SOC497H1-S Winter 2022 | Department of Sociology | University of Toronto

Computational Methods in Sociology

Instructor: Fedor A. Dokshin, PhD

E-mail: fedor.dokshin@utoronto.ca

Office: Room 372 at Sociology Dept.
725 Spadina Ave, 3rd floor

Office Hours: TBA

Class: LEC0101: Tue, 9am to 12pm
Location: FE 41 (online until further notice)

Class website: https://q.utoronto.ca

Course Description and Aims

New forms of digital data present enormous new 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. This course introduces students to a set of computational tools and their applications to question of sociological interest. It takes a practical approach, starting with the basics of programming and data management and then working through a series of computational tools and data examples.

Prerequisite

The prerequisite to take this course is 1.0 SOC FCE at the 300+ level. Students without the prerequisite can be removed at any time discovered, and without notice.

Learning Components and Course Requirements

Required text:

Salganik, Matthew J. 2019. *Bit by Bit: Social Research for the digital Age*. Princeton University Press. (Purchase or read online here: https://www.bitbybitbook.com/).

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

Course 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 weekly session will be composed of four parts:

1. Introduction to the topic (~30 minutes)
2. Discussion of research (~60 minutes)
3. Guided coding session (~60 minutes)
4. Coding challenge (~30 minutes)

Professor Dokshin will lead the first part, providing some background for the topic of the day. In the second part, we will discuss research articles that apply these methods. In the third part, the instructor will lead a guided coding exercise designed to learn basic computational skills or a particular methodological application. Each coding session will end with a "coding challenge," which asks the student to apply the tools learned during the guided session to a new problem. Time permitting, you will start working on the challenge during class time. You will continue working on the coding challenge after class and submit the completed challenge before the beginning of next class.

Article peer reviews and presentations:

Each student will write **peer reviews of two published research articles**, chosen from among the assigned readings. 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., Monday evening).

You are responsible for **presenting one of your peer reviews to the class** to facilitate discussion. A sign-up sheet will be circulated in the first week of class. Your presentation should summarize the key points from your review and raise specific questions for further class discussion. In your presentation, be sure to critically consider what computational approaches add to research on this specific topic? How do they improve upon existing measurement and analytic strategies in this topic area? What are the "big picture" opportunities and challenges that these methods present for social science research? The presentation constitutes one part of your participation grade.

Coding challenges:

You are required to submit **at least 6** completed coding challenges throughout the semester. These will not be graded for accuracy, but must be complete. A complete coding challenge consists of the two files, .Rmd file and a compiled html/pdf file.

Final project:

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 (10-15 pages). The write-up must include the following components:

- Clearly state the research objective
- Situate your effort within existing knowledge
- Describe your data and methodological approach
- Present results from your analysis
- Conclude with lessons learned and future directions

At least once before March 15th, you are required to meet with the instructor to discuss your project idea. This will allow the instructor to provide early feedback and help keep everyone stay on track.

In-class participation:

This is a seminar course and active discussion is the main mode of learning. You are expected to come to class prepared and participate actively in discussion. Your participation grade, assigned at the end of the semester, will reflect the quality of the peer review presentation, your attendance, and level of engagement in the discussion.

Software:

We will use R in this course, because it is (1) the industry standard for computing in the social sciences, (2) offers powerful tools to conduct diverse computational tasks, (3) has extensive documentation and online help resources, and (4) is free and open-source. We will use RStudio, a popular, easy to use Interactive Development Environment (IDE).

We will use UofT's JupyterHub, which makes it easy to access RStudio through a browser window and does not require you to install and configure additional software locally.

You will need access to a computer for this class. This mean you will need to bring a laptop to class, if we return to in-person teaching. If you do not own a laptop, the University libraries has some available to borrow (details: https://onesearch.library.utoronto.ca/technology).

Acknowledgements:

I also want to acknowledge specific individuals whose publicly shared materials have contributed to this class, directly and indirectly. This list includes: Rohan Alexander, Pablo Barbera, Matt Salganik, Rochelle Terman, and Tiffany-Anne Timbers,

Trevor Campbell, and Melissa Lee, whose *Data Science: A First Introduction* serves as the source material for the first several weeks of lab tutorials.

Evaluation Components

	Number of occasions	Percent value	Total percent of final mark
Coding challenges	6	5% each	30%
Article peer review	2	10% each	20%
Final project	1	35%	35%
In-class participation	Throughout semester	15%	15%
			100% (total)

Communication and Quercus

Email:

Please use your U of T email address to communicate with me about **personal matters**, or to communicate with the TAs. I make every effort to reply to emails within 24 hours, Mon-Fri 9am-5pm. Here are a couple of important points about email communication:

- Please note that the instructor and TA will not respond to emails about issues that are clearly specified in the syllabus (e.g., due dates, office hours times).
- Please use the discussion board to ask questions about course content (see below).
- Address your questions about tutorials to your TA.
- Requests for make-up tests and other accommodations should be sent to the course instructor (Professor Dokshin), not the TA.
- All emails should include the course code SOC202 in the subject line, and be signed with the student's full name and student number.

Quercus website and discussion board:

Quercus will contain the course syllabus, assignments, discussion board, and course announcements. To clarify **questions** regarding the **syllabus, assignments**, as well as **substantive questions about the readings**, please use the designated discussion boards on the Quercus site.

Office hours:

The instructor will hold online office hours through Zoom (Link will be posted in Quercus calendar).

Procedures for Late Work

Documentation:

If you are unable to turn in an assignment for medical reasons, you will need to **email the instructor and also** declare your absence on ACORN. Further, you must present one of the following documents to request an extension or another consideration:

1. **College registrar's letter:**
    - If a **personal or family crisis** prevents you from meeting a deadline, have your college registrar email me directly (it is a good idea anyway to advise your college registrar if a crisis is interfering with your studies).
2. **Letter from Accessibility Services**
    - This documentation is useful for ongoing medical issues that require special accommodation.

Academic Integrity

Academic integrity is fundamental to learning and scholarship at the University of Toronto. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that the U of T degree that you earn will be valued as a true indication of your individual academic achievement, and will continue to receive the respect and recognition it deserves.

Familiarize yourself with the University of Toronto's Code of Behaviour on Academic Matters (http://www.governingcouncil.utoronto.ca/policies/behaveac.htm). It is the rule book for academic behaviour at the U of T, and you are expected to know the rules. Potential offences include, but are not limited to:

- Obtaining or providing unauthorized assistance on any assignment including:
  - working in groups on assignments that are supposed to be individual work;
  - having someone rewrite or add material to your work while "editing";
  - crowdsourcing assignment answers through Facebook or another forum.
- Lending your work to a classmate who submits it as his/her own without your permission.
- Using or possessing any unauthorized aid, including a cell phone.
- Looking at someone else's answers.
- Letting someone else look at your answers.
- Misrepresenting your identity.
- Submitting an altered test for re-grading.
- Falsifying or altering any documentation required by the University, including doctor's notes.
- Falsifying institutional documents or grades.

Accessibility Needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom or course materials, please contact Accessibility Services as soon as possible: disability.services@utoronto.ca or http://studentlife.utoronto.ca/accessibility.

Course Schedule

Week	Date	Topic & Reading	Assignments Due
1	11-Jan	**Introduction** Evans and Foster. 2019. "Computation and the sociological imagination." *Contexts*. *Lab component:* Accessing RStudio through UofT's JupyterHub and intro to RStudio interface	
2	18-Jan	**Foundational matters** Bit by Bit. Preface & Chapter 1. Anderson. 2008. "The end of theory: The data deluge makes the scientific method obsolete." *Wired*. boyd and Crawford. 2012. "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon." *Information, Communication, & Society*. Watts. 2017. "Should social science be more solutions-oriented?" *Nature Human Behaviour*. *Lab component:* "R and the tidyverse"	
3	25-Jan	**Observing behaviour I** Bit by Bit. Chapter 2. King, Pan, and Roberts. 2014. "Reverse-engineering censorship in China: Randomized experimentation and participant observation." *Science*. *** Aral and Nicolaides. 2017. "Exercise contagion in a global social network." *Nature Communications*. *** *Lab component:* "Reading in data locally and from the web"	– Peer review by 7pm Monday before class – Coding challenge due before class
4	01-Feb	**Observing behaviour II** Lazer, Hargittai, Freelon, Gonzalez-Bailon, Munger, Ognyanova, & Radford. (2021). "Meaningful measures of human society in the twenty-first century." *Nature*. Wang, Phillips, Small, and Sampson. 2018. "Urban mobility and neighborhood isolation in America's 50 largest cities." *PNAS*. *** Hersh. 2013. "Long-term effect of September 11 on the political behavior of victims' families and neighbors." *PNAS*. *** *Lab component:* "Cleaning and wrangling data" Part 1	– Peer review by 7pm Monday before class – Coding challenge due before class

Week	Date	Topic & Reading	Assignments Due
5	08-Feb	**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:* "Cleaning and wrangling data" Part 2	– Peer review by 7pm Monday before class – Coding challenge due before class
6	15-Feb	**New ways of asking** Bit by Bit. Chapter 3. Roberts et al. 2014. "Structural topic models for open-ended survey responses." *American Journal of Political Science*. *** Lazer, Kennedy, King, and Vespignani. 2014. "The parable of google flu: Traps in big data analysis." *Science*. *Lab component:* "Effective data visualization" Part 1	– Peer review by 7pm Monday before class – Coding challenge due before class
Reading week—No Class			
7	01-Mar	**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:* "Effective data visualization" Part 2	– Peer review by 7pm Monday before class – Coding challenge due before class
8	08-Mar	**Text data I** Bail. 2014. "The cultural environment: measuring culture with big data." *Theory and Society*. 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 1	– Peer review by 7pm Monday before class – Coding challenge due before class
9	15-Mar	**Text data II**	– Peer review by 7pm Monday before class

Week	Date	Topic & Reading	Assignments Due
		Grimmer and Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis*. Rathje, Van Bavel, and van der Linden. 2021. "Out-group animosity drive engagement on social media." *PNAS*. *** Terman. 2017. "Islamophobia and Media Portrayals of Muslim Women: A Computational Text Analysis of US News Coverage." *International Studies Quarterly*. *** <u>*Lab component:*</u> "Text as data" Part 2	– Coding challenge due before class
10	22-Mar	**Network analysis** Light and Moody. 2020. "Network basics: points, lines, and positions" in *The Oxford Handbook of Social Networks*. Mukerjee et al. 2018. "Networks of audience overlap in the consumption of digital news." *Journal of Communication*. *** Farrell. 2015. "Corporate funding and ideological polarization about climate change." *PNAS*. *** <u>*Lab component:*</u> Introduction to network analysis	– Peer review by 7pm Monday before class – Coding challenge due before class
11	29-Mar	**Guided work on projects**	
12	05-Apr	**Project presentations**	