SOC 10: Introduction to Computational Social Science
Spring 2025
Mon, Wed 10:30 AM - 11:50 AM
Instructor: Yuze Sui (yuzesui@stanford.edu)
Course Website (for lab): https://yuzesui97.github.io/soc10_2025spring
Office Hour: Mon, Wed 12:00 PM – 1 PM at 380-380W; Tue 10:30 AM – 11:30 AM via ZOOM

Overview
With an ever-growing amount of data collected on our online and offline behaviors—from our purchasing habits to our social interactions—we now have the unprecedented opportunity to study social life with remarkable precision. This hands-on course explores the promises and pitfalls of leveraging "big data" and algorithms to understand modern social and economic systems. Each week, we will examine key sociological topics such as inequality, discrimination, and polarization through the lens of computational methods, as well as delve into cutting-edge techniques in computational social science. The course culminates in a group research project that will equip you with skills ranging from designing and executing large-scale surveys and experiments to fitting machine learning models and analyzing diverse data types like social networks and natural language. No prior statistical or programming experience is required.

Course Format
Classes meet twice a week and blend lectures with active discussions and hands-on lab sessions:

- **First Class (Lecture & Discussion):**
  Each session begins with a lecture on how computational social scientists approach classic sociological issues (e.g., inequality, polarization, social norms), followed by a discussion of assigned readings. Students are expected to complete the readings beforehand and come prepared with questions.

- **Second Class (Lab Session):**
  The lab focuses on the essential tools of computational social science—from data collection and cleaning to analysis. In the first half, new computational methods are introduced; in the second half, you will practice these techniques.

In the final week (Week 10), instead of our usual format, you will present your final group projects in class.

Course Requirements
There are four requirements for this class, which are about class readings, coding, participation, and final project.

Class reading and response memos (30% of total grade):
For each week, you are expected to finish the class readings before the first class of the week. In addition, you are expected to write short response memos (no more than 500 words) to the readings in 5 of the 8 weeks for which there are readings (excluding the first week). The goal of the response memo is to extend the ideas presented in the papers to a new subject or topic that is of interest to you. That can mean, for example, discussing how a computational method could be

improved or modified, developing a new idea for a method, or imagining another area, perhaps one related to your own interests, to which a method or idea could be applied. It is therefore a *creative* exercise.

Response memos should be posted on Canvas by 11:59 PM the day before the first class of each week. Late submissions incur a one-point deduction per day.

Four Coding Exercises (20% of total grade):
Beginning with Week 2, a coding exercise will be released immediately after the lab to reinforce the skills covered. Each exercise must be submitted before the following lab session. Over the course of Weeks 2 to 8, seven exercises will be available, and you are required to complete any four of them. You have the flexibility to choose which exercises to work on. These exercises are designed to be short (~30 mins of work).

Attendance and Participation (10% of total grade):
Participation is key for this course as this is a course involves a large number of discussions and hands-on practices. You are expected to come to each class in person and actively engage in class discussions. You are expected to speak up at least twice in class to earn the full participation grade.

Final project (40% of total grade):
The final requirement is to analyze data using the skills and ideas we learn in class and to write a short paper (no more than ten pages) on your study (30% of total grade). This report will consider the problem you set out to investigate, your data and data collection strategies, your analysis and results, and the limitations of your study. You will complete the report in groups of 2-3 people. You must also present your paper to the class during the last week of classes for the quarter (10% of total grade). After settling on an idea for your study, your group should schedule a one-on-one meeting with me to discuss the project and make sure it is feasible and compelling (no later than the end of the sixth week).

In life, things happen, and people miss classes and deadlines. If you think you have excusable reasons, please email me at least 24 hours in advance for a deadline extension or a permission for absence.

Grading
10% Attendance and Participation
30% Response Memos
20% Coding Assignments
30% Final Paper
10% Final Presentation

Readings and Software
Textbook:

- Salganik, M. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

It is a great reference book for aspiring computational social scientists and the digital version is openly available online https://www.bitbybitbook.com/en/1st-ed/preface/

Other reading:
All other readings, which consist of either book chapters or journal articles, will be uploaded to the Canvas page under the "Readings" folder.

Computer Software:
For this class, students are expected to run basic programming codes using R (a popular programming language for statistical analyses) and R Studio (a popular graphical user interface for R).

R and R Studio are free that you can download from https://cran.r-project.org and https://www.rstudio.com, respectively. Please make sure to download R and R Studio before the second class of the first week.

Accommodations
If you have an Office of Accessible Education letter, please present it to me at your earliest convenience, so I can ensure that the course materials and support comply with your needs.

I want you to be in our class. You are always welcome to write your preferred name on all class assignments and exams. If you have a name and/or pronoun preference that doesn't match what our class roster gets from the registrar's office, please let me know and I will make sure we use that in our class.

Detailed Class Schedule (by Week)

Week 1 Introductions on Computational Social Science and R

In our first week, we'll embark on an exciting journey into the vibrant world of computational social science—a field that marries the art of social inquiry with the power of cutting-edge computational tools. In this week, we'll discuss what computational social science is, how it offers fresh perspectives beyond traditional quantitative methods, and delve into the intriguing ethical dilemmas that come with using digital data to understand human behavior. Alongside these explorations, we'll roll up our sleeves and get hands-on with R, a dynamic and accessible tool that will be our trusty companion throughout the course. Get ready to dive in, have fun, and see how joyful and transformative the intersection of computation and social science can be!

Readings for Monday:

Required:

- Lazer, D. et al. (2009). "Computational Social Science." *Science*.
- Salganik, M. (2017). *Bit by Bit*, Chapter 1-3 & 6.

Optional:

- Blumenstock, J. et al. (2015). "Predicting poverty and wealth from mobile phone metadata." *Science*.
- Hofman, J. et al. (2021) "Integrating explanation and prediction in computational social science." *Nature*.
- Kramer, A. et al. (2014). "Experimental evidence of massive-scale emotional contagion through social networks." *PNAS*.
- Wagner, C. et al. (2021) "Measuring algorithmically infused societies." *Nature*.

For Wednesday:

- Our class will be a crash course on analyzing data with R
- **DUE:** Before class, install R and Rstudio, and work through this introductory tutorial on DataQuest: https://www.dataquest.io/blog/tutorial-getting-started-with-r-and-rstudio/#tve-jump-173bb251cef

Week 2 Inequality & Surveys and Survey Experiments

In Week 2, we delve into the multifaceted nature of inequality and explore how computational approaches serve as powerful tools for investigating social inequalities. Through our Monday readings, you'll encounter diverse studies that analyze inequality using different large scale data (Wikipedia, Cell phone signal, etc) and various computational methods (network analysis, simulations, etc). Then, in our Wednesday lab, we'll focus on online survey methods and learn how to design and interpret these critical research tools. Get ready for a week where theory and practical methods come together to illuminate the complexities of inequality in our society.

Readings for Monday:

Required:

- Adams, J. et al. (2019). "Who Counts as a Notable Sociologist on Wikipedia? Gender, Race, and the 'Professor Test'". Socius, 5, 2378023118823946.
- Chang, S. (2021). "Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening." Nature 589(7840): 82–87.
- Salganik et al. (2006). "Experimental study of inequality and unpredictability in an artificial cultural market." Science.

Optional:

- Pradhan, N. and Agrawal, A. (2025). "Mapping fine-scale socioeconomic inequality using machine learning and remotely sensed data" PNAS Nexus.
- Tripodi, F. (2021). "Ms. Categorized: Gender, notability, and inequality on Wikipedia". New Media & Society.

For Wednesday:
Our lab will be on surveys and survey experiments. Please read Salganik (2017). *Bit by Bit*, Chapter 4 before coming to lab.

Week 3 Stereotypes and Discrimination & Regression

Week 3 immerses us in the critical examination of stereotypes and discrimination. This week, we'll explore how biases influence labor market outcomes and cultural narratives, drawing on seminal field experiments and innovative studies that decode gender and ethnic stereotypes through computational techniques. As you engage with our required readings, you'll gain a deep understanding of how empirical research and predictive models converge to reveal the underlying dynamics of societal bias. In Wednesday's lab, you'll have the opportunity to learn and run basic predictive regression models (which are the foundations of machine learning!), from logistic to OLS, on large-scale data.

Readings for Monday:
Required:

- 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.
- Garg, N. et al. (2018). "Word embeddings quantify 100 years of gender and ethnic stereotypes." *Proceedings of the National Academy of Sciences*, *115*(16), E3635-E3644.
- Xu et al. (2019). "The Cinderella Complex: Word embeddings reveal gender stereotypes in movies and books." *Plos One*.

Optional:

- Correll, S., Weisshaar, K., Wynn, A., & Wehner, J. (2020). "Inside the Black Box of Organizational Life: The Gendered Language of Performance Assessment." *American Sociological Review*, 85(6): 1022-1050.
- Doleac, J. L., & Stein, L. C. (2013). The visible hand: Race and online market outcomes. *The Economic Journal*, *123*(572), F469-F492.

For Wednesday:
Our lab will be about running basic predictive linear regression models on large-scale data (including logistic regressions on binary outcomes and OLS regressions on continuous outcomes).

Week 4 Homophily, Diffusion and Social Connection & Network Analysis

Week 4 explores the dynamic interplay of homophily, diffusion, and social connection within social networks. This week, you'll delve into foundational studies that distinguish between contagion driven by social influence and diffusion stemming from inherent similarities among individuals. Our required readings will guide you through experiments and models that reveal how behaviors and social ties spread across networks, illustrating how homophilic tendencies can shape and strengthen social connections. Then, in Wednesday's lab, you'll gain practical experience with network analysis using R and the "igraph" package, equipping you with robust tools to map and interpret these complex interactions.

**In this week, we will welcome a guest speaker whose pioneering work in computational social science is reshaping the field.*

Readings for Monday:
Required:

- Aral, S. et al. (2009) "Distinguishing influence-based contagion from homophily driven diffusion in dynamic networks," *PNAS*.
- Centola, D. (2010). "The spread of behavior in an online social network experiment," *Science*, 329(5996), 1194-1197.
- Wimmer, A., & Lewis, K. (2010). "Beyond and below racial homophily: ERG models of a friendship network documented on Facebook," *American Journal of Sociology*, 116(2), 583-642.

Optional:

- Banarjee et al. (2013) "The Diffusion of Microfinance," *Science*.
- Centola, D. (2011) "An experimental study of homophily in the adoption of health behavior," *Science*.
- Godoy-lorite, A. and Jones, N. (2021). "Inference and influence of network structure using snapshot social behavior without network data," *Science*.

For Wednesday:
Our lab will be about running network analysis with R. Please install the package "igraph" before the class.

Week 5 Polarization and Misinformation & Collect Data Online

Week 5 delves into the complex phenomena of political polarization and the spread of misinformation in digital environments. This week, our required readings explore how exposure to diverse—and sometimes opposing—views on social media can paradoxically intensify political divides, while also examining the mechanisms behind the diffusion of both information and radical content online. You'll engage with studies that analyze the role of social media platforms in shaping public discourse during critical events like the US 2020 election. Then, in Wednesday's lab, you'll shift gears to the technical side by learning how to scrape data and work with APIs, empowering you to collect and analyze online data firsthand.

Readings for Monday:
Required:

- Bail, C. A., et al. (2018). "Exposure to opposing views on social media can increase political polarization," *PNAS*, *115*(37), 9216-9221.
- González-Bailón, S. et al. "The Diffusion and Reach of (Mis) Information on Facebook During the US 2020 Election," *Sociological Science*.
- Hosseinmardi et al. (2021) "Examining the consumption of radical content on YouTube," *PNAS*.
- Nyhan, B. et al. (2023). "Like-minded sources on Facebook are prevalent but not polarizing," *Nature*.

Optional:

- Cineli, M. et al. (2021). "The echo chamber effect on social media," *PNAS*.
- Martel, C. et al. (2023). "Crowds Can Effectively Identify Misinformation at Scale," *Perspectives on Psychological Science*.
- Williams-Ceci, S. et al. (2023). "Trustworthiness Evaluations of Search Results: The Impact of Rank and Misinformation," *arXiv*.

For Wednesday:
Our lab will be about scraping data and APIs.

Week 6 Culture and Social Norms & Text Analysis

Week 6 invites you to explore the intricate relationship between culture and social norms through the lens of text analysis. In this week, you'll delve into groundbreaking studies that reveal how social conventions and the unwritten rules of online communities are shaped and reinforced. Our required readings provide experimental evidence on tipping points in social behavior and offer insights into how digital platforms monitor and enforce norms. Then, in Wednesday's lab, you'll pivot to the realm of text analysis by learning to run basic multinominal language models, vector space models, and word-embedding models—tools that will empower you to decode the subtleties of language and cultural expression in digital texts.

Readings for Monday:

Required:

- Centola, D. et al. (2018). "Experimental evidence for tipping points in social convention," *Science*.
- Chandrasekharan, E. et al. (2018). "The Internet's Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales." *Proc. ACM Hum.-Comput. Interact.* 2(CSCW): 32:1-32:25. doi:10.1145/3274301.

Optional:

- Centola, D. et al. (2005). "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms," *American Journal of Sociology*.
- Eichstaedt, J., et al (2018). Facebook language predicts depression in medical records. *PNAS*, *115*(44), 11203-11208.
- Goffman, E. (1959). *The presentation of self in everyday life,* Chapter 1.
- Migliano, A. et al. (2020). "Hunter-gatherer multilevel sociality accelerates cumulative cultural evolution," *Science*.
- Rajadesingan, A. et al. (2020). "Quick, Community-Specific Learning: How Distinctive Toxicity Norms Are Maintained in Political Subreddits." *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 557-568.
- Ricart-Huguet, J. & Paluck, E. (2023). "When the Sorting Hat Sorts Randomly: A Natural Experiment on Culture." *Quarterly Journal of Political Science* 18 (1), 39-73

For Wednesday:
Our lab will be about running basic multinominal language models, vector space models, and word-embedding models.

Week 7 Job, Occupations, and Organization & Machine Learning

Week 7 delves into the intersections of job dynamics, occupational structures, and organizational behavior through the innovative lens of machine learning. This week, our required readings examine how cultural adaptation influences individual outcomes in organizations, the impact of minimum wage policies on labor markets using advanced computational techniques, and how supervised machine learning can predict trends in occupational growth and decline. In addition to exploring these multifaceted topics, Wednesday's lab will provide hands-on experience with machine learning methods—specifically, random forests and neural networks. Get ready to uncover insights at the nexus of organizational theory and modern computational techniques.

**In this week, we will welcome a guest speaker whose pioneering work in computational social science is reshaping the field.*

Readings for Monday:

Required:

- Srivastava,S. et al. (2017). "Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations," *Management Science*.
- Cengiz, D. et al. (2022). "Seeing beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes," Labor Economics.
- Khalaf, C. et al. (2023). "Predicting declining and growing occupations using supervised machine learning," *Journal of Computational Social Science*.

Optional:

- Athey, S. et al. (2024). "The Heterogeneous Earnings Impact of Job Loss Across Workers, Establishments, and Markets," *working paper*.
- Athey, S. et al. (2024). "LABOR-LLM: Language-Based Occupational Representations with Large Language Models," *working paper*.
- Reagans, R. et al. (2023). "Shared language in the team network-performance association: Reconciling conflicting views of the network centralization effect on team performance," *Collective Intelligence*.

For Wednesday:
Our lab will be on machine learning using random forests and neural networks.

Week 8 Can Large Language Models Transform Computational Social Science? & ChatGPT

Week 8 examines the transformative potential of large language models in the realm of computational social science, with a particular focus on ChatGPT. This week, you'll engage with cutting-edge research that investigates how these models can revolutionize traditional methodologies—from generating novel research ideas to enhancing interdisciplinary applications across psychology, finance, and beyond. Our required readings provide critical insights into the current capabilities and limitations of large language models, while optional texts offer broader perspectives on their impact in diverse fields. Then, in Wednesday's lab, you'll delve into the underlying algorithmic infrastructure of ChatGPT and learn strategies for leveraging its power efficiently for academic research. Get ready to explore a frontier where artificial intelligence and social science intersect in exciting new ways.

Readings for Monday:

Required:

- Ziems, C. et al. (2023). "Can Large Language Models Transform Computational Social Science?" *Computational linguistics*.
- Demszky, D. (2023) "Using large language models in psychology," *Nature Reviews Psychology*.
- Si, C. et al. 2024. "Can llms generate novel research ideas? a large-scale human study with 100+ nlp researchers," *working paper*.

Optional:

- Li, et al. (2023). "Large Language Models in Finance: A Survey," *ICAIF*.
- Tessler, M et al. (2024). "AI can help humans find common ground in democratic deliberation," *Science*.

For Wednesday:
Our lab will be about understanding the underlying algorithmic infrastructure of ChatGPT and how to properly and efficiently utilize ChatGPT for academic research.

Week 9 How Algorithms are reshaping our society (no lab)

Week 9 investigates how algorithms are reshaping various aspects of society. With Monday designated as a University Holiday, our session will reconvene on Wednesday to engage with a series of compelling readings. Required texts, including Boegen's exploration of how hiring algorithms can introduce bias, Kiviat's examination of the moral limits in predictive practices, and Tao et al.'s analysis of cultural bias in large language models, will serve as a foundation for discussing the ethical, cultural, and social implications of algorithmic decision-making. Get ready to critically analyze the revolutionary ways in which algorithms are redefining our society.

Readings for Wednesday (Note that Monday is a University Holiday):

Required:

- Boegen, M. (2019). "All the Ways Hiring Algorithms Can Introduce Bias". *Harvard Business Review*.
- Kiviat, B. (2019). "The Moral Limits of Predictive Practices: The Case of Credit-Based Insurance Scores," *American Sociological Review*.
- Tao, Y. et al. (2024). "Cultural bias and cultural alignment of large language models," *PNAS Nexus*.

Optional:

- Fleder and Hosanagar (2009) "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity," *Management Science*.
- Kleinberg, Ragavan (2021) "Algorithmic monoculture and social welfare," *PNAS*.
- Liang, W. et al. (2024). "Mapping the Increasing Use of LLMs in Scientific Papers."