SOC 225: Data and Society

Summer 2018

Instructor: Connor Gilroy
Email: cgilroy@uw.edu
Office: SAV 216B
Office hours: Wed 1:00-2:30; Thurs 3:20-3:50

Meeting times: Tues 1:10-3:20 & Thurs 1:10-2:10 (lecture); Thurs 2:20-3:20 (lab)
Location: PAR 310 (lecture); SAV 121 (lab)

Course Description

Through technologies from smartphones to social media, we produce digital traces of our everyday lives. Our traces become social data, collected at a massive scale and in unprecedented detail for individuals, corporations, and governments to analyze. The process of producing and using this "big data" has society-wide consequences.

This course is about the social transformations brought on by the digital data revolution. How might the proliferation of data enable discovery and innovation? How might it generate or reinforce inequality? We will consider our own roles as both producers and consumers of data, including as ethical practitioners of "data science."

This is an introductory course, appropriate for all students who want to understand how data shape our lives and our society. It has two components, a lecture and a lab. The lecture explores how digitization changes institutions and social relations. The lab complements the lecture by making issues in working with digital data more concrete, while allowing students to develop their analytic skills in data manipulation. There is no expectation of prior coding experience for the labs.

Objectives

Upon completing this course, students will be able to

- understand how individual records become aggregate data
- analyze how these data reshape social relations and institutions
- reflect on the ethical implications of data collection and data science
- apply modern data analysis techniques to explore digital trace data

Requirements

Grading overview

```
Reading responses          20%
Class participation        15%
Final paper                25%
[Lecture                   60%]
-------------------------------
Lab exercises              25%
Data project               15%
[Lab                       40%]
-------------------------------
[Total                    100%]
```

Timeline

Each week, you are responsible for:

- A reading response (8 total)
- In-class participation (8 total)
- Lab exercises (8 total)

The final paper and the data project are broken down into smaller milestones due throughout the quarter. These are generally **due on Thursdays**, before or in class.

```
Week 3                     Paper: crowdsourced example issues
Week 4                     Data project: proposed data source
Week 5                     Paper: proposal
Week 6                     Data project: code & peer code review
Week 7                     Paper: draft & peer review
Week 8                     Data project: write up
Week 9                     Paper: final draft and presentation
```

More details of these assignments are below. Rubrics for the two major assignments will be provided.

Lecture (60% total)

Reading responses (20%)

You will get the most out of this class if you consistently do the readings and come to class ready to engage with them and with each other. This part of your grade is meant to encourage engagement with the reading materials, and to stimulate conversation inside and outside of class.

Each week, make a post responding to at least one of the week's readings with reactions, connections, and questions. Do not simply summarize the readings. Your reading response should be no more than 100 words. **8 reading responses will give you full credit.** There are 9 weeks in the quarter, so *you may opt out of one week's reading response with no penalty.*

You can respond to either Tuesday or Thursday readings. Whichever day's readings you choose, your response is **due online by noon on that day.** I encourage you to read other students' responses, and reply to them if you have constructive commentary.

During the first class, **we will decide collectively** what technology platform we want to use for reading responses. Four options to consider are Canvas, Slack, Twitter, and Facebook, but I am open to other suggestions. We will consider ease of communication, accessibility, privacy, and assessment.

Class participation (15%)

Class participation is a way for you to contribute to a constructive learning environment, for yourself and for your peers. I recognize and value different forms of participation. Each week (on either Tuesday or Thursday), I will assign brief free-writes, small-group discussion topics, or other in-class activities. Participation in **8 activities will give you full credit.** There are 9 weeks in the quarter, so *you may opt out of one week's in-class participation with no penalty.*

Final paper (25%)

For the term paper, you will apply what you've learned in class to new situations. Use the theories and concepts from the course to analyze 1-3 examples, drawn from recent media articles, of data or technology leading to social change. This could be a social problem, or a social benefit.

To encourage you to work steadily throughout the quarter, and make it easier to complete at the end, the paper is divided into a number of milestones with distinct due dates:

- Week 3: crowdsourced examples (3)
- Week 5: paper proposal (< 1 page)
- Week 7: rough draft and peer review (3 pages)
- Week 9: final draft (5 pages)

Your paper is due on Canvas by **noon on Thursday, 8/16**. Bring a printed copy to class, and come prepared to talk about your paper to the class for three to five minutes.

Lab (40% total)

Lab exercises (25%)

There will be 8 lab activities over the course of the quarter. Generally, you should be able to complete or make substantial progress on these during the lab session. Your write-ups will be due **the following Thursday, before class** to allow you time to practice working with data.

Each assignment will include one or more **challenge problems**, which you can attempt for extra credit. If you miss a lab section, the challenge problems will mitigate that. The total number of points you can receive for the lab exercises is capped at 25.

Data project (15%)

You will demonstrate and synthesize your lab-based skills by producing a visualization, summary, and interpretation of some digital data source. This could be a relatively close replication of a figure from an existing study, or a more creative application. One approach you might take would be to write your final paper and data project on the same topic, but you can also choose to pursue two things that are unrelated.

Like the paper, this project is divided into multiple deliverables:

- Week 4: proposed question and data source
- Week 6: peer check in
- Week 8: blog post write up (2 pages)

Schedule

The nine weeks of the quarter are divided into three parts. *Foundations* establishes concepts and definitions and lays the groundwork for more complex and advanced conversations. *Issues* provides more conceptual tools for analyzing the social problems arising from big data and digital technology. *Applications* examines how different sectors of society have changed through digital data and technology. These parts of the course are cumulative and build on each other; you will need to remember and apply ideas from earlier parts of the course to later content. I may make substitutions or adjustments to the readings as the quarter goes along, but I will announce changes at least a week in advance.

The second half of class on Thursdays will be a hands-on lab activity. We will meet in the CSSCR computing lab. You do not need to use your own laptop for the labs. Generally, but not always, these activities will relate to the lectures for the week.

Part I: Foundations

Week 1: Defining and studying big data

Why do we need to think about data and society now?

Tuesday, 6/19

Wallach, Hanna. 2014. "Big Data, Machine Learning, and the Social Sciences." Medium. (link)

Recommended: David Lazer and Jason Radford. 2017. "Data Ex Machina: Introduction to Big Data." Annual Review of Sociology 43(1).

Thursday, 6/21

boyd, danah and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." Information, Communication & Society 15(5):662–79.

Neff, Gina, Anissa Tanweer, Brittany Fiore-Gartland, and Laura Osburn. 2017. "Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science." Big Data 5(2):85–97.

Lab: Data, files, and digital objects

Week 2: Digital life

How have our lives been digitized and quantified?

Tuesday, 6/26

Gregory, Karen, Tressie McMillan Cottom, and Jessie Daniels. 2017. Introduction, from *Digital Sociologies*. (link)

Lupton, Deborah. 2017. "Personal Data Practices in the Age of Lively Data." Pp. 339–54 in *Digital Sociologies*. (link)

Albury, Kath, Jean Burgess, Ben Light, Kane Race, and Rowan Wilken. 2017. "Data Cultures of Mobile Dating and Hook-up Apps: Emerging Issues for Critical Social Science Research." Big Data & Society 4(2).

Thursday, 6/28

Jurgenson, Nathan. 2011. "Digital Dualism versus Augmented Reality." Cyborgology. (link)

Wickham, Hadley. "Tidy Data." CRAN vignette for the `tidyr` package. (link)

Lab: Data manipulation and tidy data

Part II: Issues

Week 3: Surveillance and privacy

How does big data interact with institutions of power and control?

Tuesday, 7/3

Brayne, Sarah. 2017. "Big Data Surveillance: The Case of Policing." American Sociological Review 82(5):977–1008.

Levy, Karen E. C. 2015. "The Contexts of Control: Information, Power, and Truck-Driving Work." The Information Society 31(2):160–74.

Watch this ten-minute video on Karen Levy's work: "Automation is coming for truckers. But first, they're being watched." Christophe Haubersin. 2017. Vox. (link)

Thursday, 7/5

Examples for paper due before class

Nissenbaum, Helen. 2011. "A Contextual Approach to Privacy Online." Daedalus 140(4):32–48.

Tufekci, Zeynep. 2018. "The Latest Data Privacy Debacle." The New York Times, January 30. (link)

Lab: Data visualization and mapping

Week 4: Algorithmic bias and discrimination

What social decisions are encoded in algorithms, and when are they unfair?

Tuesday, 7/10

O'Neil, Cathy. 2016. "Introduction" and "Chapter 1. Bomb Parts: What Is a Model?" from *Weapons of Math Destruction*.

Agüera y Arcas, Blaise, Alexander Todorov, and Margaret Mitchell. 2018. "Do Algorithms Reveal Sexual Orientation or Just Expose Our Stereotypes?" Medium. (link)

Lohr, Steve. 2018. "Facial Recognition Is Accurate, If You're a White Guy." The New York Times, February 9. (link)

Watch Joy Buolamwini's ten-minute TED Talk, "How I'm Fighting Bias in Algorithms." November 2016. (link)

Recommended: Stitch Fix Algorithms Tour. http://algorithms-tour.stitchfix.com/

Thursday, 7/12

Proposed data source for data project due after class

Adler-Bell, Sam. 2018. "The High-Tech Poorhouse: An Interview with Virginia Eubanks." Jacobin, January 29. (link)

O'Neil, Cathy. 2016. "How Algorithms Rule Our Working Lives." The Guardian, September 1. (link)

Recommended: Eubanks, Virginia. 2018. "A Child Abuse Prediction Model Fails Poor Families." Excerpt from *Automating Inequality*. Wired. (link)

Lab: Exploratory data analysis

Week 5: Ethics and experimentation

What ethical obligations do researchers and companies have regarding digital data?

Tuesday, 7/17

Amrute, Sareeta. 2018. "What Would A Techno-Ethics Look Like?" Platypus. (link)

Salganik, Matthew J. 2017. "Ethics." Chapter 6 from *Bit by Bit: Social Research in the Digital Age*. (link)

Thursday, 7/19

Paper proposal due before class

Angwin, Julia and Terry Parris Jr. 2016. "Facebook Lets Advertisers Exclude Users by Race." ProPublica. (link)

Matias, J. Nathan. 2016. "The Obligation To Experiment." MIT Media Lab, Medium. (link)

Matias, J. Nathan. 2018. "Attributing Cause in Algorithm Audits." Medium. (link)

Lab: Ads, experiments, and audits

Part III: Applications

Week 6: Internet as community and infrastructure

What has the Internet done for the social world?

Tuesday, 7/24

Barlow, John Perry. 1996. "A Declaration of the Independence of Cyberspace." Electronic Frontier Foundation. (link)

Nakamura, Lisa. 2014. "Gender and Race Online." From *Society and the Internet: How Networks of Information and Communication are Changing Our Lives*.

Choose at least one of the following to read:

Alptraum, Lux. 2018. "Industry Standards: How the Internet Changed Sex Work." Real Life. (link)

Baym, Nancy. 2018. “Book Excerpt: How Music Fans Built the Internet.” Wired, July 10. (link)

Thursday, 7/26

Data project peer check-in in class

Plantin, Jean-Christophe, Carl Lagoze, Paul N. Edwards, and Christian Sandvig. 2016. “Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook.” New Media & Society 20(1):293–310.

Lab: Web scraping and APIs

Week 7: Politics and social media

From social movements to fake news, how have social media reshaped politics?

Tuesday, 7/31

Zeynep Tufekci. 2017. “Chapter 1. A Networked Public” and “Chapter 6. Platforms and Algorithms” from *Twitter and Tear Gas*. (link) [Zeynep has made the entire book available for free online.]

Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. 2017. “From Liberation to Turmoil: Social Media And Democracy.” Journal of Democracy 28(4):46–59.

Thursday 8/2

Paper draft peer review in class

Look at this visualization of Reddit: http://rhiever.github.io/redditviz/clustered/

Starbird, Kate. 2017. “Information Wars: A Window into the Alternative Media Ecosystem.” Medium. (link)

Lab: Twitter and social networks

Week 8: Attention and search

How has our attention been commodified through search and recommendations?

Tuesday, 8/7

Watch Zeynep Tufekci's twenty-three minute TED Talk, "We're Building a Dystopia Just to Make People Click on Ads." September 2017. (link)

Noble, Safiya. 2013. "Google Search: Hyper-Visibility as a Means of Rendering Black Women and Girls Invisible." InVisible Culture: An Electronic Journal for Visual Culture (19). (link)

Tufekci, Zeynep. 2018. "YouTube, the Great Radicalizer." The New York Times, March 10. (link)

Mellon, Jon. 2018. "Everybody Lies but Google Trends Is Not a Panacea." Blog post. (link)

Lab: Google Trends

Thursday, 8/9: Writing day, no class

Data project due Thursday by the end of the day

Week 9: The platform economy

How have data and technology "disrupted" sectors of the economy?

Tuesday, 8/14

Ingold, David and Spencer Soper. 2016. "Amazon Doesn't Consider the Race of Its Customers. Should It?" Bloomberg. (link)

Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." American Economic Journal: Applied Economics 9(2):1–22.

Thursday, 8/16

Final paper due in class

Banks, David A. 2018. "Engineered for Dystopia." The Baffler. (link)

Case, Nicky. 2018. "How To Become A Centaur." Journal of Design and Science. (link)

Read this narration of gendered language in course evaluations: http://benschmidt.org/profCloud/

Policies

Email

I will respond to emails within **two business days** (48 hours). I do not guarantee responses in the evenings or on weekends, so please contact me in advance of deadlines. It will help me if you **include "SOC 225" in the subject** of your email. Clear, concise questions and requests are appreciated. For more complicated questions, especially related to code and data, I encourage you to come to my office hours.

Late assignments

I expect assignments to be turned in on time. Late papers and projects will be penalized 1 point for each day they are late. If there are extenuating circumstances, such as illness or loss, you should notify me as soon as possible. This is especially important for assignments involving peer review. I cannot accept unexcused late work for reading responses, in-class work, or lab exercises.

Classroom environment

Please work to create a respectful environment for your fellow students. Refrain from discriminatory language, and recognize that students in the class have a diversity of backgrounds. Students will have a range of prior knowledge of computational skills, and a range of familiarity with sociological concepts. The course does not presume any background in either.

The evidence on classroom use of laptops is mixed. I permit laptop use for taking notes and for referring to readings, but ask that you be mindful of the effects of technological distractions on yourself and your peers. I also ask that you only check cell phones during breaks, or step out of the classroom to do so.

Accommodations

I want this class to be an accessible learning experience for students. If you have already established accommodations with Disability Resources for Students (DRS), please share your approved accommodations with me as soon as possible, so I can work with you to accommodate your learning-related needs. If you have not yet established services through DRS, but have a temporary health condition or permanent disability that requires accommodations, you are welcome to contact DRS at 206-543-8924 or uwdrs@uw.edu or disability.uw.edu.

Academic integrity

Please be sure you understand the University's guidelines on academic integrity (link). Instances of suspected plagiarism or cheating will be formally reported to the University.

Attribution

This syllabus builds on syllabi and courses by Anna Lauren Hoffman, Afra Mashhadi, Tyler McCormick, Sarah Quinn, and Emilio Zagheni.

Further resources

Data & Society https://datasociety.net/
AI Now Institute https://ainowinstitute.org/
Critical Algorithms Studies reading list https://socialmediacollective.org/reading-lists/critical-algorithm-studies/

R for Data Science http://r4ds.had.co.nz/
Data Visualization http://socviz.co/