SOC576
COMPUTATIONAL SOCIOLOGY
Syllabus

Thomas Davidson

Spring 2021

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 supervised 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.

Chris Bail (a sociologist at Duke University) recently published a series of video tutorials (https://sicss.io/boot_camp) and exercises designed to introduce these tools. I highly recommend that students watch these videos and review the relevant materials in R for Data Science (details on the book below) prior to the beginning of the semester.
ASSESSMENT

There will be two types of assessment used in the course. There will be several homework assignments throughout the semester that will be designed to enable students to become familiar with the various methodological techniques covered in the course. A tentative 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. 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 I will provide feedback on the paper proposal, the methodology for data collection, and the data analysis (see course outline for a timeline).

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 and useful readings

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

- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. An Introduction to Statistical Learning. Springer Texts in Statistics. (Entire book can be downloaded as a PDF via Rutgers Library)

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.
I will also be making additional accommodations due to the COVID-19 pandemic. If you or your family are affected in any way that impedes your ability to participate in this course, please contact me as soon as you can so that we can make necessary arrangements.

**COURSE OUTLINE**

*This outline is tentative and subject to change.*

**Week 1**

**Introduction to Computational Sociology**

*Readings*

- *R4DS*: Preface, C2-6
- *Bit by Bit*, C1

**Week 2**

**Data Structures**

*Readings*

- *R4DS*: C7-11

**Week 3**

**Programming Fundamentals**

*Readings*

- *R4DS*: C14-17, 21-24

*Assignment 1: The computational toolkit.*

**Week 4**

**Data Collection I: APIs**

*Readings*
Note: I have included readings that use a range of different APIs including Reddit (Hamilton et al.), Spotify (Askin and Mauskampf), Facebook (Davidson and Berezin), Google Trends (Davidson and Berezin; Bail, Brown, and Wimmer), Twitter (Mitts), and YouTube (Munger and Phillips). My intent is for you to see the breadth of research making use of APIs rather than to go in-depth into any one study. Make sure to read Deen Freelon’s article on the “post-API age.”


Week 5
Data Collection II: Webscraping
Readings

Assignment 2: Collecting and storing data

Week 6
Data Collection III: Online experiments and surveys
Readings

Paper proposal due

Week 7

Natural Language Processing I: Fundamentals

Readings
• Introduction to Information Retrieval, pp. 117-126.

Week 8

Natural Language Processing II: Word Embeddings

Readings
Week 9

Natural Language Processing III: Topic Models

Readings


Assignment 3: Natural language processing

Week 10

Machine Learning I: Prediction and explanation

Readings


Paper initial data collection due

Week 11

Machine learning II: Text classification

Readings


**Week 12**

**Machine learning III: Challenges**

*Readings*


**Week 13**

**Machine learning IV: Image classification**

*Readings*

Assignment 5: Agent-based models

Work on final projects

Final paper due (date to be confirmed)