

# SOC 225: Data and Society

*Summer 2019*

**Instructor:** Connor Gilroy

**Email:** [cgilroy@uw.edu](mailto:cgilroy@uw.edu)

**Office:** SAV 216B

**Office hours:** Wed 10:30-12:00 & 3:20-3:50, or by appointment

**Meeting times:** Mon & Wed 1:10-2:10 (lecture); Mon & Wed 2:20-3:20 (lab)

**Location:** SAV 164 (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 pay particular attention to how these social changes play out by gender, race, class, and sexuality. We will also 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

Lecture participation	20%
Reading responses (2x)	10%
Analytical memos (3x)	30%
[Lecture	60%]
-----	
Lab participation	15%
Lab homework (3x)	15%
Data project	10%
[Lab	40%]
-----	
[Total	100%]

### Timeline

This is the timeline of course assignments:

Week 3	Lab homework 1
Week 4	Analytical memo 1
Week 5	Lab homework 2
Week 6	Analytical memo 2
Week 7	Lab homework 3
Week 8	Analytical memo 3
Week 9	Data project writeup and presentation

Assignments are generally **due on Thursdays at midnight**. More details of these assignments are below. Rubrics for the memos and data project will be provided.

### Lecture (60% total)

#### Lecture participation (20%)

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.

I will assign in-class activities including reading quizzes, free-writes, and small-group discussions. There will be approximately 10-12 graded activities over the course of the quarter, and your lowest two grades will be dropped with no penalty.

### **Reading responses (10%)**

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 student will be assigned two dates during the quarter to bring an extended reading response to class. Your responses should contain reactions to the day's readings, making connections to the themes and ideas of the course and posing questions for discussion. You'll be asked to share a summary of your response with the class and pose those questions to the group.

Each response should be between 250-300 words long, which is about 1 double-spaced page.

### **Analytical memos (30%)**

You will write 3 analytical memos during the quarter, responding to prompts that ask you to apply what you've learned in class to new situations. These memos will use theories and concepts from the course to analyze examples of social change related to data and technology.

Each memo should be between 500-600 words long, which is about 2 double-spaced pages.

### **Lab (40% total)**

#### **Lab participation (15%)**

During each lab, you'll complete a brief check-in to gauge your understanding of the lab material. Check-in questions will be posted in class, and you'll answer them on Canvas. These check-ins will relate to the lab exercises, but the exercises themselves will **not** be graded. As with lecture participation, you can miss two check-ins with no penalty.

#### **Lab homework (15%)**

The lab homework will involve practical application of the lab content. If you participate and complete the in-class activities, the homework should be straightforward. There will be 3 homework assignments, and parts of each assignment will ask you to work toward your final project.

#### **Data project (10%)**

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.

You will submit your code, your visualization, and a brief writeup explaining your data source, the questions you posed using it, the strategies you adopted for exploratory analysis, and the limitations of your data and visualization.

On the final day of class, you'll present your visualization in a brief (2 minute) "lightning talk."

## Texts

There is one book you're required to purchase for this course, available through the University Book Store and other local bookstores.

**Required text:** Cathy O'Neil. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books; Reprint edition. ISBN 978-0553418835.

**Recommended text:** Zeynep Tufekci. 2018. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press; Reprint edition. ISBN 978-0300234176.

Tufekci and Yale Press have generously made a Creative Commons version of this book freely available as a PDF. You may, however, wish to purchase a physical copy of the book instead.

## 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. *Transformations* 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.

Each day, the second half of class 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? What's different about "big" data?*

**Lab topic:** data, files, and digital objects

#### Monday, 6/24

Course introduction; no readings.

#### Wednesday, 6/26

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

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.

### Week 2: Digital life

*How have our lives been digitized and quantified?*

**Lecture topic:** online dating

**Lab topic:** data manipulation and tidy data

#### Monday, 7/1

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

Fischer, Claude. 1992. "Technology and Modern Life." From *America Calling: A Social History of the Telephone to 1940*. [Read through page 21.]

#### Wednesday, 7/3

Jurgenson, Nathan. 2011. "Digital Dualism versus Augmented Reality." *Cyborgology*. ([link](#))

boyd, danah. 2014. "Identity: why do teens seem strange online?" From *It's Complicated: The Social Lives of Networked Teens*.

## Part II: Issues

### Week 3: Algorithmic bias and discrimination

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

**Lecture topic:** targeted advertising

**Lab topic:** data visualization

#### Monday, 7/8

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

D'Ignazio, Catherine. 2015. "What Would Feminist Data Visualization Look Like?" MIT Center for Civic Media. ([link](#))

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

#### Wednesday, 7/10

WMD: "Chapter 4. Propaganda Machine: Online Advertising."

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

### Week 4: Surveillance and privacy

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

**Lecture topics:** work and policing

**Lab topic:** exploratory data analysis

#### Monday, 7/15

WMD: Chapter 5.

Foucault, Michel. 1975. "Panopticism," from *Discipline and Punish*.

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

**Recommended:** Lum, Kristian and William Isaac. 2016. "To Predict and Serve?" *Significance* 13(5):14–19. ([link](#))

### **Wednesday, 7/17**

WMD: Chapters 6 and 7.

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

- In class, we'll 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](#))

### **Week 5: Ethics and violence**

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

**Lecture topic:** facial recognition

**Lab topic:** ads and audits

### **Monday, 7/22**

WMD: Chapter 10 and Conclusion.

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

**Recommended:** Keyes, Os. 2018. "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition." *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW):1–22.

### **Wednesday, 7/24**

Hoffmann, Anna Lauren. 2018. "Data Violence and How Bad Engineering Choices Can Damage Society." *Medium*. ([link](#))

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

## Part III: Transformations

### Week 6: Internet as community

*What has the Internet done for the social world?*

**Lab topic:** Web scraping and APIs

#### Monday, 7/29

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*.

#### Wednesday, 7/31

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](#))

We'll look at this visualization of Reddit together: <http://rhiever.github.io/redditviz/clustered/>

### Week 7: Politics and social media

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

**Lecture topic:** fake news

**Lab topic:** Twitter and social networks

#### Monday, 8/5

Tufekci, Zeynep. 2017. *Twitter and Tear Gas* ([link](#)) [Zeynep has made the entire book available freely online]:

- "Chapter 1. A Networked Public"
- "Chapter 5. Technology and People"
- "Chapter 6. Platforms and Algorithms"

## Wednesday 8/7

Twitter and Tear Gas: Chapter 7, Chapter 9, and Epilogue

Starbird, Kate. 2017. "Information Wars: A Window into the Alternative Media Ecosystem." Medium. ([link](#))

## Week 8: Attention and search

*How has our attention been commodified through search and recommendations?*

**Lecture topic:** YouTube

**Lab topic:** Google Trends

## Monday, 8/12: NO CLASS

On your own, watch Zeynep Tufekci's twenty-three minute TED Talk, "We're Building a Dystopia Just to Make People Click on Ads." September 2017. ([link](#))

## Wednesday, 8/14

Noble, Safiya. 2018. "1. A Society, Searching." From *Algorithms of Oppression*. ([ebook available from the UW library](#))

Twitter and Tear Gas, Chapter 2, "Censorship and Attention."

**Recommended:** Tufekci, Zeynep. 2018. "YouTube, the Great Radicalizer." The New York Times, March 10. ([link](#))

## Week 9: The platform economy

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

**Lecture topic:** Amazon and Alexa

**Lab topic:** final projects

## Monday, 8/19

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

Crawford, Kate and Vladan Joler. 2018. "Anatomy of an AI System." ([link](#))

## **Wednesday, 8/21**

### **Final data project presentations in class**

Read this narration about 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 memos, homework assignments, and projects will be penalized 1 percentage 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. For lecture participation and lab check-ins, you can miss two without any effect on your grades. If you believe you will need to miss class for longer than this (for instance, for summer travel), please communicate with me to see if we can work out an alternative or decide if it's appropriate for you to take this course.

### **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](mailto:uwdrs@uw.edu) or [disability.uw.edu](http://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. When writing code, you're encouraged to get help from your peers and online resources, but please acknowledge and attribute the help or example code you receive and use.

## **Attribution**

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

## **Further resources**

Data & Society <https://datasociety.net/>

AI Now Institute <https://ainowinstitute.org/>

Data Feminism <https://bookbook.pubpub.org/data-feminism>

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/>