POLI 172: Advanced Social Data Analytics

Summer Session II - 2024

Instructor: Jesús E. Rojas Venzor
Email: jrojasve@ucsd.edu
Office Hours: Calendly (Tu/Th 10am-12pm)
Class Hours: M/Tu/W/Th 6:30-7:50pm
Canvas: Course Link
Class Room: Zoom
TA: Roberto Bonilla (OH: F 1-2pm, Zoom)
TA Email: robonilla@ucsd.edu

COURSE DESCRIPTION

"One can learn data analysis only by doing, not by reading."

Data analysis is changing the way we engage with politics, policy, institutions, and information. This is an accelerated course in computer programming and data analytics meant for students to analyze and understand data in the social world. Students engage in hands-on learning with applied social science problems, while also developing statistical and computational skills for sophisticated data manipulation, analysis, and visualization. This course will take you through a survey of fundamental principles you need to answer a wide range of political and policy oriented questions. For example: who is most likely to win the upcoming presidential election? Do countries become less democratic when leaders are assassinated? Is there racial discrimination in the labor market? Do women introduce different policy than men to Congress? Does increasing police training decrease violent police confrontations? Are coastal states better prepared for climate change compared to inland states? These are just a few of the questions you will be able to tackle after taking this course.

PREREQUISITES

POLI 5 or 5D or ECON 5, and POLI 30 or 30D.

MATERIALS

We will utilize readings from *Data Analysis for Social Science: A Friendly and Practical Introduction* by Elena Llaudet and Kosuke Imai, which will be uploaded in the Canvas page. Please make sure to read before class time as I will use terminology and concepts from the readings.

We will also utilize R as our workhorse statistical software to run statistical models. The software is completely free and you will be required to install R and an additional platform known as RStudio. Please follow the instructions found at the *Hands-On Programming with R* website to install both of these. I encourage you to read part *B* through *E* from the same link so you familiarize with programming in R.

COURSE OBJECTIVES

The course is meant to teach you to understand your data, statistical models, and give you a sense of the issues and benefits of certain methods. I do not expect you to be an expert but rather provide you with an initial intuition about data analytics in social science and particularly in political science. More specifically, students will be able to understand quantitative data, explain its importance through data analysis, and reproduce their own analyses. Students will be able to explain concepts such as variables, constants, estimates, distributions, while also acknowledging the limitations of quantitative analysis. The course will also provide students with the language to continue learning data analysis.

COURSE STRUCTURE

There are no written midterms or finals in this class. Your course grade will be based on a combination of attendance (see *Attendance Policy* subsection below), extra credit (see *Extra Credit* subsection below), four weekly assignments, and a final project. Any assignment involving R must be submitted with accompanying code. All results should be presented so that they can be easily understood. Comment on your code or create subsections to make it much easier to read.

Assignments

This class will include 4 different weekly assignments due and a final project. Again, you will also be graded based on attendance (see *Attendance Policy* below), and will have the opportunity to submit a written assignment for extra credit. I recommend you begin the final project soon after its release halfway through the term.

Due Date: The assignments will be released every Monday at 12:00pm and will be by Saturday of their respective week, by 11:59pm. you will have approximately 5.5 days to complete each assignment.

Collaboration: Group work is allowed but students must turn in all of their work individually. In particular, students should not pass someone else's answers or computer code as their own. The student must write their full name and student ID in any assignment turned in. In addition, the student must write down the names of the other students with whom they discussed the assignments on the first sheet of their solutions.

ASSIGNMENT #1: Do Women Promote Different Policies than Men? Part 1

- The assignment covers general data manipulation, basics of coding, descriptive statistics, and interpretation.
- *Due: Saturday 8/10 by 11:59 PM*

ASSIGNMENT #2: Do Women Promote Different Policies than Men? Part 2

- The assignment covers correlations, visualizations, and sample selection.
- *Due: Saturday 8/17 by 11:59 PM*

ASSIGNMENT #3: Can We Predict Overall Class Scores?

- The assignment covers linear regression, OLS interpretation, prediction, and R^2 interpretation.
- *Due: Saturday 8/24 by 11:59 PM*

ASSIGNMENT #4: Does Leader Death Affect the Level of Democracy?

- The assignment observational studies, randomized experiments, confounders, multiple linear regression, and its interpretation.
- *Due: Saturday 8/31 by 11:59 PM*

Final Project

The final project requires you to pose a novel social science question and test it using data and the techniques learned in class. The goal of this assignment is to test your ability to understand the development of research, the components of a finely crafted study, and the tools available to answer a question. You will have in-class time to work on this project. The final project will be a 5-6 page (12pt. font, one-half spaced, not counting bibliography, not counting figures) paper and it should contain the following components in this order:

1. **Cover Page with Abstract:** 250-word single-spaced abstract with cover page including the title of your project, your name, the abstract, the course name, and any names of people you received help from in the form of acknowledgments. Examples will be provided.

2. **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? 1

    - *"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?"*

3. **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 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) informs your question and 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..."

4. **Data and 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."*
5. **Findings:** What did your analysis tell you about your research question and 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."*
6. **Contribution:** 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."*

Generally, ask yourselves these questions when completing the final project: what is your research topic? What are the key terms and their definitions/context? What are the disagreements on this topic? Why does this topic matter? What can we learn from this project?

- *Due: Thursday 9/05 by 11:59 PM*

Extra Credit

The extra credit assignment will consist of answering a series of questions regarding another scientific study ("Does Attack Advertising Demobilize the Electorate?" by Ansolabehere et al., 1994). There will be no coding involved but rather you will describe the study and its findings. Assignment will be available in Canvas. Value for the extra credit will be decided during the term based on class performance. I reserve the right to remove such extra credit or include other opportunities given class performance.

- *Due: Sunday 9/01 by 11:59 PM*

GRADING POLICY

Grades in undergraduate courses are defined as follows: A, excellent; B, good; C, fair; D, poor; F, fail; I, incomplete (work of passing quality but incomplete for good cause); and IP (In Progress). The designations P (Pass) and NP (Not Pass) are used in reporting grades for some undergraduate courses. P denotes a letter grade of C– or better. A blank grade indicates no record or report of grade was received from the instructor. W is recorded on the transcript indicating the student withdrew or dropped the course sometime after the beginning of the fifth week of a quarter. Instructors have the option of assigning plus (+) and minus (–) suffixes to the grades A, B, and C. I reserve the right to curve the scale dependent on overall class scores at the end of the course.

The grade will count the assessments using the following proportions:

- **60%** of your grade will be determined by 4 weekly assignments (15% each)
- **30%** of your grade will be determined by the final project
- **10%** of your grade will be determined by attendance

Please familiarize with our institution's Academic Regulations and Policies.

ONLINE SETTING

If the course must take place over Zoom or any other video conferencing service, lectures will be recorded for the safety of attendants and for future reviews to improve the course. Lecture recordings will not be made available to students unless instructed by the Office for Students with Disabilities. However, the instructor reserves the right to make them available for review if needed.

MEDIA RELATED TO THE COURSE

Media has been revolutionary in spreading ideas in a manageable manner. Hence, included here is media showcasing scenarios where course content is used for making decisions. I will mention them in class.

Films/Documentaries

- *The Big Short* (2015) - Free
- *Moneyball* (2012)
- *The Great Hack* (2019)
- *The Social Dilemma* (2020)
- *Brexit: The Uncivil War* (2019)

COURSE POLICIES

During Class

I encourage everyone to be active and attentive during class to better not only your experience but that of your classmates and instructor. The use of electronic devices for taking notes will be allowed in class. Please refrain from using such devices for anything but activities related to the class during the class time. Eating and drinking are allowed in class but please refrain from it affecting the course such as spills, trash, or general loud eating. If the class is virtual please keep your microphone muted during meals.

Attendance Policy

You are allowed 2 unexcused absences during the term. Any other absences will affect your grade unless they are excused by the instructor. Absences can also be accommodated where reasonable for any sincerely held religious belief, observance, or practice.

Missed/Late Assignments

Make-up exams, quizzes, and essays will only be given under valid, documented, and extreme circumstances. If you know you will not be able to complete an assignment, please let me know before class session through email. If you are not able to contact me in advance, you must contact me as soon as possible. Life gets in the way, I understand the complexities of being a student. I can work with you but it is your responsibility to contact me if you need additional support.

If an extended deadline is not authorized, a missed or late assignment will be given a zero. If an extended deadline is authorized but you fail to comply, I will treat it as a missed assignment.

Academic Integrity

Students are required to comply with UC San Diego Academic Integrity Policy. There is absolutely no tolerance for cheating in this course. Please do not cheat. The consequences are not worth it, ask for help when needed. Students agree that by taking this course any required assignments involving text can be subject to submission for textual similarity review to *Turnitin.com* for the detection of plagiarism. Please also abide by UCSD's Principles of Community and the Student Code of Conduct.

Student Accessibility

Students requesting 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. You are required to discuss accommodation arrangements with instructors and OSD liaisons in the department well in advance of any exams or assignments. The OSD Liaison for the Department of Political Science is Zain Sharifi; please connect with her via the Virtual Advising Center as soon as possible. I will be glad to meet with you privately during my office hours to discuss your special needs.

Discrimination on the basis of race, color, national origin, religion, sex, gender, gender identity, gender expression, pregnancy (including pregnancy, childbirth, and medical conditions related to pregnancy or childbirth), physical or mental disability, medical condition, genetic information, ancestry, marital status, age, sexual orientation, citizenship, or service in the uniformed services (including membership, application for membership, performance of service, application for service, or obligation for service in the uniformed services will not be tolerated. Retaliation against any person who complains about discrimination is also prohibited. If students have questions about student-related nondiscrimination policies or concerns about possible discrimination or harassment, they should contact the Office for the Prevention of Harassment & Discrimination (OPHD) at (858) 534-8298, `ophd@ucsd.edu`, or `reportbias.ucsd.edu`.

Religious Accommodation

It is the policy of the university to make reasonable efforts to accommodate students having bona fide religious conflicts with scheduled examinations by providing alternative times or methods to take such examinations. If a student anticipates that a scheduled examination will occur at a time at which his or her religious beliefs prohibit participation in the examination, the student must submit to the instructor a statement describing the nature of the religious conflict and specifying the days and times of conflict.

For final examinations, the statement must be submitted no later than the end of the second week of instruction of the quarter. For all other examinations, the statement must be submitted to the instructor as soon as possible after a particular examination date is scheduled. If a conflict with the student's religious beliefs does exist, I will attempt to provide an alternative, equitable examination that does not create undue hardship for the instructor or for the other students in the class. Please see the EPC Policies on Religious Accommodation, Final Exams, Midterm Exams

Additional Resources

- CARE at the Sexual Assault Resource Center
- Counseling and Psychological Services (CAPS)
- UCSD Basic Needs
- Writing Hub
- Supplemental Instruction
- Tutoring
- CAPS Student Health and Well-Being
- Community Centers
- Remote Access

Tentative Schedule and Learning Goals

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each class, and also as a study guide before each assignment, and at the end of the term. For any calendar-related questions refer to the Summer Session II Calendar.

———— WEEK 1 ————

Week 1 - 08/05

Class 1: Introduction

- **Topics Covered:** Syllabus, goals for class, R, RStudio installation.
- **Reading:** Syllabus, Canvas Page

Week 1 - 08/06

Class 2: Observations, Variables, Computing, and Interpreting Means

- **Topics Covered:** Dataframes, observations, variables, unit of observation, i, character vs. numeric variables, binary vs. non-binary variables, n, mean or average, $\sum$, unit of measurement; *R*: `+, -, *, /, <-, ", (), sqrt(), #, setwd(), read.csv(), View(), head(), dim(), mean()`.
- **Reading:** 1.0-1.10

Week 1 - 08/07

Class 3: Estimating Causal Effects with Randomized Experiments

- **Topics Covered:** Causal relationships, treatment (X) and outcome variables (Y), factual vs. counterfactual outcomes, problem of causal inference, individual and average causal effects, randomized experiments, random treatment assignment, treatment and control groups, pre-treatment characteristics, the difference-in-means estimator.
- **Reading:** 2.0-2.4

Week 1 - 08/08

Class 4: Does Social Pressure Increase the Probability of Turning Out to Vote?

- **Topics Covered:** *R*: `==, ifelse(), []`
- **Reading:** 2.5-2.7

ASSIGNMENT #1 DUE BY SATURDAY 8/10, 11:59 PM

———— WEEK 2 ————

Week 2 - 08/12

Class 5: Survey Research and Exploring One Variable at a Time

- **Topics Covered:** Sample, representative sample, random sampling, table of frequencies, table of proportions, histogram, descriptive statistics (mean, median, standard deviation, and variance); *R*: `table()`, `prop.table()`, `hist()`, `median()`, `sd()`, `var()`, `^`.
- **Reading:** 3-3.4 (skip: 3.2.2, 3.4.1, 3.4.2, 3.4.3, 3.4.5)

Week 2 - 08/13

Class 6: Exploring the Relationship Between Two Variables

- **Topics Covered:** scatter plot, correlation; *R*: *plot(), cor()*.
- **Reading:** 3.5-3.7

Week 2 - 08/14

Class 7: Predicting Non-Binary Outcomes Using Linear Regression

- **Topics Covered:** Prediction and correlation, predicted ($\hat{Y}$) vs. actual outcome (Y), prediction errors ($\hat{e}$), the least squares method, the linear regression model, $\hat{Y} = \hat{\alpha} + \hat{\beta}\Delta X$, interpretation of coefficients, intercept ($\hat{Y}$) and slope ($\hat{\beta}$), $\Delta\hat{Y} = \Delta\hat{\beta}X$; *R*: `lm(Y X)`, `abline()`.
- **Reading:** 4.0-4.4.1

Week 2 - 08/15

Class 8: Predicting Binary Outcomes Using Linear Regression

- **Topics Covered:** R^2, relationship between R^2 and correlation.
- **Reading:** 4.6-4.9 (skip 4.8)

ASSIGNMENT #2 DUE BY SATURDAY 8/17, 11:59 PM

———— WEEK 3 ————

Week 3 - 08/19

Class 9: Estimating Causal Effects with Observational Data and the Problem of Confounders

- **Topics Covered:** Observational studies vs. randomized experiments, confounders (Z), interpretation of and when X is binary and identifies treatment assignment.
- **Reading:** 5-5.3.1

Week 3 - 08/20

Class 10: Controlling for Confounders Using Multiple Linear Regression

- **Topics Covered:** Multiple vs. simple linear regression models, new interpretation of coefficients.
- **Reading:** 5.3.2-5.4.2

Week 3 - 08/21

Class 11: Internal vs. External Validity

- **Topics Covered:** Internal validity, external validity.
- **Reading:** 5.5-5.7

Week 3 - 08/22

Class 12: Probability

- **Topics Covered:** Probability, random variables, probability distributions, Bernoulli vs. normal distribution, the standard normal distribution, population parameters vs. sample statistics, the law of large numbers, the central limit theorem
- **Reading:** 6-6.8 (skip 6.7 and ignore code)

ASSIGNMENT #3 DUE BY SATURDAY 8/24, 11:59 PM

———— WEEK 4 ————

Week 4 - 08/26

Class 13: Hypothesis Testing with Coefficients

- **Topics Covered:** Hypothesis testing, test statistic, standard error of , *R:* (summary()$coef)
- **Reading:** 7-7.1 (skim), 7.3-7.6 (skip 7.3.1)

Week 4 - 08/27

Class 14: Do Small Classes Increase Probability of Graduating from High School?

- **Topics Covered:** Previous concepts
- **Reading:** 7.7 (PDF)

Week 4 - 08/28

Class 15: Self-Work Week

- **Topics Covered:** Learning how to allot time for research projects

Week 4 - 08/29

Class 16: Self-Work Week

- **Topics Covered:** Learning how to allot time for research projects

ASSIGNMENT #4 DUE BY SATURDAY 8/31, 11:59 PM

EXTRA CREDIT DUE BY SUNDAY 9/01, 11:59 PM

———— WEEK 5 ————

Week 5 - 09/02

Class 17: Self-Work Week

- **Topics Covered:** Learning how to allot time for research projects

Week 5 - 09/03

Class 18: Self-Work Week

- **Topics Covered:** Learning how to allot time for research projects

Week 5 - 09/04

Class 19: Self-Work Week

- **Topics Covered:** Learning how to allot time for research projects

Week 5 - 08/05

Class 20: Self-Work Week

- **Topics Covered:** Learning how to allot time for research projects

FINAL PROJECT DUE BY THURSDAY 9/05, 11:59 PM