

# POLS 8500: Computational Data Analysis\*

Instructor: [David Cottrell](#)

Spring, 2024

---

Days:	We @ 7:00pm–9:50pm
Meeting location:	Baldwin Hall 302
Office hours:	Online by appointment

---

## Course Description

The ability to use computational techniques to process and analyze data is an increasingly important skill for students of political science. It is not only a skill that is widely leveraged in political science research, but it is also a skill that can be leveraged across a range of analytical environments outside of academia. With that in mind, this course has been designed to help PhD students in political science become proficient in computational data analysis. Students will learn how to use the R programming language of to collect, clean, and analyze data. Over the course of the semester they will become familiar with the “tidyverse” approach to data science and they will acquire valuable computational skills in data visualization, data wrangling, text analysis, database management, web scraping, programming, and GIS.

## Prerequisites

This course is intended to be an intermediate R course. Students should have some familiarity with the RStudio IDE and some prior experience analyzing data using R. We will spend very little time in this class covering the basics of R computing.

## Method of instruction

The course surveys a series of topics related to computational data analysis in R. We will cover a new topic each week, as outlined in the agenda below. Students will be expected to complete a reading assignment before each class - which usually consists of one or two chapters from the textbook. In class, students will listen to a brief lecture and apply what they’ve learned through a set of hands-on data analysis activities.

## Textbook

For most of the course, we will be using the first edition e-book, [R for Data Science](#) (r4ds), by Hadley Wickham and Garret Grolemund.<sup>1</sup> This book describes how to use the popular **tidyverse** toolset to perform common data science tasks. Not only is the textbook free to access, but it uses a “learn-by-example” approach that makes it an excellent introductory textbook for inexperienced programmers.

## Computers

Bring your computer with R installed to class.

## Accessibility Needs

Students with special needs that require accommodation should notify me and the Office for Disability Services in the first two weeks of the course so appropriate arrangements can be made. All information and documentation of special needs is confidential.

---

\*The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.

<sup>1</sup>The authors published a second edition of the book in June, 2023. While most almost of all of the readings come from the first edition, I have assigned two chapters from the second edition. I refer to the second edition as “r4ds2” in the agenda.

## Grades

**Data assignments (30% of grade)** Each week, you will be given a data assignment that allows you to exercise the skills you've learned from the readings and lectures. The assignments are designed to be completed in-class, though they may require some extra work outside of class. Completed assignments must be submitted on ELC by the deadline at the end of the week. Full credit is given to completed assignments.

**Midterm (30% of grade)** The midterm exam will be made available online on Monday, March 11th. It will ask students to analyze data using the concepts covered up to that point. Students will have until Friday, March 15th to begin the exam, but only 24 hours to complete the exam once they have begun. The goal of the midterm is to test your ability to analyze data quickly before an impending deadline.

**Final (40% of grade)** The final exam will be made available online at the beginning of finals week. It is cumulative and will ask students to analyze data using the concepts covered throughout course. Students will have multiple days to begin the exam, but only 24 hours to complete the exam once they have begun. Like the midterm, the goal of the final exam is to test your ability to analyze data quickly before an impending deadline.

## Agenda

Class	Date	Topic	Reading
0	Jan 10	Review syllabus	<a href="#">Intro</a> to r4ds
1	Jan 17	Visualizing data (Part 1)	<a href="#">Ch.03</a> in r4ds
2	Jan 24	Visualizing data (Part 2)	<a href="#">Ch.28</a> in r4ds (and <a href="#">Ch.15</a> in r4ds)
3	Jan 31	Transforming data	<a href="#">Ch.05</a> in r4ds
4	Feb 07	Presenting data	<a href="#">Ch.27</a> in r4ds
5	Feb 14	Importing data	<a href="#">Ch.11</a> in r4ds
6	Feb 21	Tidying data	<a href="#">Ch.12</a> and <a href="#">Ch.13</a> in r4ds
7	Feb 28	Midterm review	no reading
8	Mar 13	Text as data	<a href="#">Ch.14</a> in r4ds
9	Mar 20	Dates/Time as data	<a href="#">Ch.16</a> in r4ds
10	Mar 27	Web scraping	<a href="#">Ch.24</a> in r4ds2
11	Apr 03	Databases	<a href="#">Ch.21</a> in r4ds2
12	Apr 10	Loops/iterations/simulation	<a href="#">Ch.21</a> in r4ds
13	Apr 17	GIS	<a href="#">Making maps with R</a>
14	Apr 24	Final review	No reading