SOC 681
COMPUTATIONAL SOCIAL SCIENCE USING R
Purdue University

Syllabus

Dr. Marcus Mann

Fall 2024

CONTACT AND LOGISTICS

E-mail: `mannm1@purdue.edu`

Website: `https://github.com/marcone84/SOC681-Introduction-to-R`

Class meetings: Wednesday 1:30-4:20 p.m.

Office hours: by appointment.

COURSE DESCRIPTION

This (crash) course is designed to provide a friendly and broad introduction to computational methods in social science using R. First, we will discuss the history and evolving role of computational methods in social science and their attendant implications for conducting ethical and generative research. Then we will discuss and practice the most basic programming skills in R including setting working directories, searching for and loading relevant packages, and basic date cleaning and manipulation. After this, the course will focus on introducing and applying a specific method of data collection or analysis every week. These will include 1) data wrangling, 2) working with API's, 3) experiments and surveys, 4) text analysis methods, 5) and data visualization. While this course does aim to broaden your awareness and skills in various kinds of data collection and analysis, it is not a statistics course, so do not expect detailed coverage of specific statistical methods or strategies. In general, we will study and discuss how computational methods have changed and broadened social scientific research, when they should and should not be applied, and - most importantly - how they can help you with your specific research agenda.

PREREQUISITES AND PREPARATION

This course assumes no experience in R or other programming languages and is meant to be a gentle introduction to using R for social scientific research. That said, online resources are readily available to attain at

1

least a rudimentary understanding of the basic skills needed to begin coding in R. One such resource is the Summer Institute in Computational Social Science websit here: https://sicss.io/overview

ASSESSMENT

100% of your grade will be based on a "final paper" (at least 5,000 words) that uses some kind of computational method covered in this course. This paper will be assessed by 1) relevance to a particular field, 2) justification of hypotheses, 3) justification of computational method, and 3) clear communication of results (including at least one data viz) and implications. This paper has no other parameters.

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 range and complexity 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 instructional readings more useful whereas others may benefit from the academic articles applying these methods to substantive research questions.

Require texts and useful references

indicates a required text. All required texts and useful references are available for free online on the listed websites.

- *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/
- *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/

COURSE POLICIES

The Purdue 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

This outline is tentative and subject to change.

Week 1 - For 8/28

Introduction to Computational Sociology

Readings

- *R4DS*: Preface, "Chapters" 4-7 on website
- *Bit by Bit*, Chapter 1
- Lazer, David, et al. 2009. "Computational Social Science." *Science* 323 (5915): 721–23. https://www.science.org/doi/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): https://www.annualreviews.org/doi/abs/10.1146/annurev-soc-121919-054621

Week 2 - For 9/4

Data Structures, Wrangling, and the Tidyverse

Readings

- *R4DS*: "Chapters" 9-13 on website
- 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 - For 9/11

Data Collection I: APIs

Readings

- *R4DS*: C11 ("Strings with stringr"), 13 ("Dates and Times with lubridate")
- *Bit by Bit*, C2
- Baumgartner, Jason, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. "The Pushshift Reddit Dataset." In *Proceedings of the International AAAI Conference on Web and Social Media*, 14:830–39.
- Freelon, Deen. 2018. "Computational Research in the Post-API Age." *Political Communication* 35 (4): 665–68. https://doi.org/10.1080/10584609.2018.1477506.

Recommended

I have included recommended readings that use a range of different APIs including Spotify (Askin and Mauskampf), Facebook (Davidson and Berezin; Bail, Brown and Mann), Google Trends (Davidson and Berezin; Bail, Brown, and Wimmer, Gross and Mann), Twitter (Mitts), and YouTube (Munger and Phillips).

- Askin, Noah, and Michael Mauskapf. 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.
- Davidson, Thomas, and Mabel Berezin. 2018. "Britain First and the UK Independence Party: Social Media and Movement-Party Dynamics." *Mobilization: An International Quarterly* 23 (4): 485–510. https://doi.org/10.17813/1086-671X-23-4-485.
- Bail, Christopher, Taylor Brown, and Andreas Wimmer. 2019. "Prestige, Proximity, and Prejudice: How Google Search Terms Diffuse across the World." *American Journal of Sociology* 124 (5): 1496–1548. https://doi.org/10.1086/702007.
- Bail, Christopher, Taylor Brown, and Marcus Mann. 2017. "Channeling Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation." *American Sociological Review* 82(6): 1188-1213.
- Mitts, Tamar. 2019. "From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West." *American Political Science Review* 113 (1): 173–94. https://doi.org/10.1017/S0003055418000618.
- Munger, Kevin, and Joseph Phillips. 2020. "Right-Wing YouTube: A Supply and Demand Perspective." *The International Journal of Press/Politics*, 34.
- Gross, Neil and Mann, Marcus.2017. "Is There a Ferguson Effect? Google Searches, Concern About Police Violence, and Crime in U.S. Cities, 2014-2016." Socius. 3, 1-16.

Week 4 - For 9/18

Data Collection II: 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.
- Munger, Kevin. 2016. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior*, November. https://doi.org/10.1007/s11109-016-9373-5.
- 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.
- Christopher Bail, Lisa Argyle, Taylor W. Brown, John Bumpus, Haohan Chen, M.B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media." Proceedings of the National Academy of Sciences.

Week 5 - For 9/25

Natural Language Processing I: Fundamentals

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.
- DiMaggio, Paul. 2015. "Adapting Computational Text Analysis to Social Science (and Vice Versa)." *Big Data & Society* 2 (2): 205395171560290. https://doi.org/10.1177/2053951715602908.
- 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.
- Austin Van Loon. 2022. "Three Families of Automated Text Analysis". Working Paper. https://osf.io/preprints/socarxiv/htnej/

Recommended

- *Speech and Language Processing*, C6, pages 1-13.
- Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. "Echoes of Power: Language Effects and Power Differences in Social Interaction." In *Proceedings of the 21st International Conference on World Wide Web*, 699–708. ACM. http://dl.acm.org/citation.cfm?id=2187931.
- 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.
- Niculae, Vlad, Srijan Kumar, Jordan Boyd-Graber, and Cristian Danescu-Niculescu-Mizil. 2015. "Linguistic Harbingers of Betrayal: A Case Study on an Online Strategy Game." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. Beijing, China: ACL.

Week 6 - For 10/2

Natural Language Processing II: Dictionaries, word embeddings, topic models, and networks

Readings

- *Text Mining with R*: C5.
- *Speech and Language Processing*, C6, pages 17-30.
- 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.
- 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.
- Arseniev-Koehler, Alina, and Jacob G. Foster. 2020. "Machine Learning as a Model for Cultural Learning: Teaching an Algorithm What It Means to Be Fat." Working Paper. *SocArXiv*. https://doi.org/10.31235/osf.io/c9yj3.
- *Text Mining with R*: C6.
- 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.
- 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.
- Blei, David 2012. "Probabilistic Topic Models." *Communications of the ACM* 55 (4): 77. https://doi.org/10.1145/2133806.2133826.
- Stoltz, Dustin S, and Marshall A Taylor. 2019. "Textual Spanning: Finding Discursive Holes in Text Networks." *Socius: Sociological Research for a Dynamic World* .
- 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.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of NAACL-HLT 2019*, 4171–86. ACL.
- Manning, Christopher D., Kevin Clark, John Hewitt, Urvashi Khandelwal, and Omer Levy. 2020. "Emergent Linguistic Structure in Artificial Neural Networks Trained by Self-Supervision." *Proceedings of the National Academy of Sciences*, June, 201907367. https://doi.org/10.1073/pnas.1907367117.
- Rodman, Emma. 2019. "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors." *Political Analysis*l, July, 1–25. https://doi.org/10.1017/pan.2019.23.

Week 7 - For 10/9

Data Visualization

Readings

- Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton University Press. https://socviz.co/

Final paper

Final paper due on Saturday 10/20 at 5pm ET