FILE 170 · POLI 170A · SUMMER 2026

Applied Data Analysis

An applied introduction to causal inference for the social sciences.

Contents

Instructor	Benjamin Noble · b2noble@ucsd.edu
Meetings	Monday and Wednesday, 11:00am to 1:50pm (fully remote; Zoom link on Canvas)
Office hours	TBD

You can come to office hours to ask me questions about the course content (especially if you're having trouble). But you can also come to office hours to say hello, ask me about my research, learn what political scientists do, and tell me about your interests (academic or otherwise). I'd love to meet you.

Course Description

Does ideological extremism cause a candidate's vote share to increase? Does inflation cause presidential approval to decline? Does democratization cause countries to become more peaceful? These are all *causal* questions. They all take the form: does X *cause* Y?

As social scientists, we often care about causal questions, but it is difficult to answer these questions definitively. Unlike a pharmaceutical company that can randomly assign people to take or not take a drug, we cannot randomly assign ideologies to candidates, inflation to presidents, or democracy to countries (but it would be nice if we could!). Instead, we have to apply creative strategies to try to learn whether X causes Y with *observational* data. That is, data we observe from the real world (not an experiment).

In this course, we are going to learn about the social science toolkit that will help us analyze data from the real world and, sometimes, draw causal conclusions about whether X causes Y.

In a typical class, you can expect:

- Lecture on a new statistical/data analysis concept.
- In-class "lab" exercise where you will work in small groups to apply concepts we learned in the lecture.
- Review solutions to the lab exercise, answer questions.

Required Texts and Materials

All textbooks and materials used in this course are available for free online.

Our primary textbook is:

- Causal Inference: The Mixtape (referred to below as "Mixtape") by Scott Cunningham (2021, Yale University Press).
  - *Note: this textbook includes a good deal of mathematical notation. Because this is an* applied *course, our focus is on gaining a conceptual understanding of the math and implementing it with code and data.*

I will also assign readings from the below textbook. This book covers many of the same concepts as our main textbook in a different, less technical way. You may want to reference this book if you get stuck or prefer its approach.

- The Effect: An Introduction to Research Design and Causality (referred to below as "The Effect") by Nick Huntington-Klein (2025). Written for an applied audience, with R, Stata, and Python code throughout.

In-class labs, problem sets, and the final project will require use of the R programming language. We will cover R on the first day of class, and programming knowledge is not required as a prerequisite. To program in R, we will be using UCSD's DataHub, which you can access with your UCSD Student SSO.

Course Format

This class is fully remote. All sessions meet over Zoom. The meeting link is available on Canvas.

Lectures will be recorded and posted to Canvas after class. However, attendance is required.

Please keep your camera on during class sessions. This is a remote class with a lot of group work and discussion. Seeing each other matters.

If you experience a technical failure during class (your connection drops, Zoom crashes, your Wi-Fi goes out), try to reconnect as soon as you can. If you miss a lab because of a tech issue, that counts toward your one excused lab submission described below. If technical problems become a recurring issue for you, get in touch and we will come up with a plan.

Assignments

- **Attendance and Participation (15%)** including attending lecture synchronously, turning your camera on, and actively participating in lab activities and discussions.
- **Lab assignments (25%)** completed in class and submitted at the end of each session, graded for effort and participation.
- **Problem sets (30%)** to be completed individually, applying key course concepts.
- **Final group project (30%)** to be completed in class and presented (privately) during the final exam period.

Attendance and Participation (15%)

You are expected to attend every class session synchronously and have your camera on for the full session.

Because life happens, you will be allowed one excused class absence with no questions asked; the lab from that day will be dropped from your final grade. You do not need to email me to use this policy. It will be applied automatically when I calculate final grades.

Beyond your one free absence, you must contact me in advance to miss additional classes without penalty, and waiving the penalty is at my discretion. If I grant an excuse, the lab from that day may be dropped or made up, depending on the situation.

Missing class without prior contact (barring exceptional emergencies) will result in a zero on that day's lab. Be thoughtful about how you use the free absence policy; I will not grant additional

exceptions for routine reasons.

Lab Assignments (25%)

All class sessions include group lab activities, which allow you to practice what we have learned in a low-stakes, group setting. Lab groups will be assigned during the first week of class and will serve as your final project groups as well (described below).

Labs are graded for participation and effort. Although these are group assignments, each member is responsible for submitting their own code and results. You will submit your `.Rmd` file and knitted `.html` file via Canvas at the end of the class session. We will cover these file types on the first day of class. You do not need to finish the entire lab to receive full credit, and solutions will be posted after class so you can check your work.

After lab, we will review solutions together as a class. During this review, you may be called upon to share your screen and walk us through what your group did. I will rotate through students over the course of the quarter, so everyone should expect to be called on at some point. This is a chance to learn from each other's approaches and to make sure each of you is engaging with the material. You will not be penalized if you aren't sure of your answer. This is simply an opportunity to share your work and demonstrate your group's approach.

Problem Sets (30%)

You will complete three problem sets over the course of the quarter, each covering material from the lectures and in-class labs. These problem sets will include a mix of conceptual and coding questions.

When you submit your problem set, you must upload the code (rmd) and html files. Your code should include brief comments explaining what the code is doing. If you cannot arrive at the correct answer, you can still earn many of the possible points provided you make an attempt, show your work, and explain what you are doing. I will provide a template that will help you fulfill these requirements.

You must work on problem sets on your own. You may use our textbooks, notes, the internet, even AI. But you may not consult with other students, tutors, etc. You may discuss your problem sets with me during office hours.

Final Project (30%)

This course will conclude with a group final project. Your group will work together to complete a timed data analysis task during the final class session on Wednesday, July 29. You will present your findings (privately, not in front of the whole class) during our final exam period on Friday, July 31.

Your final project grade will be based on three components: the quality of your group's analysis and presentation (the bulk of the grade), a short individual reflection that each group member submits describing their contributions and what they learned, and a confidential peer evaluation completed by each group member, which I may use to adjust individual grades within a group if there is clear evidence of unequal contribution.

Course Policies

Academic Integrity

I take academic honesty and integrity seriously. You must adhere to the assignment-specific requirements in terms of what you may/may not consult in completing your work. Please see the UCSD policy on academic integrity for more information.

Use of AI

In this class, I encourage thoughtful use of generative AI tools (such as ChatGPT, Claude, Gemini, TritonGPT, etc). These tools are incredibly powerful and can help you with both the statistical and coding concepts taught in this course. However, over-reliance on these tools poses some risk. They can make you *feel* like you understand something more than you do, and while they are less likely to make explicit errors than in the past, they can mislead you if you do not understand the underlying concepts. It is your responsibility to be a careful consumer of these tools and ensure that you validate anything you learn from them. It is also your responsibility, as a student in this course, to understand the answers they provide.

Here are some recommendations:

- If you are having trouble with coding, you should begin by working with one of these AI tools. They are often very good at answering basic coding questions and providing example code that can help you solve your problem. Often, if you copy/paste your error into the chat window, AI can provide a solution. Try this before asking me questions about coding (not because I don't want to help, but because this is how real data analysts solve problems today). Don't just ask for answers though. Ask the LLM to explain its work.
- These tools work best for topics and concepts you "mostly" know. If you are working with an LLM on completely new material and it makes a mistake or misleads you, you will not know or have any intuition that it is wrong. As such, I encourage use of these tools in consultation with material from lectures and the textbooks. Go back and forth between them to master these topics, and verify what you learn from the LLM in our other materials.
- You should be aware that these systems often save your conversation history and may use it to train future models. As such, never put any sensitive or private information into a prompt.

Late Submissions

As I post lab submissions after class, labs may not be submitted late. Late problem sets will incur a one letter grade penalty for each 24 hour period they are late. For example, if a problem set is due at 10:59am, a late submission delivered between 11:00am on the due date and 10:59am the following day will automatically lose one letter grade. All solutions will be posted one session after the problem set is due. After solutions have been posted, late problem sets cannot be accepted.

Requests for Re-Grades

If you believe an error has been made, you have *one week* following the return of the assignment to request a regrade. After this point, re-grades cannot be requested. To do so, please email Professor Noble with a brief explanation of why you are requesting a re-grade as well as evidence from our course materials justifying the request. I reserve the right to refuse to re-grade, and if we do re-grade, please note it may result in a lower grade.

Communication

For all questions or comments, you may get in touch with me during my office hours listed on this syllabus, or via email. If your email requires a response, you can expect one within 1–2 business days. If you email me over the weekend, the clock begins Monday morning.

Note: if you contact me the night before the homework is due, I will not respond in time to provide any advice before the deadline. Please plan and work ahead.

Accommodations

Students needing accommodations for this course due to a disability must provide a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities. Students are required to discuss accommodation arrangements with instructors, IAs, and OSD liaisons in the department.

Other resources, including the inclusive classroom statement, advising, and resources to support equity, diversity, and inclusion, and more can be found in the Additional Resources section below the reading list.

Course Schedule and Readings

Based on your learning style, you may find it helpful to complete the readings before or after lecture. Ultimately, it is up to you when you want to do the readings. You can always refer to this syllabus for the most updated information about the course.

June 29 · Introduction and R Skills

- Readings:
  - Does X cause Y? An in-depth evidence review by Holden Karnofsky.
- Optional:
  - For help with R: Data Science in R: A Gentle Introduction by James Scott, Chapter 1, Chapter 2, and Chapter 4.
- **Problem Set 1 assigned.**

July 1 · Potential Outcomes

- Readings:
  - Mixtape, Chapter 4 (stop before "4.2 Randomization Inference").

July 6 · Extraordinary Least Squares

- Readings:
  - Jared Wilber, September 2022. Linear Regression. *MLU Explain*.

July 8 · Potential Outcomes (II)

- Readings:
  - The Effect, Chapter 10.

July 13 · Instrumental Variables

- Readings:

  - Mixtape, Chapter 7, sections 7.1–7.2, 7.3.1, 7.5.
- Optional:
  - The Effect, Chapter 19.
- **Problem Set 1 due before class, at 10:59am PT.**
- **Problem Set 2 assigned.**

July 15 · Instrumental Variables (II)

- No reading.

July 20 · Regression Discontinuity

- Readings:
  - Mixtape, Chapter 6, sections 6.1–6.2.3, 6.3.
- Optional:
  - The Effect, Chapter 20.
- **Problem Set 2 due before class, at 10:59am PT.**
- **Problem Set 3 assigned.**

July 22 · Difference-in-Differences

- Readings:
  - Mixtape, Chapter 9 (stop before "9.5 The Importance of Placebos in DD").
- Optional:
  - The Effect, Chapter 18.

July 27 · Buffer / Synthesis

- No readings.
- **Problem Set 3 due before class, at 10:59am PT.**

July 29 · Final Project Session

At the start of class, you will receive a dataset and instructions. You will work in your group to complete the analysis. You will have the full three hours of class to complete this work. You will submit your completed analysis at the end of class.

July 31, 11:30am–2:29pm · Final Project Presentation

Each group will present their final project and answer questions about their analysis (privately, not in front of the class).

Grading Scale

Letter	Range
A+	96.5% and above
A	93.5% to 96.5%
A-	89.5% to 93.5%

Letter	Range
B+	86.5% to 89.5%
B	83.5% to 86.5%
B-	79.5% to 83.5%
C+	76.5% to 79.5%
C	73.5% to 76.5%
C-	69.5% to 73.5%
D	59.5% to 69.5%
F	Below 59.5%

Additional Resources

These additional resources and the language come directly from the UCSD Political Science Department.

Inclusive Classroom Statement

The IAs and I are fully committed to creating a learning environment that supports diversity of thought, perspectives, experiences, and identities. We urge each of you to contribute your unique perspectives to discussions of course questions, themes, and materials so that we can learn from them, and from each other. If you should ever feel excluded, or unable to fully participate in our class for any reason, please let me know, or please consult the Department's Report an Issue page for additional campus resources to support you, and diversity, equity, and inclusion in our classroom, and beyond.

Additional resources to support equity, diversity, and inclusion in our classroom, and beyond, may be found here:

- https://diversity.ucsd.edu/
- https://students.ucsd.edu/student-life/diversity/index.html
- https://regents.universityofcalifornia.edu/governance/policies/4400.html

Resources to Support Student Learning

- Library Help, eReserves and research tools: https://library.ucsd.edu/ask-us/triton-ed.html
- Writing Hub: https://commons.ucsd.edu/students/writing/index.html
- Supplemental Instruction: https://aah.ucsd.edu/supplemental-instruction-study-group/index.html
- Tutoring: https://aah.ucsd.edu/content-tutoring/index.html
- Mental Health Services: https://caps.ucsd.edu
- Community Centers: learn about the different ways UC San Diego explores, supports, and celebrates the many cultures that make up our diverse community: https://students.ucsd.edu/student-life/diversity/index.html

Academic Advising

Students who have academic advising questions related to the Political Science major should contact the department's Undergraduate Advisor, Zain Sharifi, via the Virtual Advising Center. Academic advising questions often include (but are not limited to): add/drop deadlines, course enrollment policies, planning major and minor requirements, quarter-by-quarter plans, department petitions and paperwork, and referrals to campus and student support services.

Equity, Diversity, and Inclusion Offices

Office of Equity, Diversity, and Inclusion. 858.822.3542 · diversity@ucsd.edu · https://diversity.ucsd.edu/

Office for the Prevention of Harassment and Discrimination. https://ophd.ucsd.edu/ · ophd@ucsd.edu · (858) 534-8298

UCSD Office of the Ombuds. https://ombuds.ucsd.edu/ · To reach a Confidential Ombudsperson, please call 858-534-0777.

UCSD's Principles of Community

To foster the best possible working and learning environment, UC San Diego strives to maintain a climate of fairness, cooperation, and professionalism. These principles of community are vital to the success of the University and the well being of its constituents. UC San Diego faculty, staff, and students are expected to practice these basic principles as individuals and in groups. The Principles of Community and the Student Code of Conduct support equity, diversity, and inclusion in our classroom.

Acknowledgements

I thank Pamela Ban, Christopher Lucas, and Tiago Ventura for inspiration. I also thank Christopher Lucas for providing some of the materials that were used in this course.