/4: Assignment 2 due

Part II: Data Analysis as Problem Solving

Week 8

Monday, 10/7: Data Visualization Part 3: Advanced Visualization

- LearnR Tutorial 6, Topic 6: Advanced ggplot commands

Wednesday, 10/9: Data Visualization Part 4: Practice

- LearnR Tutorial, Week 7: Thinking through visualization

Week 9

Monday, 10/14: Functions & Loops

- Watch the following video: <https://www.youtube.com/watch?v=p8tAQx7ijXE>

Wednesday, 10/16: Invited talk, no class

Friday, 10/18: Assignment 3 due

Week 10

Monday, 10/21: Standard deviations, Z scores, and correlations

- Reading TBA

Wednesday, 10/23: Causality, Part 1

- Reading TBA

Week 11

Monday, 10/28: Causality, Part 2: Randomization

- Reading TBA

Wednesday, 10/30: Causality, Part 3: Observational Studies I

- Reading TBA

Week 12

Monday, 11/4: Causality, Part 4: Observational Studies II

- Read the article assigned for our in-class activity

Wednesday, 11/6: Prediction, Part 1

- Reading TBA

Friday, 11/8: Assignment 4 due

Week 13

Monday, 11/11: Prediction, Part 2

- Reading TBA

Wednesday, 11/13: Prediction, Part 3

- Reading TBA

Week 14

Monday, 11/18: Prediction, Part 4

- Reading TBA

Wednesday, 11/20: Prediction, Part 5

- Reading TBA

Part III: Putting Data to Work

Week 15

Monday, 11/25: In-class final project workshopping

Tuesday, 11/26: Assignment 5 due

Wednesday, 11/27: Thanksgiving; No Class

Week 16

Monday, 12/2: In-class final project workshopping

Wednesday, 12/4: In-class final project workshopping

Saturday, 12/7, 4pm to 7pm: Final Exam (project presentation)