

A DATA FEMINISM NETWORK &  
TORONTO WOMXN IN DATA SCIENCE  
COLLABORATION

# Data Feminism Study Guide

## Chapter 6-7

---

### Meet the Authors



**Catherine D'Ignazio**

 @kanarinka



**Lauren F. Klein**

 @laurenfklein

---

## Seven Principles of Data Feminism:

*Examine Power*

*Challenge Power*

*Elevate Emotion and Embodiment*

*Rethink Binaries and Hierarchies*

*Embrace Pluralism*

*Consider Context*

*Make Labor Visible*

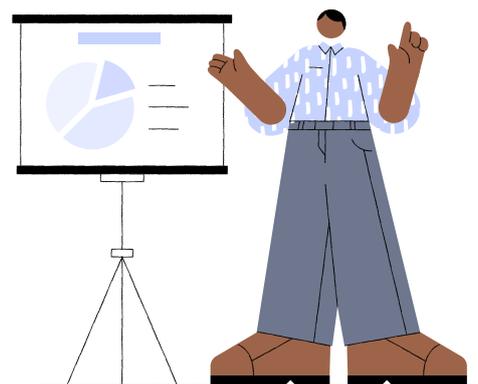
# Chapter 6 - The Numbers Don't Speak for Themselves

## *Principle: Consider Context*

→ Data feminism asserts that data are not neutral or objective. They are the products of unequal social relations, and this context is essential for conducting accurate, ethical analysis.

## Key Definitions

- **Big dick data** → a formal academic term coined by the authors to denote big data projects that are characterized by patriarchal, cis-masculinist, totalizing fantasies of world domination as enacted through data capture and analysis. Big Dick Data projects ignore context, fetishize size, and inflate their technical and scientific capabilities (151)
- **Knowledge infrastructure** → what Christine Borgmann defines as, “an ecology of people, practices, technologies, institutions, material objects, and relationships” or the context that makes data possible (153)
- **Open data** → the idea that anyone can freely access, use, modify, and share data for any purpose (155)
- **Statistical inference** → the theory that you can infer things about a population by studying a random and/or representative sample and then mapping those findings back on the population as a whole (156)
- **Subjugated knowledge** → this term describes the forms of knowledge that have been pushed out of mainstream institutions and the conversations they encourage (163). Subjugated knowledge is knowledge that is not taken seriously, discounted, or ignored, and is even seen as irrelevant and unimportant by those who are positioned to influence what is considered “acceptable”



# Key Themes

- *The issue of context (152-153)*
  - All knowledge is situated. When approaching any new source of knowledge, it's essential to ask questions about the social, cultural, historical, institutional, and material conditions under which that knowledge was produced, as well as about the identities of the people who created it. Context allows us, as data scientists, to better understand any functional limitations of the data and any associated ethical obligations, as well as how the power and privilege that contributed to their making may be obscuring the truth.
- *The open data movement (155)*
  - The open data movement is a loose network of organizations, governments, and individuals. The goals are good ones in theory: economic development by building apps and services on open data; faster scientific progress when researchers share knowledge; and greater transparency for journalists, citizens, and residents to be able to use public information to hold governments accountable. In practice, however, limited public funding for technological infrastructure has meant that governments have prioritized the "opening up" part of open data - publishing spreadsheets of things like license applications, arrest records, and flood zones - but lack the capacity to provide any context about the data's provenance, let alone documentation that would allow the data to be made accessible and usable by the general public.
- *Raw data, cooked data, cooking (159)*
  - Lisa Gitelman and Virginia Jackson have memorably explained, data enter into research projects already fully cooked - the result of a complex set of social, political, and historical circumstances.
- *Refusing to acknowledge context is a power play (162)*
  - It's a way to assert authoritativeness and mastery without being required to address the complexity of what the data actually represent.
- *Numbers cannot speak for themselves (171)*
  - Those of us who work with data must actively prevent numbers from speaking for themselves because when those numbers derive from a data setting influenced by differentials of power, or by misaligned collection incentives, and especially when the numbers have to do with human beings or their behavior, then they run the risk not only of being arrogantly grandiose and empirically wrong, but also of doing real harm in their reinforcement of an unjust status quo.

---

# Data Feminism in Action

## ***ALGORITHMS OF OPPRESSION, BY SAFIYA UMOJA NOBLE***

Noble demonstrates that Google search results do not simply correlate with our racist, sexist, and colonialist society; that society causes the racist and sexist results. (156)

→ <https://nyupress.org/9781479837243/algorithms-of-oppression/>

## ***THE IMAGE OF ABSENCE, BY LAUREN F. KLEIN***

Some may already know that Jefferson is considered the nation's "founding foodie." But fewer know that he relied upon an enslaved kitchen staff to prepare his famous food. Lauren created a visual representation of all of the work that Jefferson's enslaved staff put into preparing his meals but that he did not acknowledge - at least not directly - in the text of the letters themselves. (161)

→ <https://eng318dataasrheteric2017.web.unc.edu/wp-content/uploads/sites/13595/2017/01/American-Literature-2013-Klein-661-88.pdf>

## ***SAFELAB, RUN BY SOCIAL WORKER AND SCHOLAR DESMOND PATTON***

SAFElab uses artificial intelligence to examine the ways that youth of color navigate violence on and offline. Patton wanted to take a deeper approach to really "grasp culture, context and nuance, for the primary reason of not misinterpreting what's being said". His approach to incorporating culture, context, and nuance took the form of direct contact with and centering the perspectives of the youth whose behaviors his group sought to study. (163)

→ <https://safelab.socialwork.columbia.edu/news/turing-lecture-ai-innovative-social-work>

## ***DATA USER GUIDES BY BOB GRADECK, MANAGER OF THE WESTERN PENNSYLVANIA REGIONAL DATA CENTER***

They are simple, written documents that each contain a narrative portrait of a dataset. They describe, among other things, the purpose and application of the data; the history, format, and standards; the organizational context; other analyses and stories that have used the dataset; and the limitations and ethical implications of the dataset. (169)

→ <http://www.wprdc.org/data-user-guides/>

## **WOMANSTATS, BY VALERIE HUDSON**

Hudson, a professor of political science, worked for decades to trace the links between state security and the status of women. She and geographer Chad Emmett started the project WomanStats as a modest Excel spreadsheet in 2001. It has since grown to a large-scale web database with over a quarter of a million data points, including over 350 variables ranging from access to health care to the prevalence of rape to the division of domestic labor. (169)

→ <https://www.womanstats.org/>

### **Questions to Ask When Considering Context (172)**

1. Which power imbalances have led to silences in the dataset or data that is missing altogether?
2. Who has conflicts of interest that prevent them from being fully transparent about their data?
3. Whose knowledge about an issue has been subjugated, and how might we begin to recuperate it?

## **Discussion Questions**

1. Which actors in the data ecosystem are responsible for providing context? End users? Data publishers? Data intermediaries?
2. What steps can we take to ensure context is considered? How can we more effectively present context through data visualization?



# Chapter 7 - Show Your Work

## *Principle: Make Labor Visible*

→ The work of data science, like all work in the world, is the work of many hands. Data feminism makes this labor visible so that it can be recognized and valued.

## Key Definitions

- **Invisible labor** → encompasses the various forms of labor, unwaged, underwater, and even waged, that are rendered invisible because they take place inside of the home, because they take place out of sight, or because they lack physical form altogether (179)
- **Reproductive labor** → comes from the classical economic distinction between the paid and therefore economically productive labor of the marketplace, and the unpaid and therefore economically unproductive labor of everything else (179)
- **Racialized labor** → just as housework is structured along the lines of gender, it is also structured along the lines of race and class (179)
- **Cultural data work** → cultural data workers are responsible for the invisible labor involved in moderating the veritable deluge of content produced online every day, ensuring that your Facebook feed is free of, for example, child pornography and violent propaganda videos (182)
- **Emotional labor** → as described by feminist sociologist Arlie Hochschild, emotional labor describes the work involved in managing one's feelings, or someone else's, in response to the demands of society or a particular job (192)



# Key Themes

- *Hidden labor (178)*
  - It's not a coincidence that much of the work that goes into designing a data product - visualization, algorithm, model, app - remains invisible and uncredited. In our capitalist society, we tend to value work that we can see.
- *Slavery and colonialism (184)*
  - Our present technological infrastructure follows this same pattern of exploitation - the exploitation of Black bodies for white financial gain.
- *Citation (185-188)*
  - Feminist theorist Sara Ahmed describes this practice as a way of resisting how certain types of people - usually cis and white and male - "take up spaces by screening out others". When those other people are screened out, they become invisible, and their contributions go unrecognized.

## Data Feminism in Action

### **INTERNATIONAL FEMINIST COLLECTIVE WAGES FOR HOUSEWORK CAMPAIGN**

The group was trying to expose invisible labor, labor that was unpaid and therefore unvalued. (178)

→ <https://www.plutobooks.com/blog/wages-housework-campaign-history/>

### **WORKERS LEAVING THE GOOGLEPLEX, A DOCUMENTARY BY ANDREW NORMAN WILSON**

The documentary exposes how the workers tasked with scanning the books for the Google Books database are hired as a separate but unequal class of employee. Wilson also observes that Google's book-scanning workers are disproportionately women and people of color. (181)

→ <http://www.andrewnormanwilson.com/WorkersGoogleplex.html>

## **ANATOMY OF AN AI SYSTEM, BY TECHNOLOGY RESEARCHER KATE CRAWFORD AND DESIGN SCHOLAR VLADLAN JOLE**

Seeks to describe and diagram the human labor, data dependencies, and material resources that contribute to a single Amazon Echo. (185)

→ <https://anatomyof.ai/>

## **THE DIVERSE ECONOMIES ICEBERG (2017)**

A diagram of the multiple labor practices created by the Next System Project for a report on cultivating communities. Image courtesy of J.K. Gibson-Graham, Jenny Cameron, Kelly Dombrowski, Stephen Healy, and Ethan Miller for the Next System Project. (190)

→ <https://www.communityeconomies.org/resources/diverse-economies-iceberg>

---

## Discussion Question

1. How do we “show the work” of care workers and other forms of invisible labor?

How

do we ensure that this work is sufficiently recognized and valued? And can we do anything more to challenge the root cause of this undervalued work?



# Conclusion - Now Let's Multiply

## The Path Forward

What is most important is not that we all share the same starting point, but rather that we nurture all of these emerging ecosystems and build links between them. We will need all of them for mobilizing resistance to the differentials of power embedded in our current datasets and data systems. And we will also need them for mobilizing courage and creativity - to imagine what data science and artificial intelligence beyond the matrix of domination might look like. The best time for resistance and reimagination is before the norms and structures and regulations of the data economy have been fully determined.

So now let's multiply. (214)

## Data Feminism in Action

### **2018 WALKOUT AT GOOGLE**

: Women-led walkout to protest Google's handling of sexual misconduct cases. (204)

→ <https://www.nytimes.com/2018/11/01/technology/google-walkout-sexual-harassment.html>

### **DESIGN JUSTICE**

A networked community assembled to challenge the idea of "design for good." As co-organizer Una Lee put it, "How could we redesign design so that those who are normally marginalized by it, those who are characterized as passive beneficiaries of design thinking, become co-creators of solutions, of futures?" (205-206)

→ <https://designjustice.org/>

## **TECH WORKERS COALITION**

A group that is building bridges between the programmers who code the search engines and the cafeteria workers who prepare their food. (205)

→ <https://techworkerscoalition.org/>

## **DATA FOR BLACK LIVES (D4BL)**

A network [of] over 4,000 scientists and activists working to harness the power of data and technology to make real change in the lives of Black people. The group's emphasis on abolition and liberation, rather than a generic form of social good, leads it to design projects that actively work to overturn the data-driven discrimination experienced in Black communities. (206)

→ <https://d4bl.org/>

## **THE CENTER FOR SPATIAL RESEARCH AT COLOMBIA BY LAURA KURGAN**

Uses a uniquely transdisciplinary approach that includes data science and AI, the humanities, geography, and design to investigate complicated phenomena like urban/rural displacement due to conflict. (213)

→ <https://www.arch.columbia.edu/research/centers/3-center-for-spatial-research>

## **MODEL CARDS**

A form of documentation by Margaret Mitchell and her co-authors that would accompany machine learning models to detail their intended uses and their technical and ethical limitations. (213)

→ <https://arxiv.org/pdf/1810.03993.pdf>

## **ECONOMÍA FEMINI(S)TA**

Activist group based in Argentina that has an ongoing civic accountability project called Feminindex in which the group visualizes where each candidate stands in relation to a range of gender and LGBTQ+ issues. (213)

→ <https://economiafeminita.com/>

# Discussion Questions

1. How can we resist and reimagine the norms, structures, and regulations of the data economy?

