GET1030

Computers and the humanities

Lecture 11

Computation and society

Dr Miguel Escobar Varela

Objectives

- To summarize key ideas in the module
- To consider the implications of the course materials beyond the module

Main ideas in the module

We need large datasets to describe contemporary culture (the humanities need data science).

Humanities perspectives are crucial for a data-driven world (data science needs the humanities).

What do you think?

Let’s re-evaluate the arguments. And then you can tell me what you think.

Lecture 11

Computation and society

Part 1: The humanities need data science

A quick summary and caveat

In this module we only saw some basic introductory concepts and measurements.

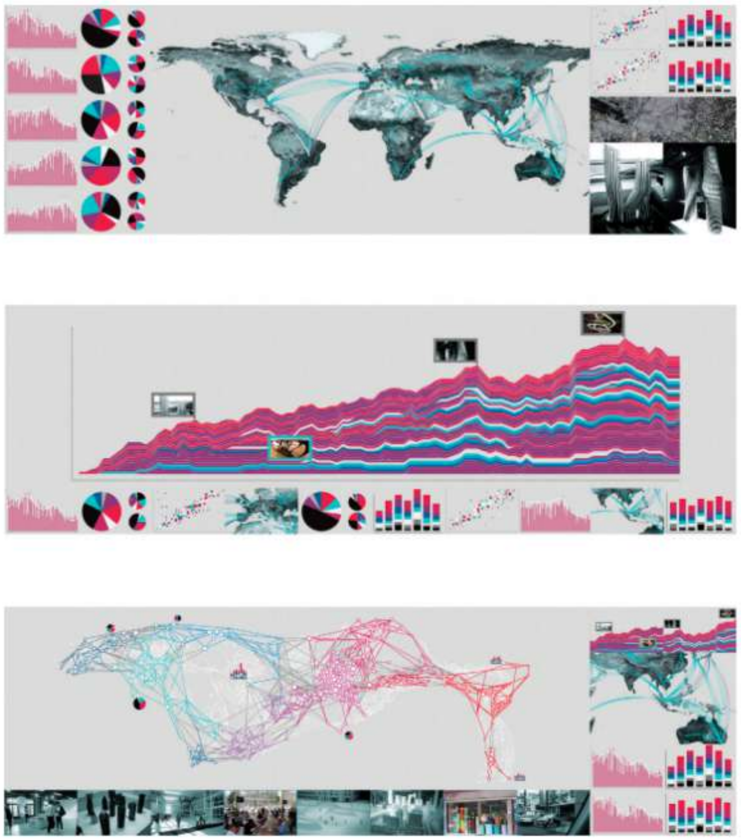
Texts. Basic corpus measurements.

Networks. Basic network-theoretical measurements.

Maps. Basic types of cartographic visualizations.

In Computational Humanities / Cultural Analytics people use more complex measurements and more rigorously statistic procedures. Machine Learning is also becoming more common.

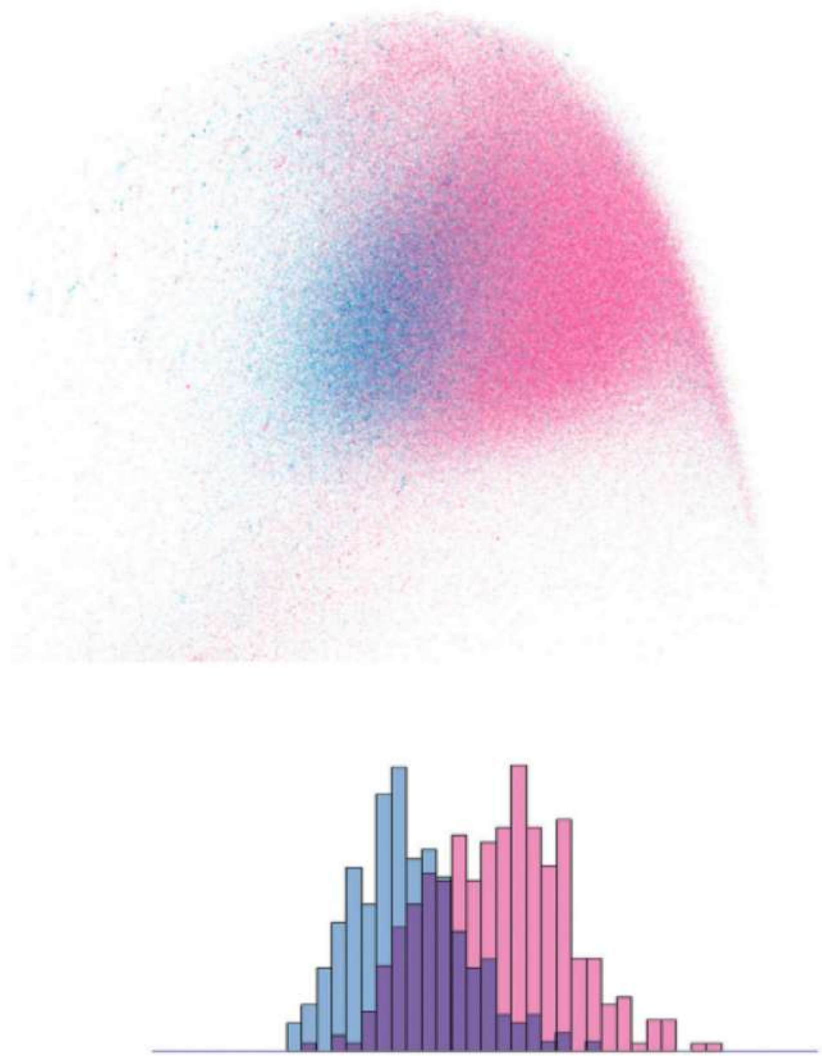
Cultural Analytics at scale



The ultimate goal of cultural analytics can be to map and understand in detail the diversity of contemporary professional and user-generated artifacts created globally (Manovich, 2020).

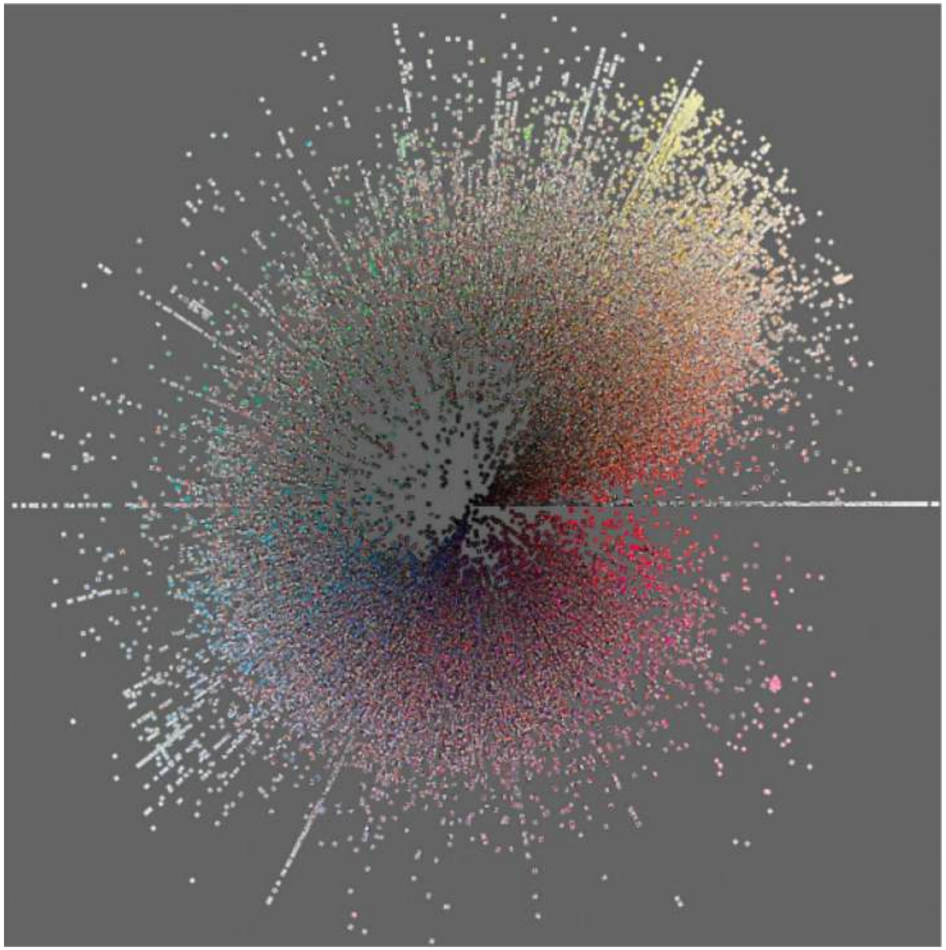
GET1030

Cultural Analytics at scale



Plots comparing shoujo and shōnen manga pages. Blue: shōnen manga; pink: shoujo manga. Top: 1,074,790 manga pages shown as points and organized according to their brightness standard deviation (x-axis) and brightness entropy (y-axis). Bottom: Histograms of mean brightness values of the same 1,074,790 manga pages (Manovich 2020).

Cultural Analytics at scale



Radial image plot showing fifty thousand Instagram images shared in Bangkok, sorted by median hue (angle) and brightness mean (distance to the center) measurements.

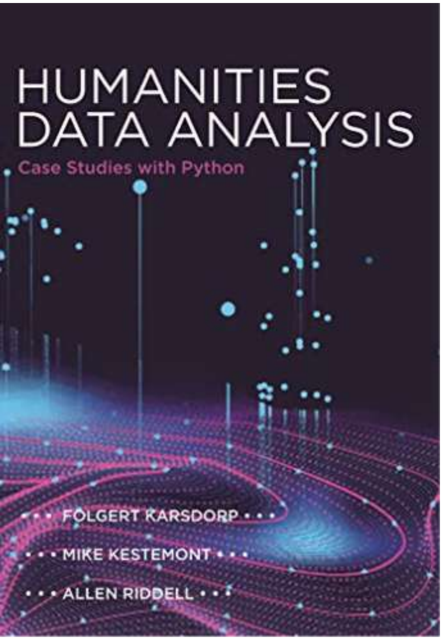
(Manovich 2020)

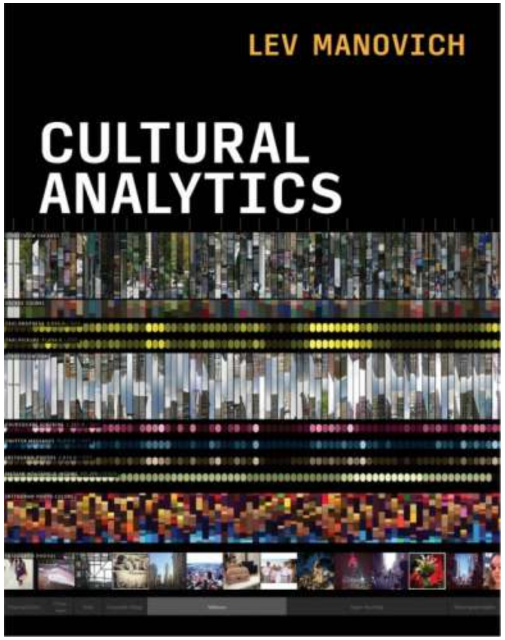
Cultural Analytics at scale

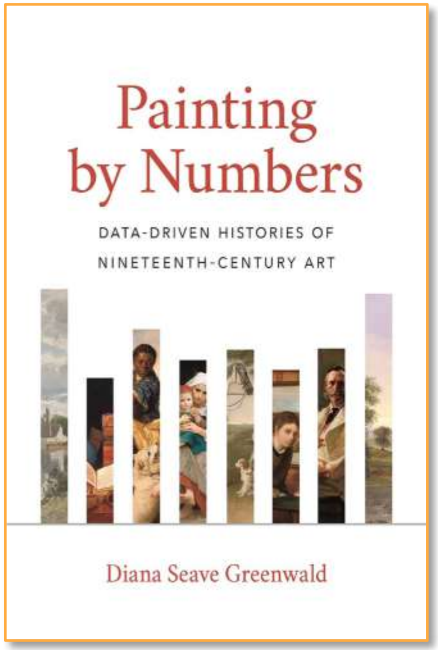
If we do not learn to see at sufficient resolution what people today create and how they behave culturally, any theories or interpretations we may propose based on our intuitions and received knowledge are likely to be misguided. This was the case when we analyzed data on 16 million Instagram images shared in 17 global cities in 2012- 2015, one million manga pages, five thousand paintings of Impressionist artists, and other cultural datasets. In each case, my assumptions about what I am going to see based on intuitions and accepted knowledge were overturned. (Manovich, 2020).

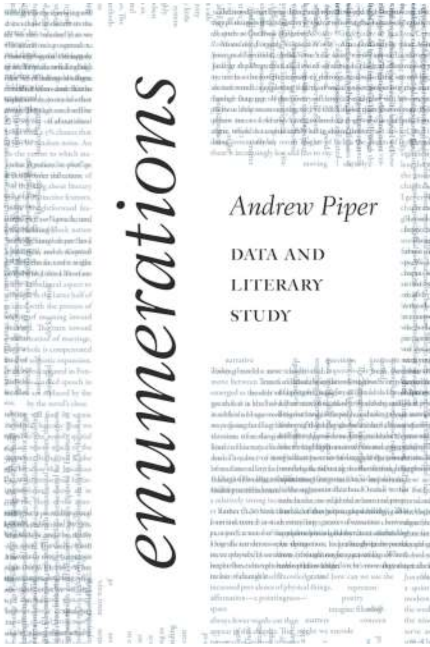
GET1030

Cultural Analytics at scale









Lecture 11

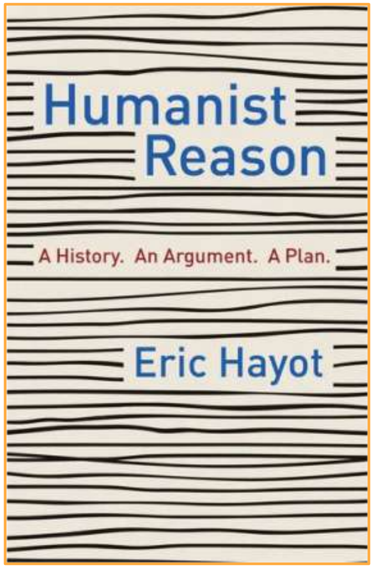
Computation and society

Part 2: Data science needs the humanities

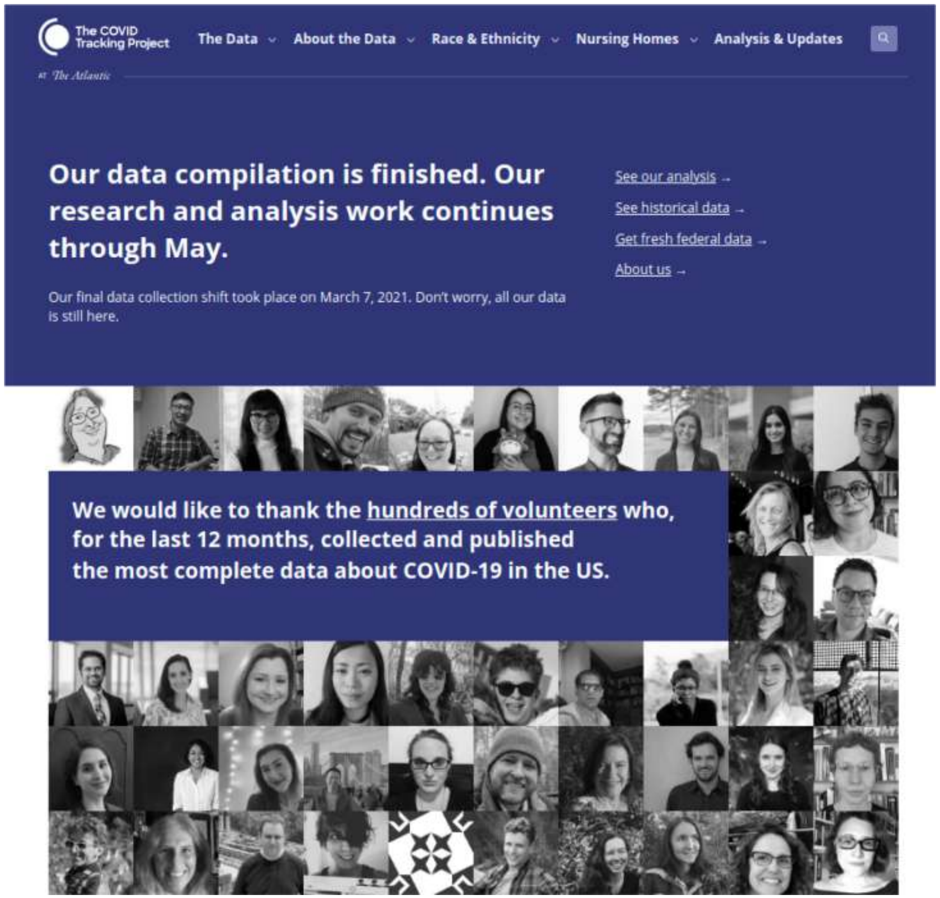
What do we mean by “humanities approaches”?

- Attention to context, history, definitions and small differences
- This is how literary scholars read texts, but this approach also needs to complement data analytics

Humanist reason (Hayot, 2021)



Example: the Covid Tracking Project



Volunteer-lead data collection project in the US.

https://covidtracking.com/

Covid data

1) All data are created; data never simply exist.

Before March 2020, the country had no shortage of pandemic-preparation plans. Many stressed the importance of data-driven decision making. Yet these plans largely assumed that detailed and reliable data would simply … *exist*. They were less concerned with how those data would actually be made.

Data systems have to be aligned very precisely to produce detailed statistics. Yet in the U.S., many states create one sort of data for themselves and another, simpler feed to send to the federal government. Both numbers might be "correct" in some sense, but the lack of agreement within a state's own numbers made interpreting *national* data extremely difficult.

The early work of the COVID Tracking Project was to understand those inconsistencies and adjust for them, so that every state's data could be gathered in one place. Consider the serpentine journey that every piece of COVID-19 data takes. A COVID-19 test, for instance, starts as a molecular reaction in a vial or lab machine, then proceeds through several layers of human observation, keyboard entry, and private computer systems before reaching the government. The pipelines that lead to county, state, and federal databases can be arranged in many different ways. At the end of the process, you have a data set that looks standardized, but may actually not be […]

The government missed the initial explosion of COVID-19 cases because, despite its many plans to *analyze* data, it assumed that data would simply materialize.

https://www.theatlantic.com/science/archive/2021/03/americas-coronavirus-catastrophe-began-with-data/618287/

Covid data

In November 2020, they stopped publishing data on recovered cases

First, several states and territories [...] don't report any kind of recovery data, and it doesn't make sense to report a national total that excludes so much of the country. A second and crucial reason is that"recovered" has no standard definition, and states report it in many different ways. Just as important, many people who have had COVID-19 and have lived to tell the tale—and many of whom are categorized as "recovered"—don't consider themselves to have actually recovered.

COVID-19 can have many long-term health consequences, and none of the definitions for counting people who have "recovered" from COVID-19 accounts for latent or ongoing health issues that can be caused by COVID-19 [...] Determining how many people have recovered from COVID-19, then, is currently more like trawling with a net than fishing with a pole: Every attempt dredges up a lot of scaly things we don't want.

In the absence of federal guidance, and as with many other COVID-19 metrics, different U.S. jurisdictions rely on different definitions for reporting recoveries [...] available definitions generally fall into one of four categories: days since diagnosis/onset; symptom improvement; hospital discharged; or definitions that are unclear.

https://www.theatlantic.com/health/archive/2021/01/how-many-have-recovered-covid-19-we-dont-know/617679/

Example: the Covid Tracking Project

Data Heroes of Covid Tracking Project Are Still Filling U.S. Government Void

The volunteer effort has become a vital source of information on the pandemic.

https://www.bloomberg.com/news/features/2020-11-20/covid-tracking-project-volunteers-step-up-as-u-s-fails-during-pandemic

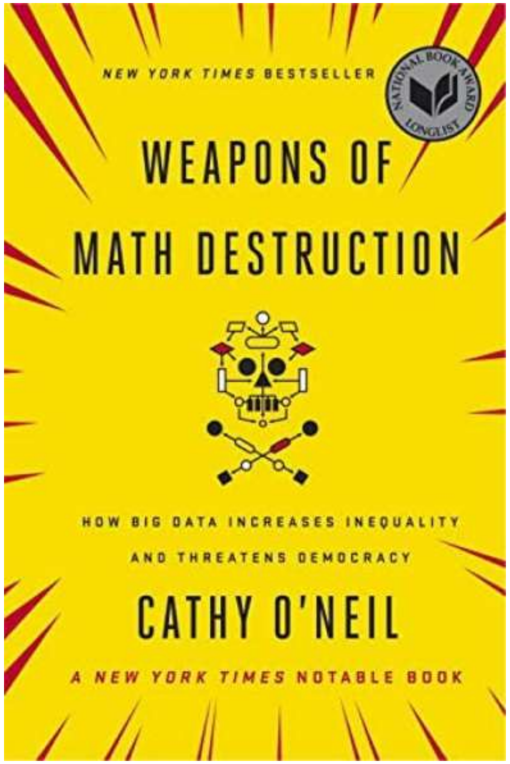
Volunteer-lead data collection project in the US.

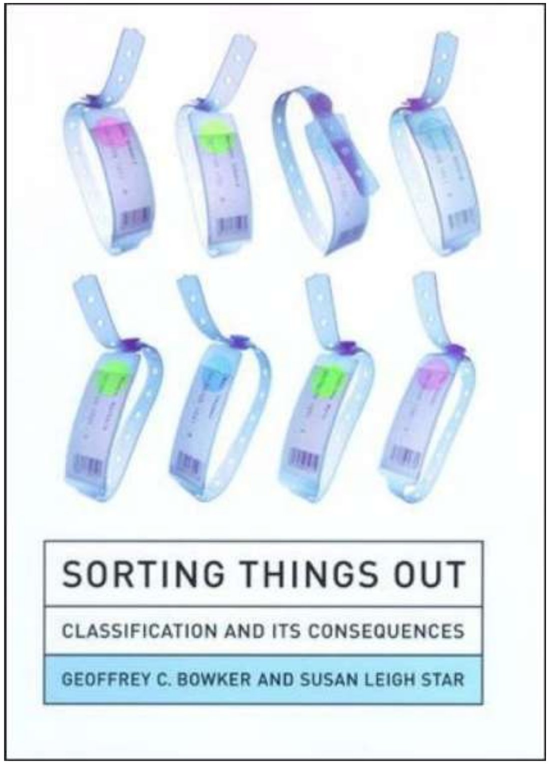
Dr Amanda French, Community Lead and Data Entry Shift Lead at The COVID Tracking Project, has a doctorate in English and is an expert in digital humanities.

Why this matters

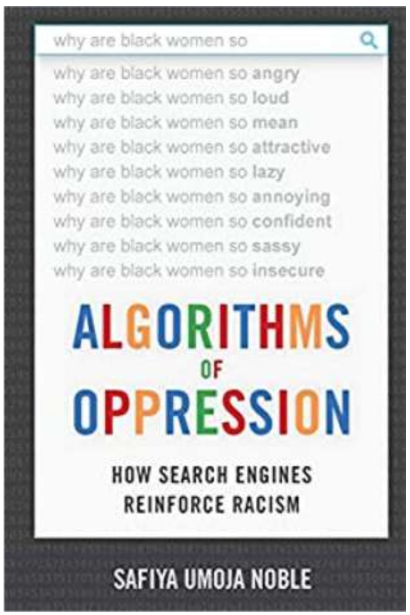
Learning the history and context of how data is collected, processed and classified will make you a better data scientist… and a more informed citizen in a data-driven world.

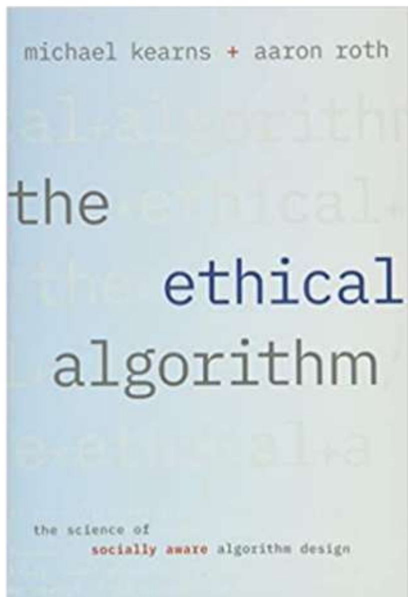
Where to keep learning

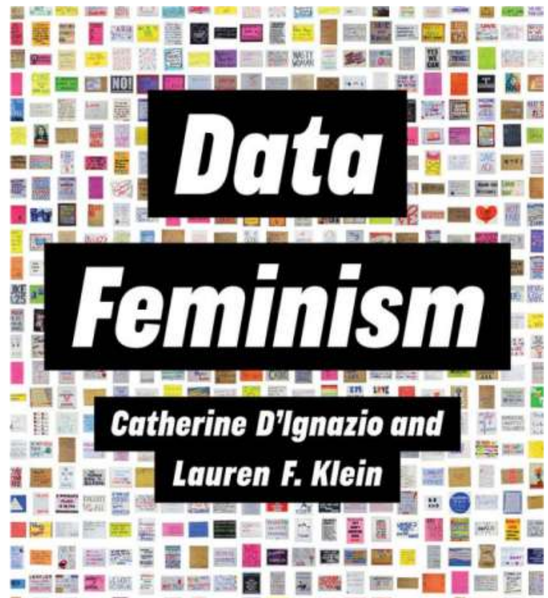


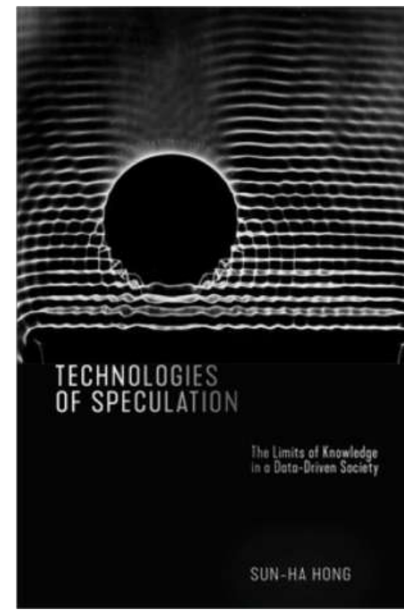


Where to keep learning









Lecture 11

Computation and society

Part 3: Conclusions

To conclude

If you are more of a humanities person, I want you to think more like a data scientist

pay attention to evidence, trends and patterns in large datasets

If you are more of a data person, I want you to think more like a humanist

consider history, context, culture and power

Over to you

Please answer the quiz.

References

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out*. Inside Technology. Cambridge: MIT Press.

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Cambridge, Massachusetts: The MIT Press.

Greenwald, Diana. 2021. *Painting by Numbers*. Princeton: Princeton University Press.

Hayot, Eric. 2021. *Humanist Reason: A History. An Argument. A Plan*. Columbia University Press.

Hong, Sun-ha. 2020. *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society*. New York: NYU Press.

Karsdorp, Folgert, Mike Kestemont, and Allen Riddell. 2021. *Humanities Data Analysis: Case Studies with Python*. Princeton: Princeton University Press.

Kearns, Michael, and Aaron Roth. 2019. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. New York: Oxford University Press.

Loukissas, Yanni Alexander. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Illustrated edition. Cambridge, Massachusetts: The MIT Press.

Manovich, Lev. 2020. *Cultural Analytics*. Cambridge, Massachusetts: The MIT Press.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. Illustrated edition. New York: NYU Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction*. Crown Publishers.

Piper, Andrew. 2018. *Enumerations: Data and Literary Study*. Chicago: The University of Chicago Press.