

Introduction to Data Science and Political Science Research Methods

Samuel Frederick

Course Overview

This course will focus on the fundamentals of data science and political science research. We will start by learning how to program in R, a statistical coding language. Being conversant in R will help us collect and analyze data about questions we are interested in. Once we've established a foundation in coding, we will delve into the probability and statistics that forms the basis of quantitative political science research. The remainder of the course will focus on understanding causality in social science research, including learning how we can ask (and answer) causal questions convincingly.

Throughout the course, we will engage with published work in political science, applying the techniques we learn to understand and advance existing work. A key assignment in this course will be the creation of a research paper which builds upon previous research to offer new insights into politics. The skills learned in this course are broadly applicable to a wide variety of potential career paths.

Learning R is like learning a new language. It will be challenging. I believe everyone is capable of learning R, and I am here to help you succeed in this class. If you have questions about topics in the class, please don't hesitate to ask in lecture, office hours, or via email; I guarantee other students will have similar questions, and asking for clarification will help me help you.

Course Learning Goals

After taking this course, students will be:

- Proficient in the R programming language
- Experienced in the collection, cleaning, visualization, and analysis of data
- Familiar with the fundamentals of probability and statistics
- Knowledgeable about the principles of causal inference and modern political science
- Skilled in the critical evaluation of published research
- Able to effectively communicate about research methods and findings

Grading

Problem Sets

Approximately every other week there will be a problem set due, which asks you to apply the materials we have covered in lecture to real-world data problems.

Presentation

Each week, a group of students will give a 10-minute presentation about a journal article from political science or another social science. Presentations should give a brief overview of the article and then describe how the article applied a technique we have learned in class. Presentations should critically engage with the work, including offering suggestions for how the paper's implementation of the technique could be developed or expanded.

Final Project

Students will, over the course of the semester, implement either (a) an original research paper or (b) a replication and extension of an existing political science article. Replications should involve a significant methodological or data-based extension of the existing work. I will set up meetings with students early in the semester to discuss potential topics.

By the end of the fourth week, students will submit a memo, outlining their project plans. For students writing an original research paper, this memo should identify the topic, a feasible data source or sources, and summaries of 10 academic papers relevant to the proposed topic. For students writing an extension of existing work, this memo should identify the work that will be extended, the steps that will be taken to extend the work (including identifying relevant data sources), and summaries of 10 academic papers relevant to the proposed extension.

By the end of the ninth week, students will submit a memo, containing summaries and visualizations of the data as well as preliminary analyses.

In the final week of class, students will give a brief presentation of their work, describing their topic, presenting their analyses, and showing their findings. Students will then submit their final projects in journal article form.

Required Textbook

There is one required book for this course:

Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press

A second required book is available online:

Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Golemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd Edition. URL: <https://r4ds.hadley.nz>

Where possible, I will provide alternative sources which are freely available online.

Other Reading Options:

Ethan Bueno de Mesquita and Anthony Fowler (2021). *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press

Elena Llaudet and Kosuke Imai (2022). *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press

R and RStudio

Over the course of the semester, we will be working in the R programming language using the RStudio environment. Students should download R and RStudio prior to the first class by following the instructions at this link: <https://posit.co/download/rstudio-desktop/>. If you are unable to download R or RStudio, please feel free to reach out for help. You may also consider trying RStudio in the cloud for free if you are unable to download the desktop version (available here: <https://posit.cloud/plans/free>).

Course Schedule

1 Overview of R

Week 1. Introduction to R

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapters 1, 2.2**
- Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Golemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd Edition. URL: <https://r4ds.hadley.nz>, **Chapters 3, 7**
- Elena Llaudet and Kosuke Imai (2022). *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press, **Chapter 1**

Available here: <https://press.princeton.edu/books/paperback/9780691199436/data-analysis-for-social-sciencepreview>

Week 2. Summarizing Data with Plots and Statistics

Reading Data in R

- Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Golemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd Edition. URL: <https://r4ds.hadley.nz>, **Chapters 8, 4, 6**

Visualizing and Summarizing Data

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapters 2.6, 3.3, 3.6**

- Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Golemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd Edition. URL: <https://r4ds.hadley.nz>, **Chapters 2, 10, 11**

Suggested Readings:

- Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen (2023). *ggplot2: Elegant Graphics for Data Analysis*. Springer. URL: <https://ggplot2-book.org>
- Elena Llaudet and Kosuke Imai (2022). *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press, **Chapter 3**

Week 3. Web Scraping

- Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Golemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd Edition. URL: <https://r4ds.hadley.nz>, **Chapter 25**

2 Introduction to Probability

Weeks 4 and 5. Probability

Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapters 3.4, 6**

3 Political Science Research

Weeks 5 and 6. Causality

Thinking Causally and the Experimental Ideal

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 2.1, 2.3, 2.4**

Suggested Readings:

- Ethan Bueno de Mesquita and Anthony Fowler (2021). *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press
 - Causation: Chapters 2 and 3
 - Experiments: Chapter 11
- Elena Llaudet and Kosuke Imai (2022). *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press, **Chapter 2**

Observational Data

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 2.5**

Week 7. Uncertainty

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 7.1-7.2**

Suggested Readings:

- Ethan Bueno de Mesquita and Anthony Fowler (2021). *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press, **Chapter 6**

Weeks 8 and 9. Linear Regression

Introduction to Regression

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 4**

Suggested Readings:

- Ethan Bueno de Mesquita and Anthony Fowler (2021). *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press, **Chapter 5**
- Elena Llaudet and Kosuke Imai (2022). *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press, **Chapter 4**

Uncertainty and Inference with Regression

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 7.3**

Suggested Readings:

- Elena Llaudet and Kosuke Imai (2022). *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press, **Chapters 5, 7**
- Ethan Bueno de Mesquita and Anthony Fowler (2021). *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press, **Chapters 9 and 10**

Visualizing Regression Results

- Vincent Arel-Bundock (n.d.). *Model Summary*. URL: <https://modelsummary.com/articles/modelplot.html>

Week 10. Causality with Observational Data

Difference-in-Differences Design

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 2.5.3**
- Nick Huntington-Klein (2022). *The Effect: An Introduction to Research Design and Causality*. Boca Raton: CRC Press. URL: <https://theeffectbook.net/index.html>, **Chapter 18.1, 18.2**

Suggested Readings:

- Ethan Bueno de Mesquita and Anthony Fowler (2021). *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press, **Chapter 13**

Week 11. Causality with Observational Data

Regression Discontinuity Design

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 4.4.3**
- Nick Huntington-Klein (2022). *The Effect: An Introduction to Research Design and Causality*. Boca Raton: CRC Press. URL: <https://theeffectbook.net/index.html>, **Chapter 20**

Suggested Readings

- Ethan Bueno de Mesquita and Anthony Fowler (2021). *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press, **Chapter 12**

Week 12. Heterogeneous Treatment Effects and Visualizing Interactions

Estimating Interactions

- Kosuke Imai and Nora Webb Williams (2022). *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press, **Chapter 4.4.2**

Visualizing Interactions

- Vincent Arel-Bundock (2023). *The Marginal Effects Zoo (0.14.0)*. URL: <https://marginaleffects.com>, Chapters 2-6

Week 13. Student Research Presentations