Info 3370 / 5371: Studying Social Inequality with Data Science. Spring 2024
Last updated: February 15, 2024

Course Staff — **Office Hours**

Ian Lundberg, ilundberg@cornell.edu, ianlundberg.org — T 10–11am, Gates 223

PhD TAs

Federica Bologna, fb265@cornell.edu, federicabologna.github.io — W 2–3pm, Rhodes 406
Daniel Molitor, djm484@cornell.edu, dmolitor.com — W 12–1pm, Rhodes 406

Undergraduate TAs

Stacy Boey, seb397@cornell.edu — M 12–1pm, Rhodes 402
Chang Chen, cc992@cornell.edu — M 3–4pm, Rhodes 408
Olaf de Rohan Willner, ocd4@cornell.edu — W 3–4pm, Rhodes 406
Joanne Hu, xh285@cornell.edu — M 1–2pm, Rhodes 408
Akira Goh, ag969@cornell.edu — Th 3:30–4:30pm, Zoom
Jonathan Gotian, jeg338@cornell.edu — Th 1–2pm, Uris G44
Elizabeth Moon, em652@cornell.edu — W 12:30–1:30pm, Rhodes 402

Where to send questions. Please post questions on Ed Discussion (accessible through Canvas). 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 Ed. When possible, use Ed rather than email!

Credits
3.0 Credits, Student Option Grading (Letter, S/U)

Time and Location
T 8:40–9:55am in Gates G01
Th 8:40–9:55am in Phillips 203 and Upson 142
28 total meetings

Course description. Inequality is high in American society. Income and wealth are concentrated in far fewer hands than in other industrialized countries. Labor market outcomes are patterned by disparities across lines of race, gender, and class. This course will introduce social science theories about the origins of inequality, emphasizing how inequality is transmitted over time and across generations. Building on these theories, 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 (3370) and MA students (5371). If you would like to better understand inequality by using the tools of data science, this course is for you.

Prerequisites. Students are expected to have basic familiarity with introductory statistics (e.g., regression, confidence intervals) and experience with some statistical programming language (e.g., INFO 2950, SOC 3010, PAM 2101). Programming in the course will be in R, but knowledge specific to R is not a prerequisite. R will be taught as part of the course.

Instructional format. Active learning combining lecture, discussion, and in-class analysis of quantitative data.

Course readings. Readings involve (1) a free textbook about using R for data science and (2) social science papers and commentaries on inequality. The full semester of readings is yet to be determined, but a few examples 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.
  - Coates, Ta-Nehisi. 2014. "The case for reparations." *The Atlantic* June 2014.

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.

Each problem set will contain a bonus challenge question.

- The bonus question is optional for undergraduate students. There is no extra credit for completing it.
- The bonus question is required for master's students.

Class participation. This class is interactive and participatory. Participation grades will be determined by attendance, participation in class, 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. If you miss any class or discussion meeting, please fill out this form to submit a short writeup for participation credit.

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
  - For undergraduates: Up to 1,000 words containing at least 1 visualization
  - For master's students: Up to 2,000 words containing at least 3 visualizations
- A reproducible package of code that produces all results

Course Management

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

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.

Late work. 5 flex days to be used on problem sets with no questions. Each day beyond your 5 deducts 10% of the assignment's total points, so that the max score after 1 days late is 90% and the max score after 2 days late is 80%. We will make exceptions to this policy in exceptional circumstances; come talk to us. Minimum grade value of 50%.

Students with disabilities.[1] Your access in this course is important to me. Please request your accommodation letter early in the semester, or as soon as you become registered with Student Disability Services (SDS), so that we have adequate time to arrange your approved academic accommodations.

- Once SDS approves your accommodation letter, it will be emailed to both you and me. Please follow up with me to discuss the necessary logistics of your accommodations.
- If you experience any access barriers in this course, such as with printed content, graphics, online materials, or any communication barriers; reach out to me or SDS right away.
- If you need an immediate accommodation, please speak with me after class or send an email message to me and to SDS at sds_cu@cornell.edu.
- If you have, or think you may have a disability, please contact Student Disability Services for a confidential discussion: sds_cu@cornell.edu, 607-254-4545, sds.cornell.edu.

Mental health and wellbeing. Your health and wellbeing are important to me. There are services and resources at Cornell designed specifically to bolster undergraduate, graduate, and professional student mental health and well-being. 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, there is a continuum of campus resources available to you: mentalhealth.cornell.edu/get-support/support-students. Help is also available any time day or night through Cornell's 24/7 phone consultation (607-255-5155). You can also reach out to me, your college student services office, your resident advisor, or Cornell Health for support.

[1] This statement is based on guidelines from Student Disability Services.