POLS 361: Data Analysis in Political Science

Cal Poly, Dept. of Political Science
Spring 2022

Meeting time:	M/W 4:10-6:00 PM
Meeting place:	Kennedy Library 0217D;

Instructor:	Dr. Kevin Simmt
Email:	`ksimmt@calpoly.edu`
Phone:	N/A
Office:	TBD
Office hours:	Anticipated: M 2:00-4:00
	Or during other times set up through email

Contents

Course Description

Many, if not most, of the major debates in modern political science revolve around questions that can be addressed with data. The sources of voting behavior, the correlates of war, the determinants of development, political economy, psychology, institutions, and conflict—all are issues that are amenable to data-based analysis.

At the same time, the amount of available data and the number of publicly-available open-source tools for cleaning, transforming, analyzing and visualizing it have increased exponentially since the turn of the millennium. With a few clicks students can compare word frequencies in books over time or construct elaborate size-weighted wordclouds—tasks that would have taken scholars weeks if not months of effort in the past.

This course introduces students to those tools and the principles behind their use in the context of applications in political science. It marries the substance of political theory to the methodologies of data visualization and exploratory data analysis. It neither requires nor imparts any statistical background: it is designed to serve either as a standalone course or as a gateway to a more advanced data-analytics class.

Special Credit is due to OSU's Professor Bear Braumoeller and Ryan Dawe for the course content. Additional Credit is due to CalPoly's Professor Nancy Arrington for syllabus construction. Errors or inconsistencies are my own.

Course Goals and Learning Objectives

Goals

Students develop skills in drawing conclusions and critically evaluating results based on data.

Guide students to understand basic concepts of statistics and probability, comprehend methods needed to analyze and critically evaluate statistical arguments, and recognize the importance of statistical ideas.

We will tie elements of quantitative analysis and computational statistics together with the logic of research design in order to give students the ability to evaluate statistical arguments, and show them examples of how to do so using data on practical applications within the field of political science.

Grading

Assessment of Expected Learning Outcomes

The effectiveness of this course in achieving the expected learning outcomes outlined above will be determined in three ways.

1. A problem set will be assigned in-class, every few weeks, to help students internalize the material delivered in the online lectures. The problem sets will be graded, students' mastery of the relevant skills will be assessed, and they will be informed of problem areas, if any.

2. During the last two weeks of class, students will work on an independent project of their choosing, one that involves answering questions using the skills they have acquired. This final assignment will count for 40% of the course grade.
3. Students will be urged to fill in the online Student Evaluation of Instruction (SEI) reports to assess the amount that they have learned in this course relative to others.

Course Grade Expectations

A numeric weighted average of grades on course assignments will translate to final letter grades according to the thresholds below. Note: cutoffs are at the full point level rather than the half point level. I will not "bump-up" or "round-up" final student grades; an 89.99 is a B+. However, if you think a grade has been miscalculated or mis-entered, please let me know ASAP so I can check and make a correction.

- A (93.0-100%) Exceptional Performance
- A- (90.0-92.99%) Very Good Performance
- B+ (87.0-89.99%) Good Performance
- B (83.0-86.99%) Good Performance
- B- (80.0-82.99%) Satisfactory Performance
- C+ (77.0-79.99%) / C (73.0-76.99%) / C-(70-72.99%) Adequate Performance
- D+ (70.0-69.99) /D(65.0-66.49%) Minimal Passing Performance
- F (Below 65.0%) Unacceptable Performance

Grade Weights

There will not be a traditional midterm or final exam. Rather, five sets of assignments will make up 60% of the grade, and the remaining 40% will come from a final project in which the student finds a dataset in his or her area of interest that is not already used in the course, analyzes it to assess the structure of the data, and works through the most appropriate, succinct, and informative summaries and visualizations. Students will be given the last two weeks of recitation sections to work in–class on final projects, with the instructor present to assist. The final course grade will be determined based on the following breakdown and grading scale:

Problem Sets	60%
Final Project	40%

Problem Sets

Assignments will be posted on Carmen prior to lecture and will be **due one week from the day assigned by 11:59 PM**. Assignments can be completed in groups of up to three people. The grade assigned to a group project will be assigned to each individual in the group. The final project will be an **individual** project, and will be based on many of the skills obtained in the group projects, so there is little sense in joining a group and not being an active contributor.

Final Project

The final project is designed to be the foundation of an independent research project that could be used as the basis of a Major Research Project or Undergraduate Thesis. This 15-20 page essay will establish a student's ability to set-up a convincing research design and to produce meaningful data analysis, including visualizations.

Course Materials

Video Lectures

Incredibly helpful supplementary lectures exist online, made by the Bear Braumoeller (to whom I previously served as Teaching Assistant for the "Data Literacy" course at the Ohio State University). He kindly has made these valuable materials freely available for the public. You should watch these videos *in compliment to* attending my lectures. To some extent, they take the place of readings in this course, of which there are relatively few.

The lectures are online and can be accessed via YouTube here. **You should listen to and view the lectures prior to class.**

The lectures are recorded in Quicktime format, which should be viewable on nearly any computer or mobile device. One advantage to the format is that, on most devices, lectures can be sped up to 1.4x to 1.5x; beyond about 2x, most of them become unintelligible. Students are advised to take advantage of this feature, as humans can generally understand speech at a higher rate than they can produce it. Be advised, though, that higher speeds generally require more focused attention, as important details are easier to miss. It's also possible to rewind and slow the recording back down if a particular section moves quickly or is difficult to understand.

Recommended Texts

Two books are recommended for additional reference. All are available either as a Kindle book, which can be read on Kindle tablet or PC apps, or as an online resource (see links below). Note that you do not likely need to purchase the texts, as online resource should suffice.

1. Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton: Princeton University Press. (Most of the book and lecture notes are available free online)
2. Wickham, Hadley and Garrett Grolemund. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol, California: O'Reilly Media. (Most of the book and lecture notes are available free online)

Software

In this course, we will use the statistical software called R (see `https://www.r-project.org`). We will also use RStudio (see `https://www.rstudio.com`), a text editor that will make it easy for you to interact with R. Both R and RStudio are free, and you should download each on your own computer (Mac or PC). Versions will also be available on university computers.

What is R and why use it?

- Widely-used in academia and industries
- Open-source and free
- Power and flexibility
- Graphical capabilities
- Learning R = learning basic programming

The *New York Times* described R as

> a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca [...] whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it. [...] "The great beauty of R is that you can modify it to do all sorts of things," said Hal Varian, chief economist at Google. "And you have a lot of prepackaged stuff that's already available, so you're standing on the shoulders of giants."

Resources for Using R

Students invariably find learning R to be the most challenging part of this course. It is also, in the long run, the most rewarding, in that it's a valuable and marketable skill that also prepares you for more advanced classes. The learning curve is very steep, however, and it is best to be prepared. Toward that end, students should take advantage of the following:

- *Attend In-Class Lecture.* Although attendance is not officially monitored, the best way to succeed in this course is by coming to the bi-weekly class sessions. During these sessions, your instructors will give helpful tips for completing assignments, and will be available to go over material and answer any questions about the lectures and homework. If you have really been struggling with a problem set, you can even attend more than one session per week. Take advantage of these times to work on assignments and get help with course material. ***Priority during office hours and over email will be given to students who regularly attend recitation.***
- *Participate in Canvas Discussions.* Canvas is the online communications hub for the course. It is where you will go to post questions about assignments, and where you can answer questions as well. If you have a question about an assignment, lecture or how to use R, chances are that others will have the same question. In many instances, you will find that your question has already been asked (and answered!) by your peers. Note that Instructors may not be monitoring the discussions on Canvas everyday.
- *Google is your friend.* Simply google your coding question, such as "how do I make a bar chart in R," and you are bound to find many useful resources on the web.

- *Statistical consulting at PRISM.* The Department of Political Science offers help through the Program In Statistics and Methodology (PRISM). You may contact the PRISM Fellows to setup a brief appointment for assistance.

- *Online tutorials and books.* You should download Shidan Murphy's excellent `R` Instructor app for iPhone or Android and bookmark, or just download and print out, Tom Short's fantastic `R` Reference Card. You may also wish to consider Paul Teetor's *`R` Cookbook* from O'Reilly. You might also consider the following book: Hrishi, Mittal. *`R` Graph Cookbook* (Kindle edition; Packt Publishing)

Course Policies

Office Hours

Office hours are times during the week set aside for me to meet with students about any questions or concerns they have about the course. I intend to hold a combination of in-person and virtual (via Zoom) office hours during the times listed at the top of this document and by appointment. (PLEASE NOTE: UPDATE PENDING – I am yet to be assigned office space.) If you are in class or scheduled to work during those times, that's A-ok, just send me an email with your availability and we will find a time where we are both free.

Late Work

I honestly don't know which late-work policy is the best combination of ethical, practical, and pedagogically effective in a covid world. In my experience, deadlines with penalties are a great way to encourage students to stay on track with their work (y'all are generally a grade-motivated bunch!), but I do not want to penalize students for things outside of their control nor do I want to put myself in the position of deciding whose illnesses and life issues are important enough to justify an "excused" label. So, y'all are the guinea pigs for a new policy: **I will automatically issue every student 10 days (240 hours) of "free" late penalities**. *Any student exceeding their cap will start to be penalized 3% per day.* **These free days cannot be used past the last day of class, DATE, nor for the final project.** As of DATE, they are null and void (I have to submit final grades, after all.)

I've set up deadlines that will make sure you are on-track to complete a solid final project at the end of the course. I ***STRONGLY*** recommend submitting work by the deadlines throughout the quarter, but if something out of your control happens and you need another day or two, go for it. If you need to exceed your 10 allowance days, then YOU MUST PROVIDE VALID DOCUMENTATION demonstrating the need for more than 10 excused days. You cannot use late days beyond the final day of class.

Some qualifications:

- The main exception is your final paper. (Because my deadline for submitting grades is firm.)

- All late work from the Assignments must be submitted by the last day of classes: Wednesday June 1st at 11:59 PM. Any work not received by midnight on the 1st will get an automatic zero. This is REGARDLESS of whether you have remaining days "free of late penalties."

Note: all of the work in this course is designed to help you produce good final projects. If you don't do the work until the end of the quarter, it won't be helpful.

- You will get more timely – and therefore more useful – feedback (which is *very* important in this class), for work that is turned in by the stated deadline.
- Don't let the "free late days" lull you into a false sense of having extra time. The quarter moves fast. And, submitting late just means your not able to advance on your final project.

Incomplete Grades

If extenuating circumstances affect your ability to finish the course material by the end of the quarter, please schedule a meeting with me as soon as possible to discuss whether taking an incomplete for the course is a good option for you. Incomplete grades will only be given for extenuating circumstances (simply falling behind because you're busy is not sufficient for an incomplete[1]) and if an agreement is made **prior** to the end of the course. Once an Incomplete has been given, it is the student's responsibility to schedule meetings with me to facilitate the completion of the work.

Communication

In addition to announcements made in class, I will communicate with you through email. You are expected to check email regularly (daily, M-F). I check my email regularly during the work week – Monday through Friday, 9:00am-5:00pm – and respond as promptly as possible. Emails sent to me in the evenings and weekends *might* get a prompt(ish) response but probably won't. Emails sent in the middle of the night definitely aren't getting a prompt response (go to bed!).

Integrity of Scholarship

I take the integrity of scholarship very seriously in this course. By taking this course, you affirm that it is against course standards to cheat on exams, to plagiarize, to deviate from my instructions about collaboration on work submitted for grades, to give false information to a faculty member, and to undertake any other form of academic misconduct. You agree that I am entitled to move you to another seat during quizzes without explanation.

Any evidence of cheating/plagiarism will be punished with an automatic 0 for the assignment. Severe offenses (defined at the instructor's discretion) will result in an automatic failure of the course. All cases of cheating will be reported to the Office of Student Rights and Responsibilities, without exception.

For more information about Cal Poly's expectations for academic integrity see
`https://academicprograms.calpoly.edu/content/academicpolicies/Cheating`

The Department's Definition of Plagiarism "Plagiarism is defined as the act of using the ideas or work of another person or persons as if they were one's own without giving proper credit to the source. Such an act is not plagiarism if it is ascertained that the ideas were arrived through independent reasoning or logic or where the thought or idea is common knowledge.

[1] I'm not trying to be harsh here. I've tried giving out incompletes for students who've fallen behind for non-specific reasons, and not a single one of them has ever finished the work late. In those cases, all the incomplete does is prevent students from re-registering for the course/taking the next courses, which ultimately hurts progress towards graduation.

Acknowledgement of an original author or source must be made through appropriate references; i.e., quotation marks, footnotes, or commentary. Examples of plagiarism include, but are not limited to the following: the submission of a work, either in part or in whole completed by another; failure to give credit for ideas, statements, facts or conclusions which rightfully belong to another; failure to use quotation marks (or other means of setting apart, such as the use of indentation or a different font size) when quoting directly from another, whether it be a paragraph, a sentence, or even a part thereof; close and lengthy paraphrasing of another's writing without credit or originality; use of another's project or programs or part thereof without giving credit." `https://politicalscience.calpoly.edu/students/advising/policy`

The Department's Definition of Cheating "Cheating is defined as obtaining or attempting to obtain, or aiding another to obtain credit for work, or any improvement in evaluation of performance, by any dishonest or deceptive means. Cheating includes, but is not limited to: lying; copying from another's test or examination; discussion at any time of answers or questions on an examination or test, unless such discussion is specifically authorized by the instructor; taking or receiving copies of an exam without the permission of the instructor; using or displaying notes, "cheat sheets," or other information devices inappropriate to the prescribed test conditions; allowing someone other than the officially enrolled student to represent same." `https://politicalscience.calpoly.edu/students/advising/policy`

Access, Disability Services, and Resources

I am committed to providing appropriate accommodations. If you have a disability-related need please contact the Disability Resource Center at Cal Poly immediately: `http://drc.calpoly.edu/content/drc-services`, (805) 756-1395. If you need immediate accommodations or physical access, please let me know. You can email me, come to office hours, or schedule a one-on-one meeting. I want everyone to learn and will do what I can to facilitate that process.

Diversity and Inclusion

My goal is to create a learning environment that enables all students to succeed and know that their experiences are valid. I expect students to treat each other with kindness and civility. For more information on resources related to diversity and inclusion, please visit the Office of University Diversity & Inclusivity website at `http://diversity.calpoly.edu`. If you notice behaviors or institutions in our class that undermine or obscure diversity and inclusion, please share those concerns with me (if you are comfortable doing so) or the department chair, Dr. Anika Leithner `aleithne@calpoly.edu`

Needs Support

If you face challenges securing food, housing, or other human needs know that you are not alone and Cal Poly can help. You can find resources available to support you through Cal Poly's Basic Needs Initiative at `https://www2.calstate.edu/impact-of-the-csu/student-success/basic-needs-initiative` and information about the Food Pantry and Meal Vouchers at at `https://studentaffairs.calpoly.edu/hunger`

Zoom Etiquette

While primarily an in-person class, we shall, on occasion, conduct meetings over Zoom. Most of you are familiar with the zoom etiquette norms at this point, but a few reminders:

- Mute yourself unless you are actively about to ask or respond to a question to minimize background noise.
- When your video is on, please minimize distractions for the other students. Life happens when we take classes from home, but – if possible – point your camera to minimize distractions behind you. Turn off any TV screens in the background, and temporarily turn your video off if you need to get up or do something other than listen and participate in the class (eating a meal, for example)
- When your environment has minimal distractions and you feel safe sharing it, please keep your video on. Having your video on is useful for a few reasons – having your video on is an incentive for you to stay focused on the course content, and having your video on is useful to me/your instructor (being able to see faces gives good feedback about whether or not what we are discussing makes sense).
- If you'd like, you're welcome to use that chat function to talk to other students about the content of the class, but move any extra-curricular conversations out of zoom (and after class).

Class Schedule

Any changes to the schedule below will be announced in class as soon as possible and an updated syllabus will be emailed and posted to Canvas.

Please recall, each class-day cites supplementary "Lecture X, Y, Z." These are online and can be accessed via YouTube here. **You should listen to and view the lectures prior to class.**

Session		Topics/Deadlines	Deadlines
Mon Mar 28	1	Syllabus and Policy + Starting with R	Read Syllabus & Submit Questions
Wed Mar 30	2	Lectures 1, 2 & 3	
Mon Apr 4	3	Lectures 4, 5, & 6	
Wed Apr 6	4	Lectures 1 - 7 Review!	Assignment 1
Mon Apr 11	5	Lecture 8	
Wed Apr 13	6	(I/III) Asynchronous Review of Topics Applied to Research Project; **Online** **Group Work**	
Mon Apr 18	7	Lectures 9, 10, & 12	
Wed Apr 20	8	More R	Assignment 2
Mon Apr 25	9	(II/III) Asynchronous Review of Topics Applied to Research Project; **Online** **Group Work**	
Wed Apr 27	10	Lectures 11 & 13	
Mon May 2	11	Theories & Hypotheses	
Wed May 4	12	Lectures 14, 15, & 16	Assignment 3
Mon May 9	13	Lecture 17 & 19	

Session		Topics/Deadlines	Deadlines
Wed May 11	14	(III/III) Asynchronous Review of Topics Applied to Research Project; **Online** **Group Work**	Assignment 4
Mon May 16	15	Lecture 18, 20, & 21	
Wed May 18	16	Research Design Recap + Work Day/Research Workshop	Assignment 5
Mon May 23	17	Group Work; Asynchronous Materials; **Online**	
Wed May 25	18	Group Work; Asynchronous Materials; **Online**	
Mon May 30	19	Memorial Day	
Wed Jun 1	20	Presentations	
Mon Jun 6			Final papers due by 11:59 pm

Description of Lectures

Recall, as per the earlier section, that online lectures have been made by the Bear Braumoeller, for whom I previously served as Teaching Assistant. He kindly has made these valuable materials freely available for the public.

The lectures are online and can be accessed via YouTube here. **You should listen to and view the lectures prior to class.**

Lecture 1: Introduction. Nature of the course, its relationship to other courses and students' plans of study; how and why data visualization can be useful.

Lecture 2: Data are All Around Us! Terminology, sources, and advice on what to do if you can't find a preexisting dataset.

Lecture 3: Data Tools, part I. Open-source solutions that require little or no additional information or effort to produce compelling results.

Lecture 4: The Good. Exemplary representations of data from our field and others, with a discussion of what makes them so compelling.

Lecture 5: The Bad and The Ugly. Ideas that seemed good at the time, why they really aren't, how we can fix them, and what general principles they imply.

Lecture 6: Data Tools, part II. These versatile tools allow you to upload, explore, and visualize your own data.

Lecture 7: Truth. Data visualization is inherently a reduction of information. The main imperative, in reducing information, is integrity: representing the data as honestly as possible. We look at some good cases and some not-so-good cases.

Lecture 8: Introduction to `R`. A powerful (and free) general statistical package, `R` is capable of producing a wide range of graphs. In these lectures we demonstrate how to download and install the package, add libraries, import data, and create visualizations.

Lecture 9: Beauty. The difference between a fairly good data visualization and an amazing one often lies in the application of a few straightforward graphical principles.

Lecture 10: `R`, part II. Data structures and data transformations.

Lecture 11: Time. Different ways of examining variables over time.

Lecture 12: Plots in `R`. Scatterplots, line graphs, boxplots, dot charts, and more.

Lecture 13: APIs. The application program interface: a wonderful way for people to make their data available to you.

Lecture 14: Maps. Less common, but more eye-catching, are ways of looking for patterns across space, especially in maps.

Lecture 15: Careers in Data I. Aaron Schwartz describes his work at Community Research Partners.

Lecture 16: Space and Time. Representing relationships and trends over both space and time.

Lecture 17: Research Design. Designs for causal inference using observational data, including natural experiments, regression discontinuity, and differences in differences.

Lecture 18: Careers in Data II. Jill Riepenhoff and Mike Wagner discuss their work at the Columbus *Dispatch*.

Lecture 19: Simulated Worlds Using simulation to predict the outcome of the 2012 Presidential election.

Lecture 20: Workshop—Partisanship. Is American politics more partisan than it was 50 years ago? If so, why?

Lecture 21: Simulations and Inference Monte Carlo techniques for understanding uncertainty.

Signature Sheet

Please sign below. (in ink or e-sign) and submit this page to canvas on to indicate that you have read and generally understand the course policies outlined in this document:

Name: _____

Signature: _____

What questions do you have about the class that are not answered in the syllabus/ what policies are unclear?

Do you have access to an electronic device that you can use to work on my project in class? (It is totally OK if you do not, I just like to know so we can best make use of in-class work time.)

If professors/folks/people-with-names-like-Nancy tend to mispronounce your name, can you provide a phonetic spelling below to help me get it right?

If you'd like to share, what pronouns should I use to refer to you in class?

In addition, is there anything else that you would like to share with me?