

Topics in Computational Social Science

University of Texas at Austin, Department of Sociology (Spring 2026)

Meeting time: Thursdays, 3:30pm-6:30pm

Location: RLP 3.106

Course Instructor: Casey Breen (casey.breen@austin.utexas.edu)

Office Hours: Tuesdays, 11:30am-2:00pm ([sign-up link](#))

Course website: caseybreen.com/teaching/css (slides, labs, and in-class demos)

Overview

The growing availability of new streams of data, expansion of computational power, and the digitalisation of our lives has created new questions and research opportunities for social and population scientists. The course will introduce students to a range of methodological and substantive topics in computational social science. We will cover topics such as digital trace and big data, machine learning, non-probability sampling, social networks, and agent-based modelling and microsimulation. The course will consist of seminar and lab sessions (taught in R), where students will engage with research in computational social science and learn to apply basic computational methods to research problems based on existing research papers.

Learning Outcomes

By the end of this course, students will be able to:

1. Critically evaluate and engage with contemporary scientific debates within the area of computational social science and identify key contributions;
2. Apply computational social science perspectives to formulate and address relevant sociological and demographic questions;
3. Demonstrate proficiency in some techniques of computational social science, including methods for processing and analysing digital and unstructured data, basic methods for machine learning, and addressing bias in non-probability samples;
4. Apply and engage with a computational social science perspective to address their own research problems and areas of interest.

Teaching Arrangement

The teaching will be organised in a combination of lectures, discussion-based seminars, and four computer labs.

In presenting a paper to the group, please focus on: What is the objective of the paper, what are the main contributions, and what are its limitations? If it is an empirical paper, does the

analysis/data seem convincing and does the strength of the argument match the data and evidence? If it is a theoretical/perspective paper, what questions and ideas does it generate for research in computational social science?

All participants, and not only those presenting, are expected to read, at minimum, all core readings for the week (marked with a star, ★, below). We strongly encourage you to review all readings for a given week.

Requirements

The course will be taught in R (using RStudio). Students are expected to have R and RStudio downloaded and be familiar with R. As an additional resource, we suggest the following text, which is freely available online:

- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (2nd Edition). O'Reilly. <https://r4ds.hadley.nz/>

Course Assessment

The course will be graded based on:

1. Attendance and participation in discussion
2. Computer labs
3. Final project, consisting of:
 - **Topic pitch** (February 19): a short presentation outlining the proposed final project (research question, substantive importance, data, and computational methods)
 - **Final presentation** (April 23): a 10-minute presentation of the final project, followed by brief Q&A
 - **Final paper** (~10 pages), due **Friday, May 1 at 11:59pm**

Course Schedule

Week	Date	Format	Topic
1	Thurs, January 15	Lecture	Introduction to Computational Social Science
2	Thurs, January 22	Lab	Lab 0: Intro to R
3	Thurs, January 29	Lecture	Social Data in the Digital Age
4	Thurs, February 5	Lab	Lab 1: Social Data Collection
5	Thurs, February 12	Lecture	Machine Learning + Prediction I
6	Thurs, February 19	Seminar	Final Paper Topic Pitches
7	Thurs, February 26	Lecture	Machine Learning + Prediction II
8	Thurs, March 5	Lab	Lab 2: Machine Learning
9	Thurs, March 12	Lecture	Probability and Non-Probability Sampling
10	Thurs, March 19	–	<i>Spring Break, no class</i>
11	Thurs, March 26	Lecture	Social Networks
12	Thurs, April 2	Lab	Lab 3: Non-Probability Sampling
13	Thurs, April 9	Lecture	Text as Data / Large Language Models
14	Thurs, April 16	Lecture	Simulation and Agent-Based Models
15	Thurs, April 23	Seminar	Final Presentations

Week	Date	Format	Topic
–	Fri, May 1	–	Final papers due (11:59pm)

Weekly Schedule and Readings

Core readings (discussed in class) are marked with a star (★). Slides, labs, and in-class demos are posted on the [course website](#).

Week 1 (Thurs, January 15): Introduction to Computational Social Science

- Lazer, David, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. 2021. “Meaningful Measures of Human Society in the Twenty-First Century.” *Nature* 595(7866):189-96. doi: 10.1038/s41586-021-03660-7. ★
- Edelman, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. “Computational Social Science and Sociology.” *Annual Review of Sociology* 46(1):61-81. ★
- Salganik, Matthew J. 2019. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 1. [Freely available online] ★
- Kashyap, Ridhi, R. Gordon Rinderknecht, Aliakbar Akbaritabar, Diego Alburez-Gutierrez, Sofia Gil-Clavel, André Grow, Jisu Kim et al. 2023. “Digital and Computational Demography.” In *Research Handbook on Digital Sociology*, pp. 48-86. Edward Elgar Publishing.
- Lazer, David M. J., Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, Alondra Nelson, Matthew J. Salganik, Markus Strohmaier, Alessandro Vespignani, and Claudia Wagner. 2020. “Computational Social Science: Obstacles and Opportunities.” *Science* 369(6507):1060-62. doi: 10.1126/science.aaz8170.

Week 2 (Thurs, January 22): Intro to R [LAB]

Please bring your laptop to class with R and RStudio installed; also install the **tidyverse** package.

Helpful resources:

- *R for Data Science*: <https://r4ds.hadley.nz/>
- *Data Visualization: A Practical Introduction*: <https://socviz.co/>

Week 3 (Thurs, January 29): Social Data in the Digital Age

- Kashyap, Ridhi. 2021. “Has Demography Witnessed a Data Revolution? Promises and Pitfalls of a Changing Data Ecosystem.” *Population Studies* 75(sup1):47-75. doi: 10.1080/00324728.2021.1969031. ★
- Salganik, Matthew J. 2019. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 2. ★
- Bruch, Elizabeth E., and M. E. J. Newman. 2018. “Aspirational Pursuit of Mates in Online Dating Markets.” *Science Advances* 4(8):eaap9815. doi: 10.1126/sciadv.aap9815. ★
- Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. “The Parable of Google Flu: Traps in Big Data Analysis.” *Science* 343(6176):1203-5. doi: 10.1126/science.1248506. ★
- Breen, Casey F., and Dennis M. Feehan. 2024. “New Data Sources for Demographic Research.” *Population and Development Review* 51(1):539-73. doi: 10.1111/padr.12671. ★

- Ruggles, Steven. 2014. “Big Microdata for Population Research.” *Demography* 51(1):287-97. doi: 10.1007/s13524-013-0240-2.
- Ruggles, Steven, Catherine A. Fitch, and Evan Roberts. 2018. “Historical Census Record Linkage.” *Annual Review of Sociology* 44(1):19-37. doi: 10.1146/annurev-soc-073117-041447.
- Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, et al. 2022. “Social Capital I: Measurement and Associations with Economic Mobility.” *Nature* 608(7921):108-21. doi: 10.1038/s41586-022-04996-4.
- Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. “Fake News on Twitter during the 2016 U.S. Presidential Election.” *Science* 363(6425):374-78. doi: 10.1126/science.aau2706.

Week 4 (Thurs, February 5): Social Data Collection [LAB]

Collect original data on the count of Facebook users broken down by characteristics using the Facebook marketing API. Lab materials: [Lab 1 Quarto Doc](#).

- Fatehkia, Masoomali, Ridhi Kashyap, and Ingmar Weber. 2018. “Using Facebook Ad Data to Track the Global Digital Gender Gap.” *World Development* 107:189-209. ★
- Zagheni, Emilio, Ingmar Weber, and Krishna Gummadi. 2017. “Leveraging Facebook’s Advertising Platform to Monitor Stocks of Migrants.” *Population and Development Review* 43(4):721-34.
- Araujo, Matheus, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. 2017. “Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations.” In *Proceedings of the 2017 ACM on Web Science Conference*, pp. 253-257.

Week 5 (Thurs, February 12): Machine Learning + Prediction I

- Hofman, Jake M., Duncan J. Watts, Susan Athey, Filiz Garip, Thomas L. Griffiths, Jon Kleinberg, Helen Margetts, Sendhil Mullainathan, Matthew J. Salganik, Simine Vazire, Alessandro Vespignani, and Tal Yarkoni. 2021. “Integrating Explanation and Prediction in Computational Social Science.” *Nature* 595(7866):181-88. doi: 10.1038/s41586-021-03659-0. ★
- Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. “Predicting Poverty and Wealth from Mobile Phone Metadata.” *Science* 350(6264):1073-76. doi: 10.1126/science.aac4420. ★
- Chi, Guanghua, Han Fang, Sourav Chatterjee, and Joshua E. Blumenstock. 2022. “Microestimates of Wealth for All Low- and Middle-Income Countries.” *Proceedings of the National Academy of Sciences* 119(3):e2113658119. doi: 10.1073/pnas.2113658119. ★
- Molina, Mario, and Filiz Garip. 2019. “Machine Learning for Sociology.” *Annual Review of Sociology* 45(1):27-45. doi: 10.1146/annurev-soc-073117-041106. ★

Week 6 (Thurs, February 19): Final Paper Topic Pitches

Students will give short presentations outlining their proposed final projects. Each pitch should clearly state the research question, explain its substantive importance, and describe the data and computational methods to be used. The session will focus on constructive feedback (be kind!) to help refine questions, assess feasibility, and ensure projects are well scoped.

Week 7 (Thurs, February 26): Machine Learning + Prediction II

- Salganik, Matthew J., Ian Lundberg, et al., and Sara McLanahan. 2020. “Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration.” *Proceedings of the National Academy of Sciences* 117(15):8398-8403. doi: 10.1073/pnas.1915006117. ★
- Lundberg, Ian, Rachel Brown-Weinstock, Susan Clampet-Lundquist, Sarah Pachman, Timothy J. Nelson, Vicki Yang, Kathryn Edin, and Matthew J. Salganik. 2024. “The Origins of Unpredictability in Life Outcome Prediction Tasks.” *Proceedings of the National Academy of Sciences* 121(24):e2322973121. doi: 10.1073/pnas.2322973121. ★
- Lundberg, Ian, Jennie E. Brand, and Nanum Jeon. 2022. “Researcher Reasoning Meets Computational Capacity: Machine Learning for Social Science.” *Social Science Research* 108:102807. doi: 10.1016/j.ssresearch.2022.102807.
- Phillips, Rachael V., Mark J. van der Laan, Hana Lee, and Susan Gruber. 2023. “Practical Considerations for Specifying a Super Learner.” *International Journal of Epidemiology* 52(4):1276-85. doi: 10.1093/ije/dyad023.

Week 8 (Thurs, March 5): Machine Learning [LAB]

Apply machine learning algorithms and Facebook user counts to predict overall levels of internet adoption and digital divisions. Lab materials: [Lab 2 Quarto Doc](#).

- Breen, Casey F., Masoomali Fatehkia, Jiani Yan, Xinyi Zhao, Douglas R. Leasure, Ingmar Weber, and Ridhi Kashyap. 2024. “Mapping Subnational Gender Gaps in Internet and Mobile Adoption Using Social Media Data.” ★

Week 9 (Thurs, March 12): Probability and Non-Probability Sampling

- Stedman, Richard C., Nancy A. Connelly, Thomas A. Heberlein, Daniel J. Decker, and Shorna B. Allred. 2019. “The End of the (Research) World As We Know It? Understanding and Coping With Declining Response Rates to Mail Surveys.” *Society & Natural Resources* 32(10):1139-54. doi: 10.1080/08941920.2019.1587127. ★
- Schneider, Daniel, and Kristen Harknett. 2022. “What’s to Like? Facebook as a Tool for Survey Data Collection.” *Sociological Methods & Research* 51(1):108-40. doi: 10.1177/0049124119882477. ★
- Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. “Forecasting Elections with Non-Representative Polls.” *International Journal of Forecasting* 31(3):980-91. doi: 10.1016/j.ijforecast.2014.06.001. ★
- Dutwin, David, and Trent D. Buskirk. 2017. “Apples to Oranges or Gala versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples.” *Public Opinion Quarterly* 81(S1):213-39. doi: 10.1093/poq/nfw061. ★
- Lehdonvirta, Vili, Atte Oksanen, Pekka Räsänen, and Grant Blank. 2021. “Social Media, Web, and Panel Surveys: Using Non-Probability Samples in Social and Policy Research.” *Policy & Internet* 13(1):134-55. doi: 10.1002/poi3.238.
- Elliott, Michael R., and Richard Valliant. 2017. “Inference for Nonprobability Samples.” *Statistical Science* 32(2):249-64.
- Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. “Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls.” *Political Analysis* 12(4):375-85. doi: 10.1093/pan/mpg024.
- Breen, Casey F., Ayesha S. Mahmud, and Dennis M. Feehan. 2022. “Novel Estimates Reveal

Subnational Heterogeneities in Disease-Relevant Contact Patterns in the United States.” *PLOS Computational Biology* 18(12):e1010742. doi: 10.1371/journal.pcbi.1010742.

Week 10 (Thurs, March 19): Spring Break, No Class

Week 11 (Thurs, March 26): Social Networks

- Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. “Network Analysis in the Social Sciences.” *Science* 323(5916):892-95. doi: 10.1126/science.1165821. ★
- Watts, Duncan J. 2004. “The ‘New’ Science of Networks.” *Annual Review of Sociology* 30(1):243-70. doi: 10.1146/annurev.soc.30.020404.104342. ★
- Eagle, Nathan, Alex (Sandy) Pentland, and David Lazer. 2009. “Inferring Friendship Network Structure by Using Mobile Phone Data.” *Proceedings of the National Academy of Sciences* 106(36):15274-78. doi: 10.1073/pnas.0900282106. ★
- Salganik, Matthew J., and Douglas D. Heckathorn. 2004. “Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling.” *Sociological Methodology* 34(1):193-240. ★
- Kossinets, Gueorgi, and Duncan J. Watts. 2006. “Empirical Analysis of an Evolving Social Network.” *Science* 311(5757):88-90. doi: 10.1126/science.1116869.
- Feehan, Dennis M., and Curtiss Cobb. 2019. “Using an Online Sample to Estimate the Size of an Offline Population.” *Demography* 56(6):2377-92. doi: 10.1007/s13524-019-00840-z.
- Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78(6):1360-80.
- Goel, Sharad, and Matthew J. Salganik. 2010. “Assessing Respondent-Driven Sampling.” *Proceedings of the National Academy of Sciences* 107(15):6743-47. doi: 10.1073/pnas.1000261107.
- Centola, Damon, and Michael Macy. 2007. “Complex Contagions and the Weakness of Long Ties.” *American Journal of Sociology* 113(3):702-34. doi: 10.1086/521848.

Week 12 (Thurs, April 2): Non-Probability Sampling [LAB]

Apply poststratification and inverse-probability weighting to a non-probability sample to estimate mortality rates using the network survival method. Lab materials: [Lab 3 Quarto Doc](#).

- Breen, Casey F., Saeed Rahman, Christina Kay, Joeri Smits, Abraham Azar, Steve Ahuka, and Dennis M. Feehan. 2026. “Estimating Death Rates in Complex Humanitarian Emergencies Using the Network Survival Method.” *American Journal of Epidemiology* 195(1):49-59. doi: 10.1093/aje/kwaf101.
- Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. 2013. “Summary Report of the AAPOR Task Force on Non-Probability Sampling.” *Journal of Survey Statistics and Methodology* 1(2):90-143. doi: 10.1093/jssam/smt008.

Week 13 (Thurs, April 9): Text as Data / Large Language Models in Social Science Research

- Bail, Christopher A. 2012. “The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks.” *American Sociological Review* 77(6):855-79. doi: 10.1177/0003122412465743. ★
- Waight, Hannah, Yin Yuan, Margaret E. Roberts, and Brandon M. Stewart. 2025. “The Decade-Long Growth of Government-Authored News Media in China under Xi

Jinping.” *Proceedings of the National Academy of Sciences* 122(11):e2408260122. doi: 10.1073/pnas.2408260122. ★

- Bail, Christopher A. 2024. “Can Generative AI Improve Social Science?” *Proceedings of the National Academy of Sciences* 121(21):e2314021121. ★
- Alvero, AJ, Dustin S. Stoltz, Oscar Stuhler, and Marshall A. Taylor. 2025. “Generative AI in Sociological Research: State of the Discipline.” *Sociological Science*. ★
- Than, Nga, Leanne Fan, Tina Law, Laura K. Nelson, and Leslie McCall. 2025. “Updating ‘The Future of Coding’: Qualitative Coding with Generative Large Language Models.” *Sociological Methods & Research* 54(3):849-88. doi: 10.1177/00491241251339188. ★
- Grimmer, Justin, and Brandon M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21(3):267-97. doi: 10.1093/pan/mps028.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. “Text as Data.” *Journal of Economic Literature* 57(3):535-74. doi: 10.1257/jel.20181020.
- Hastings, Orestes P., and Luca Maria Pesando. 2024. “What’s a Parent to Do? Measuring Cultural Logics of Parenting with Computational Text Analysis.” *Social Science Research* 124:103074.

Week 14 (Thurs, April 16): Simulation and Agent-Based Models

- Bruch, Elizabeth, and Jon Atwell. 2015. “Agent-Based Models in Empirical Social Research.” *Sociological Methods & Research* 44(2):186-221. ★
- Bruch, Elizabeth E., and Robert D. Mare. 2006. “Neighborhood Choice and Neighborhood Change.” *American Journal of Sociology* 112(3):667-709. doi: 10.1086/507856. ★
- Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. “Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market.” *Science* 311(5762):854-56. doi: 10.1126/science.1121066. ★
- Alburez-Gutierrez, Diego, Carl Mason, and Emilio Zagheni. 2021. “The ‘Sandwich Generation’ Revisited: Global Demographic Drivers of Care Time Demands.” *Population and Development Review* 47(4):997-1023. ★
- Kashyap, Ridhi, and Francisco Villavicencio. 2016. “The Dynamics of Son Preference, Technology Diffusion, and Fertility Decline Underlying Distorted Sex Ratios at Birth: A Simulation Approach.” *Demography* 53(5):1261-1281.
- Gao, Chen, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. 2024. “Large Language Models Empowered Agent-Based Modeling and Simulation: A Survey and Perspectives.” *Humanities and Social Sciences Communications* 11(1):1259. doi: 10.1057/s41599-024-03611-3.
- Spielauer, Martin. 2011. “What Is Social Science Microsimulation?” *Social Science Computer Review* 29(1):9-20.
- Zagheni, Emilio. 2011. “The Impact of the HIV/AIDS Epidemic on Kinship Resources for Orphans in Zimbabwe.” *Population and Development Review* 37(4):761-783.
- Grow, André, and Jan Van Bavel. 2015. “Assortative Mating and the Reversal of Gender Inequality in Education in Europe: An Agent-Based Model.” *PloS One* 10(6):e0127806.

Week 15 (Thurs, April 23): Final Presentations

Students will deliver 10-minute presentations of their final projects. Presentations should clearly communicate the research question, data, methods, and main findings, with special attention to the

computational approach applied. Each presentation will be followed by brief Q&A.

Final papers (~10 pages) are due by Friday, May 1 at 11:59pm.