SOC6708H: Computational Methods for Sociologists

Fedor A. Dokshin

Fall 2022

E-mail: fedor.dokshin@utoronto.ca
Office Hours: By appointment
Office: 725 Spadina Ave, Rm 372

Web: Quercus
Class Hours: Mondays 10am-1pm
Class Room: FE36

Course Description

This course introduces students to the field of computational social science, a set of commonly used computational tools, and their applications to question of sociological interest. New forms of digital data present enormous opportunities for social research. These data include the fine-grained and time-stamped records of human behaviour and interactions online, massive troves of text and other "unstructured" data, and digitized documents and administrative records. Although often defined by the specific set of data sources and tools, computational social science can also be characterized by an attitude toward social research that appreciates the substantial opportunities afforded by the increasing availability of digitized data sources, an openness to novel methodologies, and a willingness to write lots of code and learn on the fly. This course seeks to cultivate this attitude. You will be introduced to a suite of currently popular methods, but you will also be encouraged to develop an aptitude and comfort with coding and to keep pace with the rapidly evolving computational toolkit in your subfield.

Course Objectives

1. Identify the opportunities and challenges presented by digital data and computational approaches for sociological research
2. Learn common research strategies, data collection approaches, and methods used in the emerging field of computational social science
3. Practice coding in R and/or Python across a variety of tasks and grow comfortable with writing lots of code to accomplish practical research tasks

Prerequisites

You will get the most out of this course if you have general knowledge of statistical concepts taught in the graduate sociology statistics sequence (i.e., basics of descriptive and inferential statistics, including linear regression). Lab examples will primarily use R, with some Python in the mix. Proficiency in either is not required, but familiarity with R would be very beneficial, as the schedule will not allow us to spend much time introducing the basics. If you are unfamiliar with R or

are feeling rusty, you are encouraged to work through the following set of tutorials before the first session: R for Social Scientists (focus on modules 1-5).

Required Readings

All readings for this course will be available openly online, through Quercus, or through the UofT Library, with the exception of *Bit by Bit*, which you can purchase by following the links on the book's website. We will also read several chapters from James et al.'s book, *An Introduction to Statistical Learning with Applications in R*, which you may purchase, or access through the UofT Library.

Salganik, Matthew J. 2019. *Bit by Bit: Social Research for the digital Age.* Princeton University Press. - Abbreviated as *Bit by Bit*

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2021. *An Introduction to Statistical Learning with Applications in R.* 2nd Edition. Springer. - Abbreviated as *ISLR*

Other readings include journal articles, which are available through the UofT Library.

Class Structure

Lectures are an ineffective way to learn computational social science. The best way to learn computational methods is by doing! With this in mind, each session will be divided into two parts: (1) a discussion of relevant empirical research articles and (2) a lab session for introducing the methods and walking through computational examples.

Evaluation Components

- **Participation and attendance** (*15%* of final grade): You are responsible for reading the assigned materials and coming to class prepared to discuss it.
- **Lab Assignments** (*30%* of final grade): There will be three Lab Assignments, which will ask you to use R and/or Python to complete tasks related to the content covered in class. These are designed to have you practice your coding and the methods you learn in class.
- **Article peer reviews** (*15%* of final grade): Each student will write peer reviews of 2 published empirical research articles, chosen from among the assigned *empirical* studies. All eligible readings are identified in the schedule with three starts (***). The peer review should be 800-1000 words in length and provide a critical evaluation of the strengths and weaknesses of the reported research. You should start the review with a brief summary (~ 1 paragraph) of the overall research goals and findings, then spend the rest of the review on an evaluation of the research design. You must submit your peer review by 7pm the evening before the class session (i.e., Sunday evening).
- **Final paper** (*40%* of final grade): Students will develop an original research project that will apply one or more of the computational tools you learn in class to a topic and dataset of your choosing. By the end of the course, you will produce a write-up of the project (~15-20 pages). Details about the Final Paper will be provided in class.

Late submission policy Late submission will result in a 5% deduction for each day the assignment is late (starting with the day the assignment is due), unless you have a legitimate, documented reason beyond your control.

Other Important Information

Accessibility needs

It is the University of Toronto's goal to create a community that is inclusive of all persons and treats all members of the community in an equitable manner. In creating such a community, the University aims to foster a climate of understanding and mutual respect for the dignity and worth of all persons. Please see the University of Toronto Governing Council "Statement of Commitment Regarding Persons with Disabilities".

In working toward this goal, the University will strive to provide support for, and facilitate the accommodation of individuals with disabilities so that all may share the same level of access to opportunities, participate in the full range of activities that the University offers, and achieve their full potential as members of the University community. We take seriously our obligation to make this course as welcoming and accessible as feasible for students with diverse needs. We also understand that disabilities can change over time and will do our best to accommodate you.

Students seeking support must have an intake interview with a disability advisor to discuss their individual needs. In many instances it is easier to arrange certain accommodations with more advance notice, so we strongly encourage you to act as quickly as possible. To schedule a registration appointment with a disability advisor, please visit Accessibility Services, call at 416-978-8060, or email at: accessibility.services@utoronto.ca. The office is located at 455 Spadina Avenue, 4th Floor, Suite 400. Additional student resources for distressed or emergency situations can be located at distressedstudent.utoronto.ca; Health & Wellness Centre, 416-978-8030, http://www.studentlife.utoronto.ca/hwc, or Student Crisis Response, 416-946-7111.

Academic integrity

Copying, plagiarizing, falsifying medical certificates, or other forms of academic misconduct will not be tolerated. Any student caught engaging in such activities will be referred to the Dean's office for adjudication. Any student abetting or otherwise assisting in such misconduct will also be subject to academic penalties. Students are expected to cite sources in all written work and presentations. See this link for tips for how to use sources well.

According to Section B.I.1.(e) of the Code of Behaviour on Academic Matters it is an offence "to submit, without the knowledge and approval of the instructor to whom it is submitted, any academic work for which credit has previously been obtained or is being sought in another course or program of study in the University or elsewhere." By enrolling in this course, you agree to abide by the university's rules regarding academic conduct, as outlined in the Calendar. You are expected to be familiar with the Code of Behaviour on Academic Matters and Code of Student Conduct which spell out your rights, your duties and provide all the details on grading regulations and academic offences at the University of Toronto.

Equity and diversity

The University of Toronto is committed to equity and respect for diversity. All members of the learning environment in this course should strive to create an atmosphsnere of mutual respect. As a course instructor, I will neither condone nor tolerate behaviour that undermines the dignity or self-esteem of any individual in this course and wish to be alerted to any attempt to create an intimidating or hostile environment. It is our collective responsibility to create a space that is inclusive and welcomes discussion. Discrimination, harassment and hate speech will not be tolerated.

Additional information and reports on Equity and Diversity at the University of Toronto is available at http://equity.hrandequity.utoronto.ca.

Class Schedule

Important: class readings are subject to change, contingent on mitigating circumstances and the progress we make as a class. Students are encouraged to attend class and check the course website for updates.

Week 01, 09/12 - 09/16: Introduction and the syllabus

Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." *Annual Review of Sociology* 46(1):61–81. doi: 10.1146/annurevsoc-121919-054621.

Wilson et al. 2017. "Good enough practices in scientific computing." *PLOS Computational Biology* 13(6).

Lab component: Setting up R/RStudio and Python/Jupyter Notebooks

Week 02, 09/19 - 09/23: Foundational matters

Bit by Bit. Preface & Chapter 1.

Lazer, D., et al. 2009. "Computational Social Science." *Science* 323(5915):721–23. doi: 10.1126/science.1167742.

Conte R. et al. 2012. "Manifesto of Computational Social Science." *The European Physical Journal Special Topics* 214:325-346.

Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1):3-42.

Recommended:

Watts, Duncan. 2017. "Should social science be more solutions-oriented?" *Nature Human Behaviour*.

Kang, Donghyun and James Evans. 2020. "Against Method: Exploding the Boundary between Qualitative and Quantitative Studies of Science." *Quantitative Science Studies* 1(3):930-944.

Lab component: Data wrangling (R and the tidyverse)

Week 03, 09/26 - 09/30: Observing behaviour I

Bit by Bit. Chapter 2.

King, Pan, and Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(2):326-343. ***

Aral and Nicolaides. 2017. "Exercise Contagion in a Global Social Network." *Nature Communications*. ***

Lab component: Data collection (reading data in R and API basics)

Week 04, 10/03 - 10/07: Observing behaviour II

Wang, Phillips, Small, and Sampson. 2018. "Urban mobility and neighborhood isolation in America's 50 largest cities." *PNAS*. ***

Hersh, Eitan. 2013. "Long-term effect of September 11 on the political behavior of victims' families and neighbors." *PNAS*. ***

Lab component: More APIs and web scraping with Python

Week 05, 10/10 - 10/14: Thanksgiving (No class)

Week 06, 10/17 - 10/21: Computational statistics and machine learning I

Molina and Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology*.

ISL. Chapters 2 & 5.

Recommended:

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24:395-419.

Lab component: Resampling methods and an introduction to supervised learning

Week 07, 10/24 - 10/28: Computational statistics and machine learning II

ISL. Chapter 12.

Garip, Feliz. 2012. "Discovering Diverse Mechanisms of Migration: The Mexico-US Stream 1970-2000." *Population Development Review* 38(3):393-433. ***

Lab component: Clustering and unsupervised learning

Week 08, 10/31 - 11/04: Text as data I

Grimmer and Stewart. 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*.

Nelson, Burk, Knudsen, and McCall. 2018. "The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods." *Sociological Methods & Research* 50(1): 202-237.

Rathje, Van Bavel, and van der Linden. 2021. "Out-Group Animosity Drives Engagement on Social Media." *PNAS*. ***

Recommended:

Evans, James A., and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42(1):21–50.

Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." *Theory and Society* 43(3):465–82.

Lab component: "Text as data" Part 1

Week 09, 11/07 - 11/11: Text as data II

Roberts et al. 2014. "Structural topic models for open-ended survey responses." *American Journal of Political Science*.

Kozlowski, Taddy, and Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review*. ***

Lab component: "Text as data" Part 2

Week 10, 11/14 - 11/18: Agent based models

Shelling, Thomas. 1978. *Micromotives and Macrobehavior*. Chapter 1.

Bruch and Atwell. 2015. "Agent-Based Models in Empirical Social Research." *Sociological Methods and Research*.

Play through the interactive game at the following link: https://ncase.me/polygons/

Goldberg, Amir and Sarah K. Stein. "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation". *American Journal of Sociology* 83(5): 897-932. ***

Recommended:

DellaPosta and Davoodi's comment on Goldberg and Stein and Goldberg's reply

Keuschnigg, Marc, Niclas Lovsjö, & Peter Hedstöm. 2017. Analytical Sociology and Computational Social Science." *Journal of Computational Sociaal Science* 1:3-14.

Lab component: ABM example

Week 11, 11/21 - 11/25: Project work/coding session

This is an opportunity to work on your project and ask questions of instructor and fellow students

Week 12, 11/28 - 12/02: Digital experiments

Bit by Bit. Chapter 4.

Salganik, Dodds, and Watts. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science*. ***

Bail et al. 2018. "Exposure to opposing views on social media can increase political polarization." *PNAS*. ***

Lab component: Exploring digital experiment tools

Week 13, 12/05 - 12/09: Ethics

Bit by Bit. Chapter 6.

Kramer, Guillory, and Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks." *PNAS*. (and Editorial Expression of Concern: https://doi.org/10.1073/pnas.1412469111) ***

O'Neil, Cathy. *Weapons of Math Destruction* (selection)

Lab component: Effective data visualization