

SOC577  
COMPUTATIONAL SOCIOLOGY  
Rutgers University

Syllabus

Thomas Davidson

Spring 2024

## CONTACT AND LOGISTICS

E-mail: [thomas.davidson@rutgers.edu](mailto:thomas.davidson@rutgers.edu)

Website: <https://github.com/t-davidson/computational-sociology>

Class meetings: Thursday 2:00-4:40 pm (Zoom)

Office hours: Monday 2:00-3:00 pm or by appointment.

## COURSE DESCRIPTION

This course is designed to introduce students to computational methods and their applications to sociological research. We will discuss the computational toolkit from the bottom up, beginning with the fundamentals of programming and data analysis and management. Once these fundamentals are in place, we will turn to different methods for collecting data: application programming interfaces, web-scraping, and online experiments. The remainder of the course will focus on computational methods for data analysis. First, we will cover various methods for the quantitative analysis of text data including word embeddings, topic modeling, and supervised learning. Second, we will discuss machine learning in more depth, assessing the relationship between prediction and explanation in social science, discussing bias and other limitations of these methods, as well as the opportunities these techniques present to work with images and multimodal data. Finally, we will explore the role of simulation and agent-based modeling in sociological research. Throughout the course students will gain hands-on experience with these different techniques, as well as an understanding of how these techniques are being used in cutting-edge sociological research. Overall this course will provide students with a strong conceptual foundation in computational sociology and the ability to apply various techniques for data collection and analysis in their own research. All assignments will be conducted using the R programming language.

## PREREQUISITES AND PREPARATION

This course is designed for students without any experience using computational methods or advanced statistics. Nonetheless, the course will proceed more efficiently if students without any such experience

are willing to undertake some independent learning prior to the beginning of the course. In particular, I recommend students familiarize themselves with Github, the R programming language, the RStudio computing environment, and RMarkdown documents, as we will be using these tools throughout the course. We will review these topics over the first few weeks of class, but the more familiar students are with these tools, the more time we can spend focusing on their sociological applications. Information on learning resources is provided on the course website.

## ASSESSMENT

There will be two types of assessment used in the course. There will be four homework assignments (each worth 10% of the final grade) designed to help students to become familiar with the various methodological techniques covered in the course. A schedule of these assignments can be found in the course outline below. Students will also write an empirical paper over the course of the semester, worth 60% of the final grade. The paper will involve the collection of original data and preliminary analyses using one or more of the approaches covered in the course. I intend for the paper to be an opportunity for students to develop the basis for a qualifying exam, master's thesis, or dissertation chapter. Students are expected to make progress on the paper over the course of the semester. At various points in the semester there will be three submissions related to the final paper (each worth 5% of the final grade): the paper proposal, initial data collection and descriptive analysis, and the implementation of methodological approach (see course outline for a timeline). Each of these stages will be an opportunity to gain feedback on the final paper.

## READINGS

There are weekly reading assignments for this course. These readings include methodological texts, reviews of relevant methodological and theoretical considerations, and examples of how sociologists and other social scientists apply computational approaches in their research. Given the complexity and unfamiliarity of some of the approaches we will cover in the course, I have included a diverse set of readings for each topic. Some students may find the technical readings more useful whereas others may benefit from the more contextualized applications.

### Require texts

- \*Matthew Salganik. 2017. *Bit by Bit*. Princeton University Press. <https://www.bitbybitbook.com/en/1st-ed/preface/>
- \*Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. (R4DS). O'Reilly Media, Inc. <https://r4ds.had.co.nz/>
- \*Martin, James H., and Daniel Jurafsky. 2020. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd edition. Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3/>
- Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.
- \*Silge, Julia, and David Robinson. 2017. *Text Mining with R: A Tidy Approach*. O'Reilly Media. <https://www.tidytextmining.com/dtm.html>.
- \*Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton University Press. <https://socviz.co/>

\* Available for free online at website provided.

## RESOURCES

There will be a Github repository containing all files related to this course (A link will be shared in the first week of class). I will also use this as a space to maintain a wiki links to various related resources. Students will also use Github to submit assignments. We will be using Slack to communicate with one another during this course, all enrolled students will receive an invite to join the Slack channel.

## COURSE POLICIES

The Rutgers Sociology Department strives to create an environment that supports and affirms diversity in all manifestations, including race, ethnicity, gender, sexual orientation, religion, age, social class, disability status, region/country of origin, and political orientation. This class will be a space for tolerance, respect, and mutual dialogue. Students must abide by the Code of Student Conduct at all times, including during lectures and in participation online.

All students must abide by the university's Academic Integrity Policy. Violations of academic integrity will result in disciplinary action.

In accordance with University policy, if you have a documented disability and require accommodations to obtain equal access in this course, please contact me during the first week of classes. Students with disabilities must be registered with the Office of Student Disability Services and must provide verification of their eligibility for such accommodations.

## COURSE OUTLINE

### Week 1

#### Introduction to Computational Sociology

##### *Readings*

- *R4DS*: Preface, C2-6, 21
- *Bit by Bit*, C1
- Lazer, David, et al. 2009. "Life in the Network: The Coming Age of Computational Social Science." *Science* 323 (5915): 721–23. <https://doi.org/10.1126/science.1167742>.
- Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." *Annual Review of Sociology* 46 (1): annurev-soc-121919-054621. <https://doi.org/10.1146/annurev-soc-121919-054621>.
- Nelson, Laura 2017. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research*, November. <https://doi.org/10.1177/0049124117729703>.

### Week 2

#### Data Structures

##### *Readings*

- *R4DS*: C7-10, 16
- Golder, Scott A., and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." *Annual Review of Sociology* 40 (1): 129–52. <https://doi.org/10.1146/annurev-soc-071913-043145>.
- Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." *Theory and Society* 43 (3–4): 465–82. <https://doi.org/10.1007/s11186-014-9216-5>.

## Week 3

### Programming Fundamentals

#### Readings

- R4DS: C14-15, 17
- Liu, David, and Matthew Salganik. 2020. "Successes and Struggles with Computational Reproducibility: Lessons from the Fragile Families Challenge," *Socius: Sociological Research for a Dynamic World*.

**Assignment 1 released: The computational toolkit. Due 2/9 at 5pm.**

## Week 4

### Application Programming Interfaces

#### Readings

- R4DS: C11 ("Strings with stringr"), 13 ("Dates and Times with lubridate")
- *Bit by Bit*, C2
- Askin, Noah, and Michael Mauskopf. 2017. "What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music." *American Sociological Review* 82 (5): 910–44. <https://doi.org/10.1177/0003122417728662>.
- Freelon, Deen. 2018. "Computational Research in the Post-API Age." *Political Communication* 35 (4): 665–68. <https://doi.org/10.1080/10584609.2018.1477506>.

## Week 5

### Webscraping

**Assignment 2: Collecting and storing data released. Due TBD.**

#### Readings

- *Bit by Bit*, C6
- Fiesler, Casey, Nate Beard, and Brian C Keegan. 2020. "No Robots, Spiders, or Scrapers: Legal and Ethical Regulation of Data Collection Methods in Social Media Terms of Service." In *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media*, 187–96.
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107 (02): 326–43. <https://doi.org/10.1017/S0003055413000014>.

## Week 6

### Online Experiments and Surveys

#### Readings

- *Bit by Bit*, C3-5
- Salganik, Matthew J., and Duncan J. Watts. 2008. "Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market." *Social Psychology Quarterly* 71 (4): 338–55. <https://doi.org/10.1177/019027250807100404>.
- Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks." *Proceedings of the National Academy of Sciences* 111 (24): 8788–90. <https://doi.org/10.1073/pnas.1320040111>.

- Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31 (3): 980–91. <https://doi.org/10.1016/j.ijforecast.2014.06.001>.
- Munger, Kevin. 2016. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior*, November. <https://doi.org/10.1007/s11109-016-9373-5>.

## Week 7

### Machine Learning and Statistics

#### Readings

- Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45: 27–45.
- Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24 (1): annurev-polisci-053119-015921. <https://doi.org/10.1146/annurev-polisci-053119-015921>.
- Salganik, Matthew, Ian Lundberg, Alexander Kindel, et al. 2020. "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration." *Proceedings of the National Academy of Sciences*.

#### Recommended

## Week 8

### Fundamentals of Natural Language Processing

#### *Paper proposals due 3/5 at 5pm.*

#### Readings

- *Text Mining with R*, C1, 3-5
- Grimmer, Justin, and Brandon Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97. <https://doi.org/10.1093/pan/mps028>.
- Evans, James, and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42 (1): 21–50. <https://doi.org/10.1146/annurev-soc-081715-074206>.

#### Recommended

- *Speech and Language Processing*, C6, pages 1-13.

## Week 9

### Supervised text classification

#### *Paper initial data collection due TBD.*

#### Readings

- Barberá, Pablo, Amber E. Boydston, Suzanna Linn, Ryan McMahon, and Jonathan Nagler. 2020. "Automated Text Classification of News Articles: A Practical Guide." *Political Analysis*, June, 1–24. <https://doi.org/10.1017/pan.2020.8>.

#### Recommended

- Hanna, Alex. 2013. "Computer-Aided Content Analysis of Digitally Enabled Movements." *Mobilization: An International Quarterly* 18 (4): 367–388.

- Davidson, Thomas, Debasmita Bhattacharya, and Ingmar Weber. 2019. "Racial Bias in Hate Speech and Abusive Language Detection Datasets." In *Proceedings of the Third Workshop on Abusive Language Online*, 25–35. Florence, Italy: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-3504>.
- Miller, Blake, Fridolin Linder, and Walter R. Mebane, Jr. 2019. "Active Learning Approaches for Labeling Text: Review and Assessment of the Performance of Active Learning Approaches." *Political Analysis*. [http://www-personal.umich.edu/~wmebane/Paper\\_Active\\_Learning\\_Approaches\\_for\\_Labeling\\_Text.pdf](http://www-personal.umich.edu/~wmebane/Paper_Active_Learning_Approaches_for_Labeling_Text.pdf).
- King, Gary, Patrick Lam, and Margaret E. Roberts. 2017. "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text." *American Journal of Political Science* 61 (4): 971–88.
- Dun, Lindsay, Stuart Soroka, and Christopher Wlezien. 2020. "Dictionaries, Supervised Learning, and Media Coverage of Public Policy." *Political Communication*, June, 1–19. <https://doi.org/10.1080/10584609.2020.1763529>.
- Dixon, Lucas, John Li, Jeffrey Sorensen, Nithum Thain, and Lucy Vasserman. 2018. "Measuring and Mitigating Unintended Bias in Text Classification." In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES '18*, 67–73. New Orleans, LA, USA: ACM Press. <https://doi.org/10.1145/3278721.3278729>.

## Week 10

### (Large) Language Models

#### Readings

- Danescu-Niculescu-Mizil, Cristian, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. "No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities." In *Proceedings of the 22nd International Conference on World Wide Web*, 307–318. <http://dl.acm.org/citation.cfm?id=2488416>.
- Wankmüller, Sandra. 2022. "Introduction to Neural Transfer Learning With Transformers for Social Science Text Analysis." *Sociological Methods & Research*. doi: 10.1177/00491241221134527.
- Bonikowski, Bart, Yuchen Luo, and Oscar Stuhler. 2022. "Politics as Usual? Measuring Populism, Nationalism, and Authoritarianism in U.S. Presidential Campaigns (1952–2020) with Neural Language Models." *Sociological Methods & Research* 51(4):1721–87. doi: 10.1177/00491241221122317.
- Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 14. Canada: ACM Press.

#### Recommended

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." P. 11 in *NIPS*. Long Beach, CA, USA.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." Pp. 4171–86 in *Proceedings of NAACL-HLT*. ACL.
- Brown, Tom B., and many others. 2020. "Language Models Are Few-Shot Learners." <https://arxiv.org/pdf/2005.14165.pdf>

## Week 11

### Word Embeddings and Computational Semantics

#### Readings

- *Text Mining with R*: C5.
- *Speech and Language Processing*, C6, pages 17-30.

- Hvittfeldt, Emil and Julia Silge. 2020 *Supervised Machine Learning for Text Analysis in R*. Chapter 5: <https://smldtar.com/embeddings.html>.
- Kozlowski, Austin, Matt Taddy, and James Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review*, September, 000312241987713. <https://doi.org/10.1177/0003122419877135>.
- Stoltz, Dustin S., and Marshall A. Taylor. 2021. "Cultural Cartography with Word Embeddings." *Poetics* 101567. doi: 10.1016/j.poetic.2021.101567.
- Arseniev-Koehler, Alina, and Jacob G. Foster. 2022. "Machine Learning as a Model for Cultural Learning: Teaching an Algorithm What It Means to Be Fat." *Sociological Methods & Research* 51(4):1484–1539. doi: 10.1177/00491241221122603.
- Zhou, Di. 2022. "The Elements of Cultural Power: Novelty, Emotion, Status, and Cultural Capital." *American Sociological Review* 000312242211230. doi: 10.1177/00031224221123030.

#### Recommended

- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." In *Advances in Neural Information Processing Systems*, 3111–3119. <http://papers.nips.cc/paper/5021-distributed-representations>.
- Hamilton, William, Jure Leskovec, and Dan Jurafsky. 2016. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1489–1501.
- Rodriguez, Pedro L., and Arthur Spirling. 2021. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research." *The Journal of Politics*. doi: 10.1086/715162.
- Rodman, Emma. 2019. "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors." *Political Analysis*, July, 1–25. <https://doi.org/10.1017/pan.2019.23>.
- Gonen, Hila, and Yoav Goldberg. 2019. "Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But Do Not Remove Them." In *Proceedings of NAACL\_HLT*, 609–14. Minneapolis, Minnesota: Association for Computational Linguistics.
- De-Arteaga, Maria, Alexey Romanov, Hanna Wallach, Jennifer Chayes, Christian Borgs, Alexandra Chouldechova, Sahin Geyik, Krishnaram Kenthapadi, and Adam Tauman Kalai. 2019. "Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting." In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT '19*, 120–28. Atlanta, GA, USA: ACM Press. <https://doi.org/10.1145/3287560.3287572>.

## Week 12

### Topic Modeling

#### **Assignment 3: Natural language processing released. Due TBD.**

#### Readings

- *Text Mining with R*: C6.
- Blei, David 2012. "Probabilistic Topic Models." *Communications of the ACM* 55 (4): 77. <https://doi.org/10.1145/2133806.2133826>.
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41 (6): 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>.
- Roberts, Margaret, Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses: Structural Topic Models for Survey Responses." *American Journal of Political Science* 58 (4): 1064–82. <https://doi.org/10.1111/ajps.12103>.
- Karell, Daniel, and Michael Freedman. 2019. "Rhetorics of Radicalism." *American Sociological Review* 84 (4): 726–53. <https://doi.org/10.1177/0003122419859519>.

### Recommended

- Mohr, John, and Petko Bogdanov. 2013. "Introduction—Topic Models: What They Are and Why They Matter." *Poetics* 41 (6): 545–69. <https://doi.org/10.1016/j.poetic.2013.10.001>.

## Assignment 4: Machine learning released. Due 4/26 at 4pm

### Week 13

#### Computer Vision and Multimodal Machine Learning

##### Readings

- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems*, 1097–1105. <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>.
- Olah, Chris, Alexander Mordvintsev, and Ludwig Schubert. 2017. "Feature Visualization." *Distill*. <https://doi.org/10.23915/distill.00007>.
- Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353 (6301): 790–94. <https://doi.org/10.1126/science.aaf7894>.
- Gebru, Timnit, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. 2017. "Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods across the United States." *Proceedings of the National Academy of Sciences* 114 (50): 13108–13. <https://doi.org/10.1073/pnas.1700035114>.
- Zhang, Han, and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media." *Sociological Methodology* 49 (1): 1–57. <https://doi.org/10.1177/0081175019860244>.
- Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of Machine Learning Research*, 81:1–15.

##### Recommended

- Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. 2021. "Learning Transferable Visual Models From Natural Language Supervision." OpenAI, 47. [Blog: <https://openai.com/blog/clip/>, code: <https://github.com/openai/CLIP>].  
– See <https://microscope-azure-edge.openai.com/models> for model visualizations

### Week 14

#### Agent-Based Models

##### Readings

- Macy, Michael, and Robert Willer. 2002. "From Factors to Factors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28 (1): 143–66. <https://doi.org/10.1146/annurev.soc.28.110601.141117>.
- Bruch, Elizabeth, and Jon Atwell. 2015. "Agent-Based Models in Empirical Social Research." *Sociological Methods & Research* 44 (2): 186–221. <https://doi.org/10.1177/0049124113506405>.
- Watts, Duncan. 1999. "Networks, Dynamics, and the Small-World Phenomenon." *American Journal of Sociology* 105 (2): 493–527. <https://doi.org/10.1086/210318>.
- Centola, Damon. 2015. "The Social Origins of Networks and Diffusion." *American Journal of Sociology* 120 (5): 1295–1338. <https://doi.org/10.1086/681275>.
- DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" *American Journal of Sociology* 120 (5): 1473–1511. <https://doi.org/10.1086/681254>.



- Goldberg, Amir, and Sarah K. Stein. 2018. "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation." *American Sociological Review* 83 (5): 897–932. <https://doi.org/10.31235/osf.io/uqvvd3>.

#### *Recommended*

- Shirado, Hirokazu, and Nicholas A. Christakis. 2017. "Locally Noisy Autonomous Agents Improve Global Human Coordination in Network Experiments." *Nature* 545 (7654): 370–74. <https://doi.org/10.1038/nature22332>.
- Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. 2017. "Mastering the Game of Go without Human Knowledge." *Nature* 550 (7676): 354–59. <https://doi.org/10.1038/nature24270>.

## **Week 15**

### *Paper preliminary analyses due TBD*

#### **Student presentations**

- 5 minute presentation, plus Q&A
  - Five slides (Add to shared Google Slides presentation)
    - \* Title
    - \* Motivation
    - \* Data
    - \* Methodology
    - \* Results

## **Final paper**

*Final paper due TBD*