SOCIOL 114: Social Data Science.   Winter 2025
Last updated: January 4, 2025

Faculty   **Office Hours**
Ian Lundberg, ianlundberg@ucla.edu, ianlundberg.org   Th 2–3pm, Haines 241C

PhD TAs
Christina Wilmot, cwilmot@ucla.edu   TBD
Nathan Hoffmann, nathanihoff@ucla.edu, nathanhoffmann.com   T 1:30–2:30, Haines A55

Where to send questions. Please post questions on Piazza (accessible through BruinLearn). This includes substantive questions about the material as well as administrative questions about the course. You will get a faster answer because all of the course staff and your peers are on Piazza. When possible, use Piazza rather than email!

Credits
4.0 Units, Student Option Grading (Letter, S/U)

Time and Location
Lecture: TTh 3:30–4:45pm in Perloff 1102
Discussion: Each student assigned to one of

- T 9:00-9:50am Haines A6

- T 5:00-5:50pm Public Affairs 1278

- T 6:00-6:50pm Public Affairs 1278

- Th 11:00-11:50am Mathematical Sciences 3915D

- Th 12:00-12:50pm Mathematical Sciences 3915D

- Th 5:00-5:50pm Bunche 2156

Course description. Lecture, three hours; discussion, one hour. Data analysis, and way social theory and data are linked. Covers data and computing environment, regression analysis, causal analysis, and machine learning. Offers tools for conducting quantitative analyses of social phenomenon, including emerging computational methods. Integrates substance and method. Draws on literature in social inequality to demonstrate applications of studied methods. P/NP or letter grading.

Longer course description. Rising computational power is transforming the way social scientists answer questions with data. This course explores how to use the tools of data science for social science goals. As a concrete area of application, we focus on social inequality including disparities by race, gender, and class and the transmission of inequality over time and across generations. Students will deploy tools for data science to visualize inequality, understand inequality, and evaluate hypothetical policy interventions that might reduce inequality. We will use the R programming language. A theme of the course is that applied work in this area can give rise to new data science tools, which may help solve some of society's most pressing challenges.

Course objectives. As a result of participating in this course, students will be able to

- visualize economic inequality with graphs that summarize survey data

- connect theories about inequality to quantitative empirical evidence

- evaluate possible interventions to reduce inequality

- conduct data analysis using the R programming language

Who should take this course? The course is designed for upper-division undergraduate students. If you would like to better understand inequality by using the tools of data science, this course is for you.

Prerequisites. No enforced prerequisites. Basic familiarity with introductory statistics (e.g., regression, confidence intervals) is useful. The R programming language will be taught as part of the course.

Instructional format. Lecture and discussion.

Course readings. Readings involve (1) a free textbook about using R for data science and (2) social science papers and commentaries on inequality. Example readings include:

- Data science

  - Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for Data Science*. Edition 2. O'Reilly Media, Inc.

- Social science papers and commentaries on inequality

  - Jencks, C. (2002). "Does inequality matter?" *Daedalus*, 131(1), 49-65.

  - England, P., Levine, A., & Mishel, E. (2020). Progress toward gender equality in the United States has slowed or stalled. *Proceedings of the National Academy of Sciences*, 117(13), 6990-6997.

Typesetting. Problem sets will be typeset using Quarto, which embeds code and results in a single reproducible document. Quarto will be taught.

Method of assessing student achievement. Grades will be determined by:

Problem sets	50%
Class participation	15%
Peer grading	10%
Final project presentation (10 minutes)	5%
Final project write-up (1000 words)	20%

For details, see Assignments.

Grading scale. Course grades will be assigned on the following scale:

		$87\% \leq x < 90\%$	B+	$77\% \leq x < 80\%$	C+	$67\% \leq x < 70\%$	D+
$93\% \leq x \leq 100\%$	A	$83\% \leq x < 87\%$	B	$73\% \leq x < 77\%$	C	$63\% \leq x < 67\%$	D
$90\% \leq x < 93\%$	A-	$80\% \leq x < 83\%$	B-	$70\% \leq x < 73\%$	C-	$60\% \leq x < 63\%$	D-
						$0\% \leq x < 60\%$	F

Assignments

Problem sets. Students will complete problem sets on their own which apply the data analysis ideas carried out in groups during the class session. Problem sets will involve a combination of data analysis, visualization of results, and written summaries.

Class participation. This class is interactive and participatory. Participation grades will be determined by attendance, participation, and engagement with in-class group work.

We are all absent sometimes. If you are feeling unwell, for example, then please stay home and feel better. There may also be other valid reasons for being absent. If you would like an absence to be excused, please fill out an excused form which is available on the Forms page of the course website.

Peer grading. A principle on which this course is built is that you will offer feedback to one another. Following this principle, after each problem set is submitted you will be assigned to anonymously grade the problem sets of three peers in Canvas. You will assign points using a rubric and offer constructive comments.

Final project. There will also be a group research project at the end of the semester, involving original analysis of data to answer a question of your choosing. This project will involve:

- A 10-minute presentation

- A writeup of the results containing up to 1,000 words containing at least 1 visualization

- A reproducible package of code that produces all results

Course Management

Late work. Each day an assignment is late deducts 10% of the assignment's total points. This deduction is applied automatically in BruinLearn. We will make exceptions to this policy in exceptional circumstances; come talk to us. Minimum grade value of 50%.

Reproducibility. A key principle of science is that we be transparent about the procedures that produced any reported result. In this course, all statistical results will be accompanied by the code that produces them (e.g., via Quarto). If a reported result should be generated by code and is clearly not generated by the accompanying code, we will impose a reproducibility penalty by subtracting off 20% of the total possible points on the assignment.

Academic integrity. Each student in this course is expected to abide by the UCLA Academic Integrity policies. Any work submitted by a student in this course for academic credit must be the student's own work.

Statement on accessible education.[1] Your access in this course is important to me. If you are already registered with the Center for Accessible Education (CAE), please request your Letter of Accommodation in the Student Portal. If you are seeking registration with the CAE, please submit your request for accommodations via the CAE website. Students with disabilities requiring academic accommodations should submit their request for accommodations as soon as possible, as it may take up to two weeks to review the request. For more information, please visit the CAE website (www.cae.ucla.edu), visit the CAE at A255 Murphy Hall, or contact them by phone at (310) 825-1501.

Mental health and wellbeing. Your health and wellbeing are important to me. There are services and resources at UCLA designed specifically to bolster well-being: see bewellbruin.ucla.edu. Remember, your mental health and emotional well-being are just as important as your physical health. If you or a friend are struggling emotionally or feeling stressed, fatigued, or burned out, it is always a good idea to connect with someone through one of the services at that link. If you find yourself experiencing a mental health crisis, call CAPS Crisis Counseling at (310) 825-0768.

---

[1]This statement is based on guidelines from the Center for Accessible Education.