

## SOCIAL 114: Social Data Science.

Winter 2026

Last updated: January 2, 2026

### Faculty

Ian Lundberg, [ianlundberg@ucla.edu](mailto:ianlundberg@ucla.edu), [ianlundberg.org](http://ianlundberg.org)

### PhD TA

Taylor Aquino, [taquino7@ucla.edu](mailto:taquino7@ucla.edu)

### Time and Location

Lecture: MW 10–11:15am in Dodd 147

Office hours:

- MW 11:15am-11:50am in Dodd 147 (Ian)
- TBD (Taylor)

### Pattern of the Week

This course has a very consistent pattern each week. Each MW, lecture ends at 11:15am but the room is reserved through 11:50am. We encourage you to stay and work on coursework and quizzes. We will be there to provide help!

Every MW at 5pm, a quiz is due on the day's lecture materials.

Every F at 5pm, a problem set is due. All material will be covered by the end of lecture on T.

**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)

**Course description.** Lecture, 150 minutes (2 x 75); optional post-lecture collaborative office hours, 70 minutes (2 x 35). 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.

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

- 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 office hours.

**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. Examples include:

- 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.
- Bertrand, M & Mullainathan, S. 2004. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94(4):991–1013.

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

Quizzes	50%
Problem sets	50%

For details, see [Assignments](#).

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

$97\% \leq x \leq 100\%$	A+	$87\% \leq x < 90\%$	B+	$77\% \leq x < 80\%$	C+	$67\% \leq x < 70\%$	D+
$93\% \leq x < 97\%$	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

**Quizzes.** Each lecture will be paired with a brief quiz in BruinLearn. Attendance is not mandatory, but attending lecture will make it much easier for you to do well on these quizzes.

Recognizing that some absences are to be expected, BruinLearn will automatically drop your lowest 2 quiz scores at the end of the quarter.

**Problem sets.** Students will complete problem sets that involve a combination of data analysis, visualization of results, and written summaries. Problem sets will be due on Fridays at 5pm.

## Course Management

**Late work.** Each hour an assignment is late deducts 0.5% of the assignment's total points. Thus for example an assignment submitted 1 minute after the deadline has a 0.5% penalty, and an assignment submitted 23 hours 1 minute after the deadline has a ( $24 \text{ hours} \times 0.5\% \text{ per hour} = 12\%$  penalty). This deduction is applied automatically in BruinLearn. No assignments will be accepted after Mar 20 at 5pm.

**Software requirements.** This course will involve coding in the R programming language. For most students, the easiest way to do this is to install R and RStudio on your laptop (instructions at [posit.co/download/rstudio-desktop](https://posit.co/download/rstudio-desktop)). Some students may prefer to use a cloud server, for example if your laptop is a Chromebook. One option is Posit Cloud ([posit.cloud/plans](https://posit.cloud/plans)), for which the Basic Plan (free) is likely sufficient or you can sign up for the Student Plan (\$5 per month) if you find the Basic Plan is insufficient.

**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 30% 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.

**Collaboration.** You are encouraged to work together. Consulting with each other on problem sets is often a good idea. However, this should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an email, an email attachment file, or a hard copy.

**Statement on accessible education.**<sup>1</sup> 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](http://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](http://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.

---

<sup>1</sup>This statement is based on [guidelines](#) from the Center for Accessible Education.