

# Political Science 150A/355A: Data Science for Politics (Summer Quarter, 2021) Tuesdays and Thursdays, 10:30–11:50am PT

## Teaching Staff

**Professor:** Yiqing Xu

[yiqingxu@stanford.edu](mailto:yiqingxu@stanford.edu)

Office Hours: Tuesday 2-4pm PT on Zoom (please sign up below first)

<https://yiqingxu.youcanbook.me>

**TA:** Valentín Figueroa

[valfig@stanford.edu](mailto:valfig@stanford.edu)

Office Hours: Monday 8:30am-11:30am PT

<https://valentinfoigeroa.youcanbook.me>

**TA:** Aliz Tóth

[aliztoth@stanford.edu](mailto:aliztoth@stanford.edu)

Office Hours: Thursday 2:30-5:30pm PT

<https://aliztoth.youcanbook.me/>

***Overview.*** Data science is quickly changing the way we understand and engage in politics, how we implement policy, and how organizations across the world make decisions. In this course, we will learn the fundamental tools of data science and apply them to a wide range of political and policy oriented questions. How do we predict presidential elections? How can we guess who wrote each of the Federalist Papers? Do countries become less democratic when leaders are assassinated? These are just a few of the questions we will work on in the course.

Students are not expected to have any programming knowledge, and the course will be centered around bite-size assignments that will help build R coding and statistical skills from scratch. Students will leave the course equipped for work in any setting that requires a social scientific approach to data science, from policy non-profits to government, from Silicon Valley to Wall Street and beyond. There are no formal prerequisites for the course, but introductory statistics is recommended.

***Learning Goals.*** The course has three basic learning goals for students. At the end of this course, students should:

1. Be comfortable using basic features of the R programming language.
2. Be able to combine political data with statistical concepts to answer political questions.
3. Know how to create visual depictions of statistical patterns in data.

***Learning Approach.*** Statistical and programming concepts do not lend themselves to the traditional lecture format, and in general, experimental research on teaching methods shows that combining active learning with lectures outperforms traditional lecturing (see for example: [http:](http://)

[//doi.org/10.1126/science.1201783](https://doi.org/10.1126/science.1201783)). We will teach each concept in lecture using applied examples that encourage active learning. Lecture will be broken up into small modules; first, I will explain a concept, and then we will write code to implement the concept in practice. Students are asked to bring their laptops to class so that we can actively code during lecture. This will help students “learn by doing” and it will ensure that the transition from lecture to problem sets is smooth.

## Course Websites

In this course, we will utilize an online discussion board called *Piazza* as the course website. We will use this site to ask questions and also to provide homework assignments, datasets, and links to reading materials. Below is an official blurb from the Piazza team:

Piazza is a question-and-answer platform specifically designed to get you answers fast. They support LaTeX, code formatting, embedding of images, and attaching of files. The quicker you begin asking questions on Piazza (rather than via individual emails to a classmate or one of us), the quicker you’ll benefit from the collective knowledge of your classmates and instructors. We encourage you to ask questions when you’re struggling to understand a concept.

In addition to sections and office hours, please use the Piazza Q & A board when asking questions about lectures, problem sets, and other course materials. You can access the Piazza course page at:

<https://piazza.com/stanford/summer2021/polisci150a>

Using Piazza will allow students to see other students’ questions and learn from them. Both the TA and the instructor will regularly check the board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. A student’s respectful and constructive participation on the forum will count toward his/her class participation grade. *Do not email your questions directly to the instructors or TAs* (unless they are of a personal nature).

## Required Text

*Quantitative Social Science: An Introduction*, by Kosuke Imai (Princeton University Press) (ISBN-10: 0691175462)

## Evaluation

40% Problem Sets (10% each)  
35% Midterm Exams (15% and 20%, respectively)  
25% Final Exam  
3% Bonus Problem Set

There will be **four** problem sets drawn from the textbook and from other course materials, each focusing on applying the concepts covered in lecture and in lab to new datasets in R.

- **Problem sets will be submitted electronically to Canvas before the due time.**

- With the exception of extraordinary circumstances, late submission will be penalized with 0.5 point of your final total grade for each day of late submission. In other words, submitting your homework 20 days after the due date will not earn you any points for that problem set in your final grade. We will strictly enforce this rule.
- Problem sets will be graded for three criteria: correct discussion of concepts; correct output of statistical code; and code “style,” meaning how well commented and explained the submitted code is.
- Working in groups is encouraged for conceptual and sometimes technical discussion, but each student must submit their own writeup of the solutions that shows their independent work on the assignment. In particular, you should not copy someone else’s answers or computer code. We also ask you to write down the names of the other students with whom you solved the problems together on the first sheet of your solutions.

The two midterm exams will be administered in class.

The final exam will be cumulative and will test knowledge developed throughout the course. There will be a final exam review session conducted in lecture on the final day of class. The final exam will take place on **Tuesday, November 19th during class**.

Class participation is awarded based on involvement in lecture and on the online discussion site.

## **Students with Documented Disabilities**

Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is being made. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. The OAE is located at 563 Salvatierra Walk (phone: 723-1066, URL: <http://studentaffairs.stanford.edu/oea>).

## **Financial Resources for Students**

Stanford University and its faculty are committed to ensuring that all courses are financially accessible to all students. This course requires a *laptop* and a *textbook*. If you cannot afford them, the Diversity & First-Gen Office may be able to provide assistance. To inquire, please complete their questionnaire on course textbooks & supplies: <http://tinyurl.com/jpqbarn> or contact Joseph Brown, the Associate Director of the Diversity and First-Gen Office ([jlbrown@stanford.edu](mailto:jlbrown@stanford.edu); Old Union Room 207). Dr. Brown is available to connect you with resources and support while ensuring your privacy.

## **Policy Concerning Exam Scheduling**

Rescheduling of midterm and final exams is usually not allowed without a signed doctor’s note or similar written proof of extraordinary circumstances. When such events happen, please consult with the teaching staff as soon as possible such that we can take alternative measures to compensate for the grade.

## **Policy Concerning Student Athletes**

Student athletes at Stanford compete in national and international arenas at the highest level. Per university policies, we will accommodate schedule conflicts for student athletes who participate in competitions (not applied to practices). When a potential conflict arises, student athletes should discuss with the instructor on alternative arrangements at least *two weeks* prior to the conflict.

## **Honor Code and Collaboration Policy**

Students are expected to adhere to the Stanford Honor Code (<http://studentaffairs.stanford.edu/communitystandards/policy/honor-code>) at all times. Collaboration is encouraged on problem sets, but should be limited to two students, in which case you *must identify the student with whom you collaborate at the top of their submitted problem set*.

For the coding portions of the problem sets, “plagiarism” is defined as copying computer code from any source other than the teaching materials without citation. Writing code can be difficult, especially with course deadlines and multiple demands on one’s time. In difficult situations, when code refuses to do what the student wants it to, and when a deadline is looming, it can be tempting to copy and paste code from other sources. Students should be aware that it is incredibly easy to detect code plagiarism using automated techniques. Please start problem sets early and contact the teaching staff about code difficulties to avoid these situations.

## **Schedule (Subject to Change as Quarter Progresses)**

Note: Class meets twice a week, and each class is one hour and 20 minutes long. Because these lectures are long, we will have a scheduled break halfway through lecture each time.

A tentative session-by-session schedule is below.

**Tentative Course Plan (subject to change)**

Day	Topic	Application	Readings	
Tuesday, 6/22:	Course Introduction	—	—	
Thursday, 6/24:	What is Data/Intro to R		1	
Tuesday, 6/29:	Intro to RMarkdown	<i>Wealth of nations</i>	—	
Thursday, 7/1:	Random variables	<i>Wealth of nations</i>	6.3	(PSet 0 Due)
Tuesday, 7/6:	Causality	<i>Racial discrimination</i>	2.1-2.4	
Thursday, 7/8:	Review for Midterm I	<i>Campaign finance</i>	2.6	(PSet 1 Due)
Tuesday, 7/13:	<b>Midterm I (in class)</b>			
Thursday, 7/15:	Selection bias		2.5	
Tuesday, 7/20:	Prediction	<i>Presidential election</i>	4.1	
Thursday, 7/22:	Prediction	<i>Presidential election</i>	4.1	
Tuesday, 7/27:	Regression, Part 1	<i>Candidate appearance</i>	4.2	(PSet 2 Due after class)
Thursday, 7/29:	Regression, Part 2		4.2	
Tuesday, 8/3:	Review for Midterm II	Assassinations and democracy		(PSet 3 Due after class)
Thursday, 8/5:	<b>Midterm II (in class)</b>			
Tuesday, 8/10:	Regression and Causation	<i>Social pressure and turnout</i>	4.3	
Thursday, 8/12:	Regression and Uncertainty	<i>Monitoring Corruption</i>	4.3	
Tuesday, 8/17:	Prediction Errors	<i>Gov. Responsiveness in China</i>	7.3	(Bonus Pset Due)
Thursday, 8/19:	Hypothesis Testing		7.3	
Tuesday, 8/24:	Putting It All Together		—	
Tuesday, 8/26:	<b>Final Exam (in class)</b>		—	(PSet 4 Due.)