**Poli 170A: Applied Data Analysis for Political Science**
Spring 2024
(Syllabus version updated 4/5/24)

**Instructor:** Rachel Skillman ([rskillman@ucsd.edu](mailto:rskillman@ucsd.edu))
**Lectures:** MWF 11:00 – 11:50a in CSB 002
**Office Hours:** Wednesdays 12:00 – 1:00p in SSB 343, sign up for a slot in Rachel’s OH [here](https://example.com)

**TA:** Bert Wilden ([bwilden@ucsd.edu](mailto:bwilden@ucsd.edu))
**TA Office Hours:** Fridays 1:00 – 3:00p on [Zoom](https://zoom.com), sign up for a slot in Bert’s OH [here](https://example.com)

**Online Content:** https://canvas.ucsd.edu/courses/56028

**Overview:** The growth of data analytics is changing the way policymakers, businesses, and individuals operate. Across all these categories, people are increasingly able to pose and answer complex questions using new evidence.

This course introduces students to methods and tools for empirical data analysis in social science research, methods that can be applied to a large range of questions ranging from policy to business to your everyday life. How do we predict the outcome of presidential elections? Do countries become less democratic when leaders are assassinated? How can data analysts detect election fraud? Does the internet encourage democratic backsliding? These are a few among the many social science questions that data analysis gives us the power to explore, and this course will give you the tools to analyze existing research like this and conduct your own.

Part of this task is conceptual: helping students to think sensibly and systematically about research design. To this end, students will learn how theory, measurement, and data fit together to help us understand the concepts we care about. But part of our task is practical as well: students will interact with data, analyze it using the various techniques we cover, and present their findings in a paper and presentation for class. As part of this process, students will utilize R which is a powerful, flexible, and commonly known statistical tool in social science research. Our overall goal is to provide students with the foundation necessary to analyze data in their own research and to become critical consumers of statistical claims made in policy reports, the news media, and academic research.

As a purely practical matter, this class is highly recommended for students who plan to write senior theses or pursue graduate work in the social sciences: the statistical and computing material you learn will be helpful for those undertaking such projects. This course will also prepare students for work in any setting that requires a social scientific approach to data analytics, from policy non-profits to government, from Silicon Valley to Wall Street and beyond.
**Prerequisites:** POLI 30/POLI 30D or instructor consent. Although the class is relatively heavy on applied statistical techniques, the mathematical demands are relatively light; our focus will very much be on helping students understand the intuition of the methods we use and interpret the outputs of these techniques rather than their internal workings. Because some students may be newcomers to statistical analysis, we will begin with an overview of foundational material that is necessary for all social science research. However, because this is an upper-division course and many students will already have completed other coursework in data science or statistics, our pace will be much quicker than POLI 30. It is assumed that the student has the mathematical background to progress through the materials a bit faster than in a true introductory course.

**Software:** We will use the open-source statistical software R (http://www.r-project.org, download here). R can be more powerful and flexible than other statistical software, such as SPSS and Stata, but it can also be more complex to learn. Although the learning curve is steeper, the reward is worth it. R uses an intuitive base language and because it is open-source, there are tons of packages available to users and packages are constantly updating and innovating. We will also use RStudio (http://www.rstudio.com), a user interface that simplifies common operations. **You must install both programs before the second week of class (Mon, April 8).** If you have trouble, contact the professor or TA, and time will be available during class in Week 2 for troubleshooting.

For help with R, I recommend the following free resources:

- https://r4ds.had.co.nz
- https://www.jaredknowles.com/r-bootcamp/
- https://www.youtube.com/user/TheLearnR/videos
- https://education.rstudio.com/learn/beginner/
- https://www.coursera.org/collections/learn-r
- https://www.datacamp.com/courses/free-introduction-to-r
- https://stackoverflow.com/

For those of you that are new to R or coding more generally, I **highly recommend** that you familiarize yourself with the interface by briefly exploring these resources before the first day of class.

**Reading and Textbooks:** There are no required textbooks for this course. For those students that would like additional practice or a more in-depth consideration of the topics that we discuss in class, I recommend Kosuke Imai’s *Quantitative Social Science* textbook, which is used in several courses at UCSD, as well as Peter Dalgaard’s *Introductory Statistics with R* and Andrew Gelman and Jennifer Hill’s *Data Analysis Using Regression and Multilevel/Hierarchical Models*. I will sometimes assign readings from these textbooks (or other online resources) but all required readings will be made freely available. You are expected to have done the readings before class – keeping up each week will be integral for your success in a fast-paced course like this one. Reading the assigned articles will also be used to assess your participation grade (see details below).
**Requirements:** The course requirements consist of:

- **3 Online Homeworks (40%),** due April 26, May 10 & May 24 at 11:59pm PST  
  10% for the first two and 20% for the third which will serve as a take-home final
- **Research In the Wild Analyses & Participation (20%)**  
  Research In the Wild Analysis – 6 analyses (9 to choose from) worth 3% each (18%) + Introduction survey (1%) + 1 comment on final presentation (1%)
- **Final Research Project (40%),** due June 12 (Wednesday!) at 11:59pm PST  
  20% for the in-person presentation and 20% for the final paper

I use the following grading scale: “A-” = [90-93.3), “A” = [93.3-96.6), “A+” = [96.6-100], with other letter grades following analogous intervals.

**Homeworks (50%)**

There will be 3 online homework assignments (available on Canvas) throughout the course, designed to ensure that you are staying up to date with the course material in a fast-paced summer session. The first two homeworks will be relatively short and weighted identically (10% each, due 4/26 and 5/10 at 11:59pm PST) and the last will be longer and serve as a take-home final (weighted 20%, due 5/24 at 11:59pm PST).

- The homeworks will be released after lecture (at 12pm) on Mondays the week they are due, meaning you will have ~4.5 days to complete each homework.
- All homeworks (including the final) are open-book, open-note, open-web. However, all homework is to be completed independently and all work submitted should be your own.
- All homeworks will be cumulative.
- Late submission will not be accepted unless you receive special permission from the instructor in advance.
- Homeworks will be multiple-choice or short answer (sometimes asking for proposed code or analytical steps) and completed electronically on Canvas. They will not be designed to trick you and most of the answers will be readily available in the class slides or code.
- For analytical questions, you should include your intermediate steps, as well as comments on those steps when appropriate. For data analysis questions, include annotated code as part of your answers. All results should be presented so that they can be easily understood. **As much work as you can show will always help us to better understand your reasoning and will often allow us to give you more partial credit.**
Final Research Project (40%)
The final project will consist of a brief in-person presentation in class in Week 9 or 10 (13 min and worth 20%) and a short research paper due 6/12 at 11:59pm PST (min: 1000 words and worth 20%) that applies a method learned in this course in an original way to an empirical problem of substantive interest to you.

- Though not required, I strongly encourage you to work with a partner or group on your final project.
- If you work with a group/partner, you will each receive the same grade on your final project.
- The most time-consuming part of the research project will likely be locating, collecting, and analyzing the data. Please begin thinking about what data you might like to use as soon as possible.
- Some class sessions will be devoted to workshopping your final projects. The instructor and TA will help guide you through the intermediate steps of completing your research project, both individually and collectively.
- Please turn in pdfs containing your research paper and presentation slides electronically through Canvas.
- Presentations must be given IN-PERSON. Failure to present in-person during class time in Weeks 9 and 10 will result in a 0 for the Presentation grade, unless previous arrangements are made with the instructor.

The final project requires you to pose a novel social science question and test it using data and the techniques we’ve learned in class. When you present your original research it should contain the following components in this order:

#1. Question: What research question are you trying to answer? What is the variation you are trying to explain? This section should also contain a broad statement about why we should care about your research question. What is puzzling or interesting about your question? Is it a topic of social, economic or political importance? Is the phenomenon of interest prevalent? Have other people trying to answer this question found conflicting or puzzling results? Will your question/answer suggest something in opposition to an established theory? Are you using data to answer your question that no one else has used before?

⇒ “Despite the growing research exploring public opinion toward climate change, we still know little about how people respond to environmental destruction in their communities. Does exposure to deforestation change individual political preferences?”

#2. Answer: What is your proposed theoretical answer to your research question? Your answer should include the causal mechanism by which change in the independent variable leads to change in the dependent variable. This section should also contain a literature review: what have other

---

1 Why your question is important can be answered in a variety of ways. It should not be answered by relating this to a personal characteristic, interest or specific life experience (i.e. “Studying women’s political representation is important to me because I am a woman.” or “Studying the effect of protected wildlife habitats on biodiversity is important because I believe protecting animals is important.”) We are often motivated to study topics that feel personally important or relevant to us, but we have to communicate the broader importance of our inquiry to the scientific community at large to be taken seriously as researchers.
scholars said about your topic or question? Finding literature that speaks to your question is not always easy, so feel free to consider how similar research on adjacent topics (or research from different fields, like public health, law or economics) inform your question & answer.

→ “I argue that deforestation in one’s community should lead people to increasingly vote for the Democratic party because they are more likely to have a preference for pro-environment policies that protect forests. As Author 1 (2020) demonstrates, people express more pro-climate preferences after experiencing a climate disruption. In contrast, Author 2 (2020) shows the effect is conditional on partisanship...”

#3. Data & Methods: What are the sources of data you use to answer your research question? What statistical method do you use to estimate the causal effect? You can even include information about the number of observations in your analysis. Importantly, this component should mention your unit of analysis (country, state, county, zip code, census tract, individual, village, project level?) and temporal information about which observations are included in the estimation (which year(s), month, etc. are the data from?).

→ “Using Hansen et al.’s Global Forest Change dataset and data on individual vote choice from the ANES, I estimate the effect of forest cover loss on reported votes for the 2008 Democratic presidential candidate for 30,000 non-metropolitan zip codes using linear and geographically-weighted regression techniques.”

#4. Findings: What did your analysis tell you about your research question & theory? Get specific. Include, if possible, the size of the effect and any interesting nuances you found. It’s okay if the answer you proposed ends up being wrong! Null results can be important theoretical contributions, especially if they are surprising.

→ “I find that deforestation does increase the likelihood of voting for the democratic candidate in 2008 by 0.39 percentage points, controlling for other predictors of vote choice. However, this effect is conditional on region – respondents in the Southern U.S. were not more likely to vote for the democratic candidate even after exposure to high rates of deforestation.”

#5. Contribution & ways to improve: What does your work contribute to social science research overall? What was interesting or surprising about your findings? If there was anything you would have liked to include but could not, or further research needed, mention it here.

→ “As environmental shocks and habitat destruction become increasingly common, my findings suggest that partisanship may shift as a result. Future research should also explore the effect of state-level policies that regulate forest conservation on people’s political response to deforestation in their communities.”

You will receive further instructions on how to conduct and present compelling social science research using this structure throughout the course.

---

2 Note that my findings now discuss another variable: region. I am stating that the treatment effect of deforestation is conditional on which region someone lives in. Yet, my proposed answer did not make any suggestions about the “heterogeneous treatment effects” of region on my relationship of interest (exposure to deforestation → presidential vote choice). This is not good! You should be careful to theorize explicitly about all conditional effects you estimate and report in your findings.
Participation (20%)
Participation will be measured through 6 Research “In the Wild” Analyses that you may select from 9 possible lectures (6 analyses worth 3% each for 18% total) and 2 smaller tasks worth 1% each (2% total). Each Friday during Weeks 1-9 we will either watch a short video or read an article and then discuss it. In order to practice coming up with good research questions and identifying the components of social science research, you are responsible for. Research in the Wild Analyses are due Friday the week of the in-class discussion (i.e. analyses for Week 1 are due Friday 11:59p PST of Week 1) and are generally graded for completion unless it is clear you did not watch/read the material. Possible answers will be discussed in class to encourage attendance, and you may use answers discussed in class as part of your submission.

Research in the Wild Analyses (18%)
- If we watch a video, you are responsible for coming up with a research question inspired by the video content, and identifying a potential independent and dependent variable that would allow you to test your research question.
- If we read an article, you are responsible for identifying the components of social science research in each article (Question, Answer, Data & Methods, Findings, Contribution). Most of this information should be available in the abstract of the article, but incorporating information from paper itself is also encouraged. Each component can and should be a sentence from the abstract.
- Research Analyses will be graded for completion and effort.
- Analyses will be due by Friday at 11:59pm PST each week. For example, if you choose to submit an analysis for the video assigned Friday 4/5, you must submit it on Canvas by 4/5 at 11:59pm PST.

Introduction Survey (1%)
- By the end of the second week of class, students will complete a short, ungraded survey available on Canvas to introduce themselves and certify students’ academic activity (Introduction Survey: Tell Me About Yourself #FinAid). Completing your introduction survey by Friday 4/12 at 11:59pm PST is worth 1% of your participation grade.

Comment on Final Presentations (1%)
- After in-class final presentations in Weeks 9 and 10, students must make one constructive comment to another group/student on their project. This is to get students accustomed to providing and receiving feedback on their work. Please send the student(s) an email with your comment by Friday 6/7 at 11:59pm PST and cc the instructor and TA.³

---
³ If, for whatever reason, you cannot attend any of the final project presentations, you must submit a 1 paragraph self-reflection to me via email by Friday 6/7 at 11:59p PST.
Class policies:

Attendance. In-person attendance is strongly encouraged. Lectures will not be recorded. During class, we will also take time to workshop your final projects and you will receive personalized feedback and have an opportunity to discuss your project with the instructor, TA and other students. These intermediate work sessions will be extremely helpful for completing your final project successfully and attendance is highly recommended. If you must miss a class, it is your responsibility to contact the instructor or fellow students to review the material you missed. Final project presentations are in-person. Failure to present in person will result in a 0% on the final project presentation grade.

Academic Honesty and Plagiarism. All of your graded work must be done by you. If you are unfamiliar with the University’s policy on academic integrity, please consult this resource. Any student who is caught cheating or plagiarizing will receive a failing grade for the course and will be reported to the Academic Integrity Office for administrative sanction.

Late Assignments. Make-up assignments are only offered under valid and documented circumstances. If you know you will miss a deadline for a legitimate reason, notify me at least a week in advance. Email is perfectly acceptable. If you cannot contact me in advance, you must do so as soon as possible. I will work with you to resolve reasonable problems, but it is your responsibility to arrange these solutions with me. No make-up work will be graded after 6/14 at 11:59pm PST.

Emails. The instructor and TA will make every effort to reply to you within 24 hours of receiving any email correspondence. However, you should not expect a reply outside of working hours (8am - 5pm, M - F). While we will always try to assist you as soon as possible, including on weekends, do not count on receiving a response outside of working hours. Please work to begin assignments with plenty of time to get assistance if needed with access, clarification, etc. IF A SUBMISSION LINK IS NOT WORKING, for example when you try to hand in a problem set at 11:58pm the day it is due, simply email it to the instructor and TA and your submission will not be marked as late.

Grades and Appeals. You will be graded solely on your academic performance. This includes clarity of thought, knowledge of the material, composition, spelling, and grammar. Students can appeal grades that they believe are incorrect. Grade appeals will consist of a single typed page that identifies the problem and presents a reasoned argument that the grade fits the appeal criteria.

Disability Accommodations. Students who will request accommodations should register with the Office for Students with Disabilities (858-534-4382 or osd@ucsd.edu, please find more info here) and provide me with documentation outlining appropriate accommodations. I am happy to meet with you during my office hours to discuss your needs.
Course Schedule *(subject to change)*

**WEEK 1**

**Mon. April 1 – Lecture 1: Introduction**
- Course introduction and logistics

**Wed. April 3 – Lecture 2: The Formula**
- Course introduction and logistics

**Fri. April 5 – Lecture 3: Research in the Wild**
Video: *Inside Rio's favelas, the city's neglected neighborhoods*
- Asking good research questions
- Independent and dependent variables
- Unit of analysis

*_Research in the Wild Option #1 due 4/5 at 11:59pm PST_*

**WEEK 2 [Project goal: identify research question]**

**Mon. April 8 – Lecture 4: Measurement**
- Operationalization
- Reliability, validity and measurement error
- Types of variables

**Wed. April 10 – Lecture 5: R Workshop**
- Installation troubleshooting and introduction to the interface

*Please have R and RStudio downloaded before class*

**Fri. April 12 – Lecture 6: Research in the Wild & Hypotheses**
Article: *International Sports Events and Repression in Autocracies: Evidence from the 1978 FIFA World Cup* (Scharpf, Glabel, & Edwards, 2022) also available on Canvas
- Moving from research questions to hypotheses & testable implications

*_Introduction survey due 4/12 at 11:59pm PST_*

*_Research in the Wild Option #2 due 4/12 at 11:59pm PST_*

**WEEK 3 [Project goal: propose a theory & locate at least one variable]**

**Mon. April 15 – Lecture 7: Causality**
- Correlation versus causation
- Counterfactuals & inference errors

**Wed. April 17 – Lecture 8: Data & R Workshop**
- Finding data
- Cleaning data in R

**Fri. April 19 – Lecture 9: Research in the Wild**
Video: *How India runs the world’s biggest election*
- Random assignment & experimentation

*_Research in the Wild Option #3 due 4/19 at 11:59pm PST_*
**WEEK 4** [Project goal: specify testable implication, have both variables downloaded]

**Mon. April 22 – Lecture 10: Sampling & Surveys**
- Samples and sample statistics
- Selection effects, representativeness, and randomization, survey challenges

**Wed. April 24 – Lecture 11: Distributions & Visualizing Distributions Using R**
- Introduction to probability distributions
- Use R to simulate distributions and visualize and summarize data

**Fri. April 26 – Lecture 12: Research in the Wild**
Video: *Is the US Headed Towards Another Civil War?*

*Homework #1 due 4/26 at 11:59pm PST*
*Research in the Wild Option #4 due 4/26 at 11:59pm PST*

---

**WEEK 5** [Project goal: have both variables cleaned & merged]

**Mon. April 29 – Lecture 13: Hypothesis Testing I**
- Normal distributions and z-scores
- Sampling distributions and standard errors
- The central limit theorem and t-scores

**Wed. May 1 – Lecture 14: Hypothesis Testing II**
- Null and alternative hypotheses
- Statistical significance, confidence intervals, and type-I / type-II errors
- Chi-square tests, difference of means, difference of proportions tests

**Fri. May 3 – Lecture 15: Research in the Wild & Merging Data in R**
Article: *The Effect of Gender on Interruptions at Congressional Hearings* (Miller & Sutherland, 2023) also available on Canvas
- Merging data in R and visualizing bivariate relationships

*Research in the Wild Option #5 due 5/3 at 11:59pm PST*

---

**WEEK 6** [Project goal: create data visualizations]

**Mon. May 6 – Lecture 16: Regression I**
- Regression and multiple regression
- Interpreting regression coefficients & goodness of fit

**Wed. May 8 – Lecture 17: Regression II & Regression Diagnostics in R**
- Regression assumptions & diagnostics
- Robust standard errors

**Fri. May 10 – Lecture 18: Research in the Wild**
Video: *Why 99% of ocean plastic pollution is “missing”*
- Regression limitations in our own research

*Homework #2 due 5/10 at 11:59pm PST*
*Research in the Wild Option #6 due 5/10 at 11:59pm PST*
**WEEK 7** [Project goal: conduct regression analysis]

**Mon. May 13 – Lecture 19: Research Design Specifics**
- Interaction effects, fixed effects

**Wed. May 15 – Lecture 20: Practical Research Design & Formatting Regression Tables in R**
- Troubleshoot specific issues setting up research designs in R
- Stargazer and other table formatting options

**Fri. May 17 – Lecture 21: Research in the Wild**
Article: *Blood Avocados: Trade Liberalization and Cartel Violence in Mexico* (Erickson & Owen, 2022) also available on Canvas

*Research in the Wild Option #7 due 5/17 at 11:59pm PST*

---

**WEEK 8** [Project goal: format research findings & prepare for presentations]

**Mon. May 20 – Lecture 22: Causal Inference Approaches**
- Matching
- Regression Discontinuity Design
- Difference-in-Difference
- Instrumental Variables
- Selection Models

**Wed. May 22 – Lecture 23: Special Topics**
- Sample final project presentations from Austin Beacham and Bertrand Wilden

**Fri. May 24 – Lecture 24: Research in the Wild & Intro to Spatial Data**
Video: *Why buying a house in the US is so hard right now*
Article: *Women politicians reduce violence against women: Evidence from Mexico* (Alcocer, Skillman & Torres-Beltran, 2023) also available on Canvas
- Spatial data & autocorrelation

*Homework #3 due 5/24 at 11:59pm PST*

*Research in the Wild Option #8 & 9 due 5/24 at 11:59pm PST*

---

**WEEK 9** [Project goal: present research & receive feedback]

**Mon. May 27 – Lecture 25: Final Project Workshopping Session**

**Wed. May 29 – Lecture 26: Final Project Presentations**

**Fri. May 31 – Lecture 27: Final Project Presentations**

---

**WEEK 10** [Project goal: present research & receive feedback]

**Mon. June 3 – Lecture 28: Final Project Presentations**

**Wed. June 6 – Lecture 29: Final Project Presentations**

**Fri. June 7 – Lecture 30: Final Project Presentations**
*Comment on final presentation due 6/7 at 11:59pm PST*

---

*Final Projects (paper + presentation) due Wednesday 6/12 at 11:59pm PST*