Quantitative Political Science I: Computing

PSCI 2300, Vanderbilt University, Fall 2025

Professor Brenton Kenkel

How can we use polls to predict election outcomes? What's the relationship between campaign spending and electoral success? Which countries would we predict are most likely to experience civil wars?

PSCI 2300 will teach you how to address these and other social science questions through data analysis. The course will introduce the basic principles of statistical analysis and the baseline programming skills you need to analyze data. The goal is to give you the foundational tools you need to analyze data in your own research, as well as to be an informed and critical consumer of statistical claims made in the news media, policy reports, and academic research.

Here is a checklist to consider when deciding whether to take this class:

- ☐ I am a political science major or someone interested in quantitative social science.
- ☐ I want to learn about statistical concepts, as well as the programming skills needed to manipulate and analyze data in R.
- ☐ I am willing to spend time outside of class each week to keep up with the material.
- ☐ I would like to use data analysis in a paper (e.g., a political science honors thesis) or a job in the future.
- ☐ I understand that this is an introductory course, so we will focus mainly on developing a conceptual understanding. We will not dive too deep into mathematical foundations or advanced computational methods.

General information

Place and time. We meet on Tuesdays and Thursdays from 2:45–4:00pm in Cohen Hall 203.

Contact. You can email me at brenton.kenkel@gmail.com. I try to respond to all student questions within one business day.

Office hours and meetings. My office hours are Thursdays from 1:00–2:30pm in Commons Center 326. You don't need to make an appointment — just show up during my office hours. If you have a question but can't meet during my office hours, email me first. From there we can decide if we can work out the issue over email or if we need to meet.

Teaching assistant. The TA for PSCI 2300 is Mason Auten. You can email him at mason.auten@vanderbilt.edu. His office hours are Tuesdays from 9:00–10:30am in Commons Center 317.

Logistics

Lectures. The typical class will consist of me introducing concepts and code motivated by real-world political science questions, then writing code and running the analysis in real time while you follow along on your own computer. This means you should bring your laptop every day — a real laptop, not an iPad or other tablet that likely cannot run RStudio.

Readings. The main source for readings will be the lecture notes posted at https://bkenkel.com/qps1. Depending on the week's content, I may assign some academic articles or news sources as supplemental readings. Any such material will be posted to Brightspace.

Software. All the software we use is free and open source.

- R programming language: https://www.r-project.org
- RStudio development environment for R: https://posit.co/download/rstudio-desktop
- Quarto scientific publishing system: https://quarto.org

Grading

Problem sets (50%). There will be five problem sets distributed throughout the semester. These will be based on applying concepts and code from the lectures to a new question or problem using different real-world data. You will have at least a week to complete each problem set.

- All problem sets will be written using Quarto, which we'll cover in the first week of class. When you submit the assignment on Brightspace, submit both the rendered PDF output and the `.qmd` source code file.
- There is a one-time-use, no-questions-asked, 72-hour extension policy. The first time you miss a deadline on a problem set, you can turn in the assignment anytime within 72 hours for no penalty. You don't have to ask or inform me that you are taking the extension; it will be applied automatically.

  After the first 72 hours, or if you have used the extension on a previous assignment — there'll be a tracker on Brightspace in case you've forgotten if you used it — each day late is a 5 percentage point grade reduction.
- Your lowest problem set score will be dropped, and the remaining four will each be worth 12.5% of your final grade.
- **AI policy:** In the first few weeks of the class, while you're still learning R, I ask that you refrain from using ChatGPT or similar AI tools for help with your assignments. After we cover AI prompting in class, you are free to use AI for problem sets. I only ask that you use AI as a *complement* to learning data analysis and R, not as a *substitute* for your personal thinking or learning. Keep in mind that you won't have access to ChatGPT on the in-class exams.

Exams (50%). There will be two midterms, each worth 15% of your final grade, and a final exam worth 20%. These will be in-class, closed-book exams that test your understanding of data analysis concepts and principles, as well as your ability to interpret and debug small portions of R code.

Outline of topics

The exact number of lectures on each topic is subject to change, as I may slow things down if there's something important the class is stuck on.

1. **20th century computing.** Basics of R, RStudio, and Quarto. Navigating directories and loading data from the R console. (1 lecture.)

2. **Data wrangling.** `select |> filter |> mutate |> group_by |> summarize |> pivot.` (2 lectures.)
3. **Univariate analysis.** Continuous versus categorical variables. Measures of central tendency and spread. (3 lectures.)
4. **Prompting AI models.** Basics of how large language models work. Best practices for coding with AI. (1 lecture.)
5. **Data visualization.** ggplot2. Bar charts, histograms, and density plots for univariate visualization. Box plots, violin plots, and scatterplots for conditional relationships. Making maps. (4 lectures.)
6. **Correlation and regression.** Correlation versus causation. The statistical idea of "control." Bivariate and multivariate regression. Using regression models for prediction. (4 lectures.)
7. **Simulation and resampling.** Using simulation to gauge the sampling error in statistical procedures. Random number generation and for loops in R. Bootstrap standard errors. (4 lectures.)
8. **Text analysis.** Common text pre-processing steps. Extracting features from text. Feature selection and cross-validation for predictive models. (4 lectures.)

Important dates

Date	Time	Event
Friday, September 5	11:59pm	Problem Set 1 due
Friday, September 26	11:59pm	Problem Set 2 due
Thursday, October 2	2:45pm	Midterm 1 in class
Friday, October 17	11:59pm	Problem Set 3 due
Thursday, November 6	2:45pm	Midterm 2 in class
Friday, November 14	11:59pm	Problem Set 4 due
Tuesday, December 2	11:59pm	Problem Set 5 due
Wednesday, December 10	3:00pm	Final exam