Political Science 150A/355A:
Data Science for Politics
(Fall Quarter, 2016)
Tuesdays and Thursdays, 9:30–11:20am
Hewlett 201

Teaching Staff

Professor Andrew Hall
andrewbhall@stanford.edu
Office Hours: Mondays, 9-11am
Encina Hall West 412

Teaching Assistant: Chloe Wang
cnwhang@stanford.edu
Tuesdays, 2:30–4:30pm
Encina Hall 416

Teaching Assistant: Scott Williamson
scottw3@stanford.edu
Office Hours: Fridays, 1:30–3:30pm
Encina Hall 317

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://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

The dedicated course website is located at `https://canvas.stanford.edu/courses/49249`.

We will use the discussion forum on Canvas in lieu of Piazza. *Please ask all course-related questions on Canvas rather than via email.* This way all students will be able to benefit from seeing the question and answer.

Required Text

A First Course in Quantitative Social Science, by Kosuke Imai (Princeton University Press)

The required textbook is forthcoming at Princeton University Press. PUP and the author have kindly given us permission to use the textbook free of charge in advance of its official release. The PDF of the book will be made available on the course website. Please repay their generosity by not circulating it!

Evaluation

40% Problem Sets
25% Two Midterm Exams (12.5% each)
30% Final Exam
5% Class Participation

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. Students are encouraged to collaborate but will be responsible for writing their own R code, producing their own statistical output, and writing their own interpretations of what the resulting estimates mean from a substantive standpoint.

Problem sets will be submitted digitally to `data.science.for.politics@gmail.com` AND in hard copy in lecture, and 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. Problem sets will not be accepted after the time at which they are due.

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 **Wednesday, December 14th from 8:30–11:30am**.

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/oae).

Financial Resources for Students

Stanford University and its faculty are committed to ensuring that all courses are financially accessible to all students. 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`; Old Union Room 207). Dr. Brown is available to connect you with resources and support while ensuring your privacy. [note: this course only requires a free textbook, but access to computing resources is essential].

Policy Concerning Exam Scheduling

Rescheduling of midterm and final exams is not allowed without a signed doctor's note or similar written proof of extraordinary circumstances. There are absolutely no exceptions to this policy. Please do not contact the teaching staff about rescheduling an exam unless you fall under this exception and have already prepared the necessary documentation.

Policy Concerning Homework Submissions

Students have many conflicting demands on their time. Sometimes, it can be difficult to let go of a hard problem set question even though "triage" principles suggest it would be better for the student to move on to other tasks. To prevent this problem, **late problem sets are not accepted**. Just hand in whatever you have by the deadline. There will be ample opportunities for partial credit.

Homeworks submitted with incorrect or unopenable attachments will be given a 0, unless the issue is corrected before the due date and time. **It is the student's responsibility to ensure that the correct materials are attached to the submission.** To be clear, a student can make as many electronic submissions as she pleases up until the deadline (so, for example, one can email in to correct a faulty attachment as long as this correction occurs before the deadline). Students must also submit a hard copy of their problem sets in lecture; **the hard copy must be identical to what is emailed to the teaching staff.**

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 students *must identify the students with whom they collaborated 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 fifty minutes long. Because these lectures are long, we will have a scheduled break halfway through lecture each time.

A tentative day-by-day schedule is below.

Tentative Course Plan (subject to change)

Day	Topic	Application	Readings (date=completion date)
Tuesday, 9/27:	Course Introduction	—	—
Thursday, 9/29:	Intro to R	*Campaign Finance*	1
Tuesday, 10/4:	Causality	*Assassinations*	2.1–2.5
Thursday, 10/6:	Describing a Single Variable	*Wartime Casualties*	2.6, 3.1–3.3
Tuesday, 10/11:	Relating Variables to Each Other	*Roll-Call Voting*	3.4–3.7
Thursday, 10/13:	Prediction	*Predicting Elections*	4.1
Tuesday, 10/18:	**Midterm I (in class)**		
Thursday, 10/20:	Regression, Part 1	*Candidate Appearance*	4.2
Tuesday, 10/25:	Regression, Part 2	*Presidential Elections*	4.2, 4.3
Thursday, 10/27:	Regression, Part 3	*Firm Regulation*	4.2, 4.3
Tuesday, 11/1:	Regression and Causation	*Effect of Women Politicians in India*	4.3
Thursday, 11/3:	**Midterm II (in class)**		
Tuesday, 11/8:	Text as Data	*Federalist Papers*	5.1
Thursday, 11/10:	Probability		6.1–6.6
Tuesday, 11/15:	Uncertainty of Estimation		7.1
Thursday, 11/17:	Hypothesis Testing		7.2
Tuesday, 11/22:	**Thanksgiving Break**		
Thursday, 11/24:	**Thanksgiving Break**		
Tuesday, 11/29:	Regression and Uncertainty		7.1–7.3
Thursday, 12/1:	Uncertainty: Review		7.3
Tuesday, 12/6:	Putting It All Together		—
Thursday, 12/8:	Final Review Session		—