Introduction to Feminist Data

Taught by: Avery Blankenship and Johan Arango-Quiroga

Lauren Savit
Digital Feminisms
Spring 2025

Workshop Objectives:

- Understand the basics of data + metadata
  - Explore “tangible” aspects of data
  - Prepare for the “methodological” aspects of data for next class
- Learn about intersectional feminist data
- Discuss the introduction to Catherine Knight Steele et al.’s *Doing Black Digital Humanities with Radical Intentionality*

All materials can be accessed here: https://bit.ly/3Fjd6Wl

What is data?

What is data?

- A “traditional” definition of data explains it as “a collection of facts, numbers, words, observations, or other useful information.” (Badman 2024).
- There are several types of data, including:
    - Quantitative data: Numerical information, such as “1,000 bees” or “Ten books sold”
    - Qualitative data: Non-numerical information, such as survey responses or interview transcripts
    - Metadata: Data that provides information about or describes data, like the fields “author,” “title,” and “place of publication” for a book in a library catalog system.
    - Big data: Complicated and large data sets that require special handling and processing

What is data? Cont.

- Data, however, is defined in different ways, especially among BIPOC and LGBTQ+ communities.
- For example, Dr. Stephanie Russo Carroll (Director of the Collaboratory for Indigenous Data Governance at the University of Arizona) defines Indigenous data as "elements of our lived experience as humans," including:
  - Data about non-human relations: Land, water, geology, sacred ecosystems, etc.
  - Data about individuals: Administrative, legal, social, etc.
  - Data about Indigenous peoples as collectives: Traditional and cultural info, language systems, etc.

What is Feminist Data?

- Feminist Data challenges the oppressive structures of traditional data. Truly feminist data is intersectional, and includes the following principles:
  - **Intersectionality + Equity:** Gender, race, disability, class, and more are considered in data-power relations. Those with privilege need to recognize it, listen, and learn.
  - **Prioritize Proximity:** Prioritize the experiences, voices, and needs of those within the community being worked with.
  - **Acknowledge the Humanity of Data:** Reduction of the human experience into data is often a tool of oppression.
  - **Remain accountable:** Receive feedback and criticism with an open mind and make changes.

Metrics for Implementation of Feminist Data

Table 1
Aspirational, draft, and final metrics and the structural problems they address

Structural problem	Aspirational metrics to live our values for this book	Draft metrics (open peer review)	Final metrics (copyedited manuscript)
Racism	• 75 percent of citations of feminist scholarship from people of color • 75 percent of examples of feminist data projects discussed led by people of color	Scholarship: 36 percent from people of color Projects: 49 percent led by people of color	Scholarship: 32 percent from people of color Projects: 42 percent led by people of color
Patriarchy	• 75 percent of all citations and examples from women and nonbinary people	67 percent of citations and examples from women and nonbinary people	62 percent of citations and examples from women and nonbinary people
Cissexism	• Center trans perspectives in discussions of the gender binary • Use transinclusive language throughout the book • Example or theorist in every chapter from a transgender perspective	Three of ten chapters feature transgender example and/or theorist	Nine of nine chapters feature transgender example and/or theorist
Heteronormativity	• Resist assumptions about family structure and gender roles • Example or theorist in every chapter that illustrates the power of communal (vs. family) support networks	Ten of ten chapters feature communal example and/or theorist	Nine of nine of ten chapters feature communal example and/or theorist
Ableism	• Challenge the dominance of visualization in the presentation of data • Example or theorist in every chapter that employs nonvisual methods of presenting data	Nine of ten chapters feature nonvisual example and/or theorist	TK of ten chapters feature nonvisual example and/or theorist

Table 1 (continued)

Structural problem	Aspirational metrics to live our values for this book	Draft metrics (open peer review)	Final metrics (copyedited manuscript)
Colonialism	• 30 percent of projects discussed come from the Global South • Example or theorist in every chapter about Indigenous knowledges and/or activism	Projects: 8.5 percent from the Global South Five of ten chapters feature indigenous example and/or theorist	Projects: 7 percent from the Global South Seven of nine chapters feature indigenous example and/or theorist
Classism	• Acknowledge that data science, as a field, is premised on economic, educational, and technological privilege • 50 percent of feminist projects discussed come from outside the academy • Example or theorist in every chapter that demonstrates how the ideas can be applied without expensive technology and/or formal training	Projects: 88 percent from outside academy Ten of ten chapters feature nonacademic example and/or theorist	Projects: 78 percent from outside academy Nine of nine chapters feature nonacademic example and/or theorist
Proximity	• 50 percent of feminist projects discussed feature and quote people directly impacted by an issue (vs. those who study or report on the phenomena from a distance)	Projects: 49 percent feature people directly impacted	Projects: 34 percent feature people directly impacted

From Catherine D'Ignazio and Lauren F. Klein's *Data Feminism*

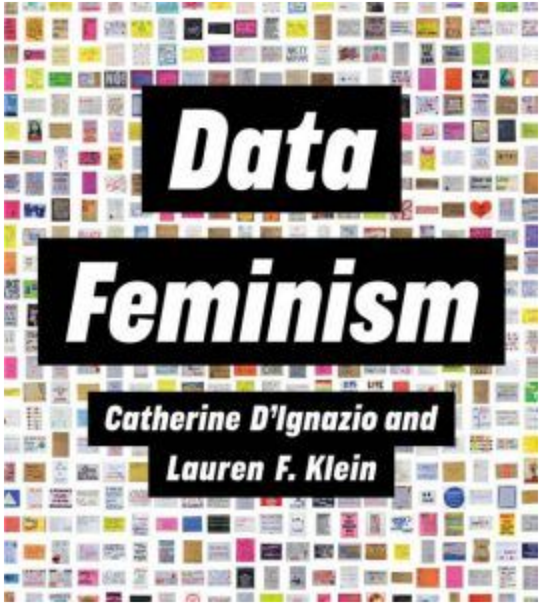
Theory & Methods: Types of Feminist Data Science

Data and Power

- Data is the consolidation of knowledge or information into some symbol, usually numeric
  - Data shares a close relationship with power because of its ability to represent, consolidate, and symbolize people, governments, economies, and ideologies
- Nothing is outside of datafication — everything and everyone is able to be transformed into data

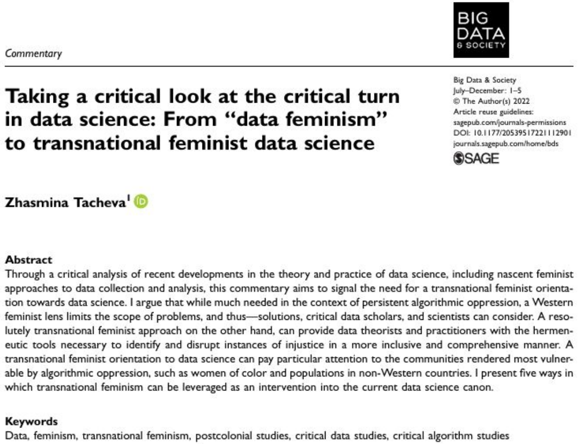
Data Feminism

- *Data Feminism* (2020) by Catherine D'Ignazio and Lauren F. Klein is centered on the idea of power and power relations
- According to D'Ignazio and Klein, our society is driven by data, and entwined with structures like class, gender, and race. Data has benefited white, male, heteronormative power relations.
    - To be a data feminist, then, is to examine power through an intersectional lens and challenge these structures.



Transnational Feminist Data Science

- Zhasmina Tacheva’s article “Taking a critical look at the critical turn in data science” (2022) argues that *Data Feminism* takes a Western approach to feminism, omitting the experiences of marginalized women outside of the West.
- Tacheva instead advocates for transnational feminist data science, which incorporates the diversification of knowledge and non-Western ways of “doing” data science.



Commentary

Big Data & Society
July–December: 1–5

DOI: 10.1177/20539517221112901
journals.sagepub.com/home/bds
SAGE

Taking a critical look at the critical turn in data science: From “data feminism” to transnational feminist data science

Zhasmina Tacheva[1]

Abstract
Through a critical analysis of recent developments in the theory and practice of data science, including nascent feminist approaches to data collection and analysis, this commentary aims to signal the need for a transnational feminist orientation towards data science. I argue that while much needed in the context of persistent algorithmic oppression, a Western feminist lens limits the scope of problems, and thus—solutions, critical data scholars, and scientists can consider. A resolutely transnational feminist approach on the other hand, can provide data theorists and practitioners with the hermeneutic tools necessary to identify and disrupt instances of injustice in a more inclusive and comprehensive manner. A transnational feminist orientation to data science can pay particular attention to the communities rendered most vulnerable by algorithmic oppression, such as women of color and populations in non-Western countries. I present five ways in which transnational feminism can be leveraged as an intervention into the current data science canon.

Keywords
Data, feminism, transnational feminism, postcolonial studies, critical data studies, critical algorithm studies

Anti-Racist Data Science

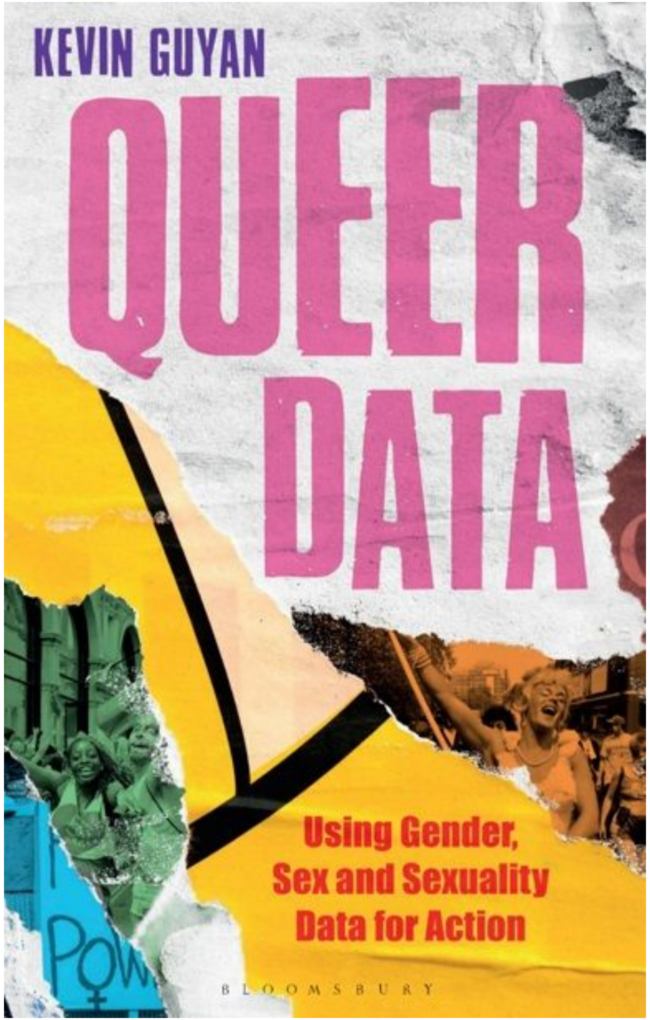
- Anti-Racist Data Science seeks to challenge data structures that oppress BIPOC communities.
- This may be done by challenging the structures in data practices that actively harm BIPOC people, supporting programs and tech initiatives for BIPOC data practitioners, and recognizing contributions of BIPOC data practitioners in these fields
- See for example, the Data for Black Lives movement

- Above: Mathematician and programmer Annie Easley (1933-2011), one of nascent NASA's "computers" and one of only four Black employees

Northeastern University
NULab for Digital Humanities and Computational Social Science

Queer Data Science

- *Queer Data* (2022) by Kevin Guyan highlights how data (specifically, biases in data) are used to "delegitimize the everyday experiences of queer people."
- Guyan's work is centered on using data science to better understand Queer identities, therefore allowing for better "resource allocation, changes to legislation, access to services, representation and visibility."



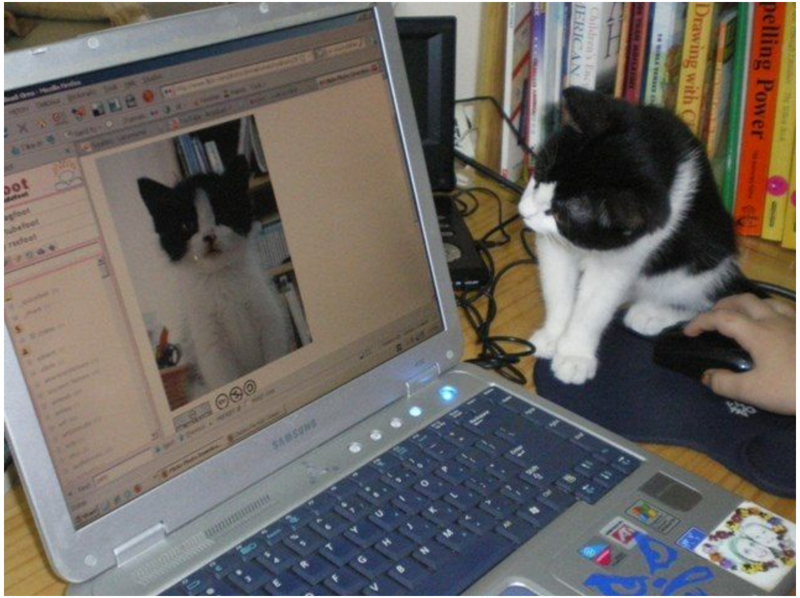
Discussion

- In what ways does “datafication” impact your day-to-day life?
- Let’s think of a common set of data/collection of data and try to think of ways that we could apply data feminim to it
    - Who holds the power? How is the data collected? Who is impacted by its collection?
- How can data feminism help us make better decisions with our data?
- How can we apply some of the responses to Data Feminism to our own work?

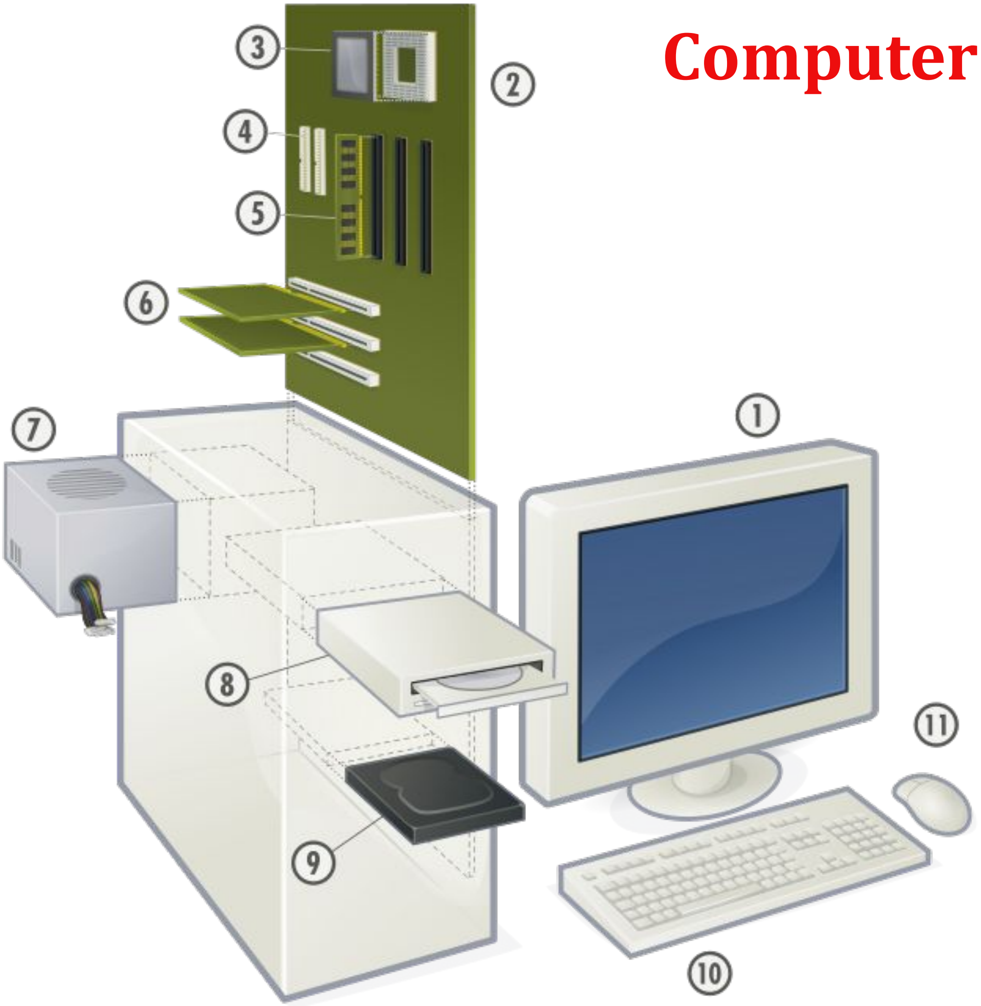
Physical Technology

What does computer data look like?

- Data comes in many forms, such as a number of things (one hundred boats) a collection of interviews, or the temperature of a lake.
- The actual technology used to interpret or process this data, however, is a computer.



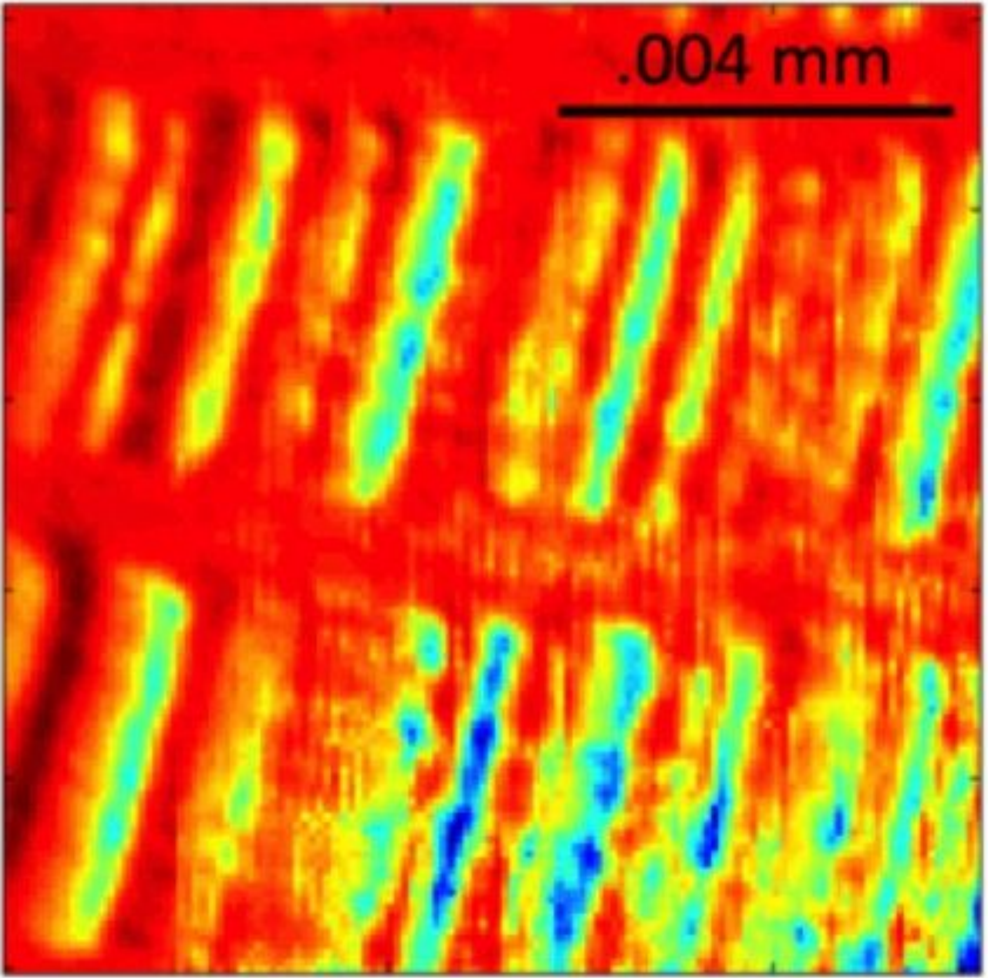
Computer Components



1. Monitor
2. Motherboard
3. CPU (Microprocessor)
4. ATA sockets
5. Main memory (RAM)
6. Expansion cards
7. Power supply unit
8. Optical disc drive
9. Hard disk drive (HDD)
10. Keyboard
11. Mouse

What does computer data look like? Cont.

- We know what a computer looks like and we know where the data is stored (on the hard drive). But what does data look like within the drive?
    - The drive has a magnetic disk with “bits” which look like stripes
    - These bits turn data and digital files into binary code, which look like zeroes and ones (0,1)
    - The code acts like a “switch” for electricity, turning signals on and off to display text, images, videos, etc.



Close up of the 0s and 1s in the hard drive. The 0s are blue, and the 1s are red.

(Berger 2019).

But what does data really look like?

- One gigabyte of data on punch cards used in the 40s-50s weighed more than thirty-five tons
- While nowhere near as heavy, a 1TB hard drive weighs about 230 grams, a server 20-25 pounds, and a server rack about 250 pounds

What is a data center?

- A data center houses hardware like servers, data storage devices, network connection devices and other computing infrastructure which makes the maintenance and storage of data and its connection to networks like the internet and cloud computing possible
- The very first data center was opened by the University of Pennsylvania to support the computing needs of the first digital computer, the ENIAC, in 1945
    - Today, there are 5,381 centers open in the US

Data Technology and the Environment

- Because a great deal of electricity is needed to power and water to cool the hardware to run generative AI, the generative AI boom is having a massive impact on the environment, more than doubling the electricity demands on data centers
- Carbon emissions are also a major concern—There is limited data on the carbon footprint of a single generative AI query, but some industry figures estimate it to be four to five times higher than that of a search engine query.

Discussion: Preparation for Next Class

Case Study: *Doing Black Digital Humanities with Radical Intentionality*

- Read the introduction to Catherine Knight Steele et al.'s *Doing Black Digital Humanities with Radical Intentionality.*
    - How do the authors imagine data?
    - What is radical intentionality?
    - What might radical intentionality look like when applied to data?
    - What connections can you draw between this reading and digital feminisms?

For Further Exploration

[Copyright and fair use handout]

[Data Ethics handout]

[Data Privacy handout]

Thank you!

Taught by: Avery Blankenship and Johan Arango-Quiroga

Developed by: Claire Lavarreda, Shannon Peifer

- For more information on DITI, please see: https://bit.ly/diti-about
- Schedule an appointment with us! https://bit.ly/diti-meeting
- If you have any questions, contact us at: nulab.info@gmail.com

Northeastern University
NULab for Digital Humanities and Computational Social Science