

A DATA FEMINISM NETWORK &
TORONTO WOMXN IN DATA SCIENCE
COLLABORATION

Data Feminism Study Guide

Created by



Ali Dunn
DFN



Kishawna Peck
Toronto Womxn



Jade Greer
DFN

Seven Principles of Data Feminism:

Examine Power

Challenge Power

Elevate Emotion and Embodiment

Rethink Binaries and Hierarchies

Embrace Pluralism

Consider Context

Make Labor Visible

Table of Contents

Introduction	-----	01
Chapter 1	-----	03
Chapter 2	-----	06
Chapter 3	-----	10
Chapter 4	-----	13
Chapter 5	-----	15
Chapter 6	-----	19
Chapter 7	-----	23
Conclusion	-----	26

Introduction

Key Definitions

- **Feminism** → a shorthand for the diverse and wide-ranging projects that name and challenge sexism and other forces of oppression, as well as those which seek to create more just, equitable, and livable futures. (6)
- **Intersectionality** → the idea of intersectionality describes not only the intersecting aspects of any particular person's identity, but it also describes the intersecting forces of privilege and oppression at work in a given society. (7)
 - The term was coined by legal theorist Kimberlé Crenshaw in the late 1980s
- **Oppression** → the systematic mistreatment of certain groups of people by other groups which happens when power is not distributed equally. (8)
- **Data Feminism** → a way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought. (8)
- **Co-liberation** → the idea that oppressive systems of power harm all of us, that they undermine the quality and validity of our work, and that they hinder us from creating true and lasting social impact with data science. (9)
- **Datafication** → a technological trend turning many aspects of our life into data which is subsequently transferred into information and monetized. (12)

Key Themes

DATA FEMINISM IS FOR EVERYBODY

- *Data feminism isn't only about women:* It takes more than one gender to have gender inequality and more than one gender to work toward justice. (14)
- *Data feminism isn't only for women:* Men, non-binary, and genderqueer people are proud to call themselves feminists and use feminist thought in their work. (14)
- *Data feminism isn't only about gender:* Intersectional feminists have keyed us into how race, class, sexuality, ability, age, religion, geography, and more, are factors that together influence each person's experience and opportunities in the world. (14)
- *Data feminism is about power - about who has it and who doesn't:* Intersectional feminism examines power. And in our contemporary world, data is power too. Because the power of data is wielded unjustly, it must be challenged and changed. (14)

Data Feminism in Action

CHRISTINE MANN DARDEN, DATA ANALYST AT NASA

Darden was faced with the fact that in her field, men with math credentials were promoted while women with the same credentials were sent to computing pools where they would languish until they quit or retired. She faced intersecting oppressions as a Black woman in a white male dominated sector. Darden, however, championed for herself and went on to be the first Black woman to advance to the top rank in the federal civil service. Her story reflects several decades of feminist activism and critical thought. You may know her story from the movie "Hidden Figures". (8)

→ <https://www.quantamagazine.org/the-nasa-engineer-whos-a-mathematician-at-heart-20210119/>

YESHIMABEIT MILNER, FOUNDER OF DATA FOR BLACK LIVES (D4BL)

This organization is dedicated to "using data science to create concrete and measurable change in the lives of Black Communities." D4bl uses explicitly feminist and antiracist methods to quantify and challenge invasive data collection by law enforcement. (13)

→ <https://d4bl.org/>

PREDPOL MODEL, EXAMINED BY CATHY O'NEIL IN *WEAPONS OF MATH DESTRUCTION*

Predpol is a predictive policing company founded in California that is used to determine which neighborhoods to patrol more heavily and which neighborhoods to largely ignore. This model is based on historical crime data and reflects policies that disproportionately surveilled and monitored neighborhoods of color. This is one example of what O'Neil describes as a pernicious feedback loop which amplifies the effect of racial bias and of the criminalization of poverty that are already endemic to the U.S. (13)

→ <https://ideas.ted.com/justice-in-the-age-of-big-data/>



Chapter 1 - The Power Chapter

Principle: **Examine Power**

→ Data feminism begins by analyzing how power operates in the world.

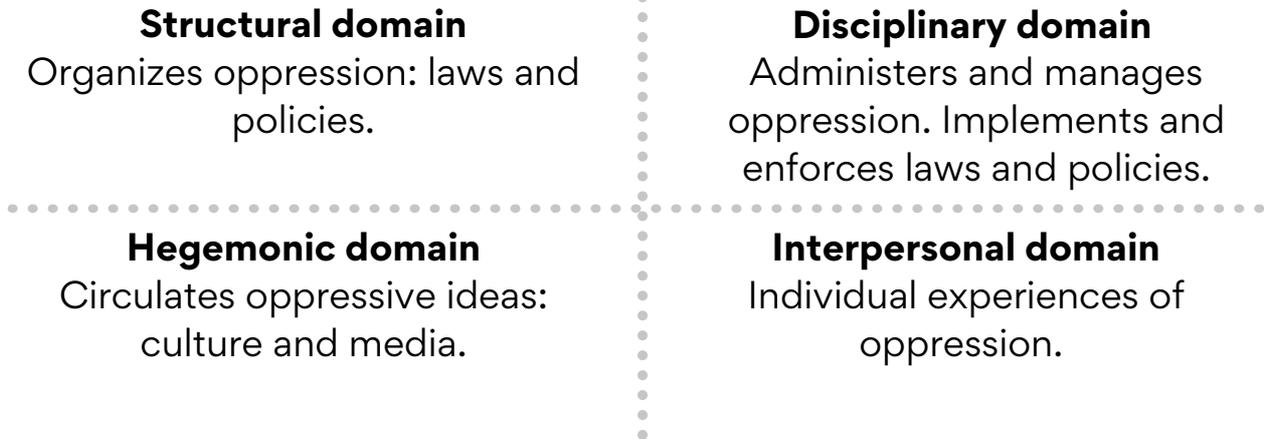
Key Definitions:

- **Examining power** → naming and explaining the forces of oppression that are so baked into our daily lives - and into our datasets, our databases, and our algorithms - that we often don't even see them. (24)
- **Power** → the current configuration of structural privilege and structural oppression, in which some groups experience unearned advantages - because various systems have been designed by people like them and work for people like them - and other groups that experience systematic disadvantages - because those same systems were not designed by them or with people like them in mind. (24)
- **Minoritized** → describes groups of people who are positioned in opposition to a more powerful social group. (26)
- **Privilege hazard** → the phenomenon that makes those who occupy the most privileged positions among us - those with good educations, respected credentials, professional accolades - so poorly equipped to recognize instances of oppression in the world. Their lack of lived experience profoundly limits their ability to foresee and prevent harm, to identify existing problems in the world, and to imagine possible solutions. (29)
- **Scarcity bias** → the idea that there are not enough resources for everyone so we should think small and allow technology to fill the gaps. (40)

Key Themes

- *Data science by whom?* (26)
 - Who is doing the work of data science (and who is not)?
- *Data science for whom?* (33)
 - Who benefits from data science (and who is either overlooked or actively harmed?)
- *Data science with whose interests and goals?* (39)
 - Whose goals are prioritized in data science (and whose are not)?

The Four Domains of the Matrix of Domination



Power and the Matrix of Domination → The matrix of domination works to uphold the undue privilege of dominant groups while unfairly oppressing minoritized groups. (25)

Data Feminism in Action

TENNIS STAR SERENA WILLIAMS' PREGNANCY COMPLICATIONS

Williams faced life threatening pregnancy complications and her self-advocacy played a crucial role in her survival. Her doctors did not believe her when she said she was in pain and thought something was wrong. Through social media, Williams found many other women faced similar experiences. She posted that "Black women are over 3 times more likely than white women to die from pregnancy or childbirth related causes". The privilege she experiences as a tennis star intersected with the oppression she experienced as a Black woman, enabling her to avoid becoming a statistic herself. As Williams asserted, "that's not fair". (23-24)

→ These disparities are well known to Black-women led reproductive justice groups like Sister Song (<https://www.sistersong.net/>), Black Mamas Matter Alliance (<https://blackmamasmatter.org/>), and Raising Our Sisters Everywhere (<http://rose.kellimcnair.com/>)



JOY BUOLAMWINI, A GHANIAN-AMERICAN GRAD STUDENT AT MIT, FACIAL ANALYSIS SOFTWARE PROJECT

The software Boulamwini was working with couldn't see her dark skinned face even though it had no problem seeing her lighter-skinned collaborators. Once she put on a white mask, the software recognized her. This story exemplifies the privilege hazard, as only 4% of faces in the data set for this software were women and dark skinned thus the problem was simply not identified by the coders. (30)

→ <https://www.ajl.org/>

MIMI ONUOHA, ARTIST, DESIGNER, AND EDUCATOR, AND HER PROJECT, THE LIBRARY OF MISSING DATASETS

A list of datasets one might expect to already exist in the world, because they help to address pressing social issues, but have never actually been created. (33)

→ <https://mimionuoha.com/the-library-of-missing-datasets>

MARÍA SALGUERO'S COMPREHENSIVE DATA SET ON FEMICIDES IN MEXICO - GENDER RELATED KILLINGS OF WOMEN AND GIRLS

She has logged more than 5,000 femicides since 2016. Her work provides the most accessible information on the subject for journalists, activists, and victim's families seeking justice. (35)

→ <https://iqlatino.org/2019/one-only-woman-is-mapping-the-femicides-of-all-of-mexico-her-name-is-maria-salguero/>

SEALS ALLERS AND HER SON, MICHAEL'S APP, IRTH

Irth stands for birth, but with the b for bias removed. This app is a data driven contribution to the maternal and infant health conversation. One of the major contributing factors to poor birth outcomes, as well as maternal and infant mortality, is biased care. Hospitals, clinics, and caregivers routinely disregard Black women's expressions of pain and wishes for treatment. (46)

→ <https://irthapp.com/>

Chapter 2 - Collect, Analyze, Imagine, Teach

Principle: Challenge Power

→ Data feminism commits to challenging unequal power structures and working toward justice

Key Definitions:

- **Redlining** → a term used to describe how banks rated the risk of granting loans to potential homeowners on the basis of neighborhood demographics (specifically race and ethnicity), rather than individual credit worthiness. (51)
 - Redlining initially entailed the literal drawing of red lines on a map
 - Redlining maps were very similar to “big data” approaches today where it is ensured that wealth remain attached to the racial category of whiteness. (52)
- **Racial capitalism** → coined by philosopher Cedric Robinson, describes the implementation of data driven methods for granting (or denying) policies to customers based on their demographics.
- **Challenging power** → requires mobilizing data science to push back against existing and unequal power structures and to work toward more just and equitable futures. (53)
- **New Jim Code** → coined by sociologist Ruha Benjamin, where software code and a false sense of objectivity come together to contain and control the lives of Black people and other people of color. (55)
 - The danger of the New Jim Code is that these findings are actively promoted as objective, and they track individuals and groups through their lives and limit their future potential. (55)
- **Auditing algorithms** → using data science to challenge data science
 - Collecting missing data and reverse-engineering the algorithm to prove systemic bias. (57)

- **Deficit narratives** → reduce a group or culture to its “problems,” rather than portraying it with the strengths, creativity, and agency that people from those cultures possess. (58)
 - While it is important to collect counterdata and analyze data to provide proof of oppression, it is equally important to remain aware of how the subjects of oppression are portrayed. (59)
 - Key too is a sustained attention to the ways in which communities themselves are already addressing the issues. (59)
- **Imagined objectivity** → coined by Cedric Benjamin in *Race After Technology: Abolitionist Tools for the New Jim Code*, emphasizes the role that cultural assumptions and personal perceptions play in upholding this false belief: one imagines (wrongly) that data sets and algorithms are less partial and less discriminatory than people and thus more “objective”. (59-60)
- **Data ethics** → growing interdisciplinary effort- both critical and computational- to ensure that the ethical issues brought about by our increasing reliance on data-driven systems are identified and addressed. (60)
- **Equity** → equality assumes everyone at the same starting point and can actually be systematically unfair (for example in the U.S. where Black babies are dying at twice the rate of white babies). Working toward a world in which everyone is treated equitably, not equally, means taking into account these different power differentials and distributing (or redistributing) resources accordingly. (62)
- **Co-liberation** → requires a commitment to and a belief in mutual benefit, from members of both dominant groups and minoritized groups.
 - Needs not only transparency but reflexivity: the ability to reflect on and take responsibility for one’s own position within the multiple, intersecting dimensions of the matrix of domination. (64)

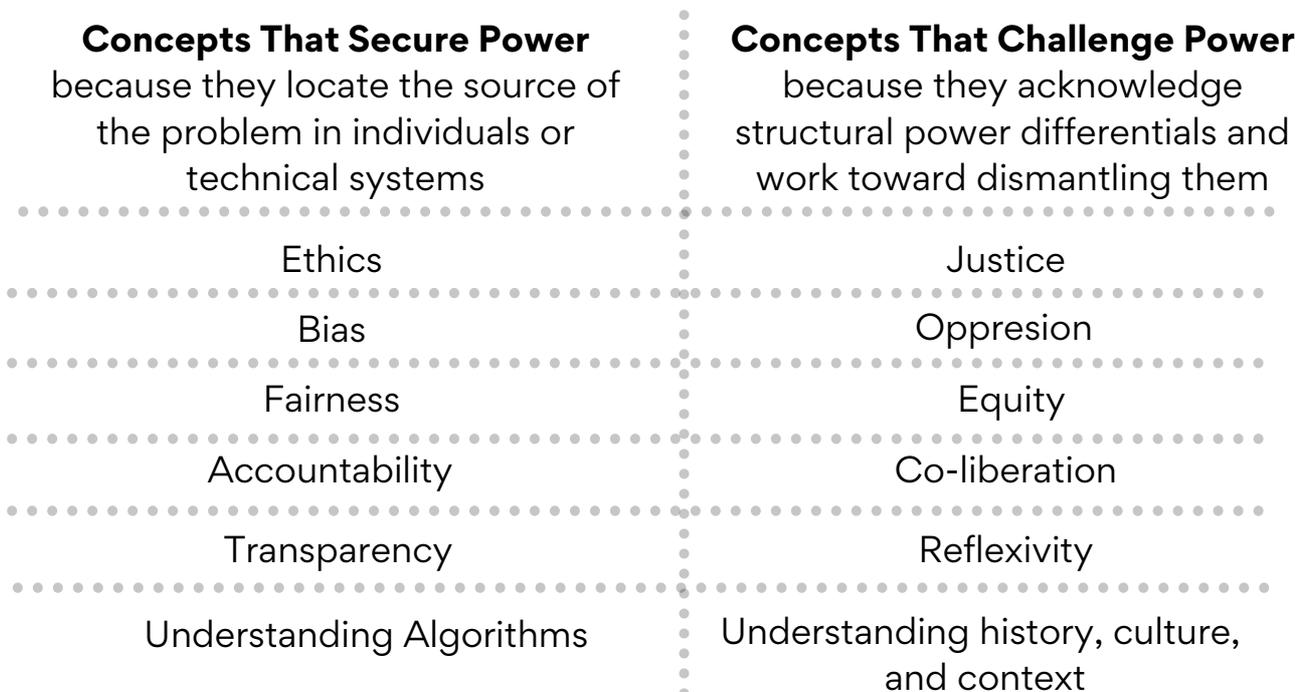


Key Themes

Four Ways to Challenge Power (53)

- *Collect* → Compiling counterdata - in the face of missing data or institutional neglect.
- *Analyze* → Challenging power often requires demonstrating inequitable outcomes across groups, and new computational methods are being developed to audit opaque algorithms and hold institutions accountable.
- *Imagine* → We cannot only focus on inequitable outcomes, because then we will never get to the root cause of injustice. In order to truly dismantle power, we have to imagine our end point not as “fairness” but as co-liberation.
- *Teach* → The identities of data scientists matter, so how might we engage and empower newcomers to the field in order to shift the demographics and cultivate the next generation of data feminists?

From Data Ethics to Data Justice



Data Feminism in Action

DETROIT GEOGRAPHIC EXPEDITION AND INSTITUTE (DGEI) MAP, WHERE COMMUTERS RUN OVER BLACK CHILDREN ON THE POINTES-DOWNTOWN TRACK

The map uses sharp black dots to illustrate the places in the community where the children are killed. Those who lived along this route were aware of the problem and its profound impact on friends and neighbors, but gathering data turned out to be a real challenge. No one kept detailed records of the deaths and no one was making basic information of what happened publicly available. In lieu of missing data, DGEI compiled counter-data to ensure the problem be properly addressed. (52)

→ <https://medium.com/nightingale/gwendolyn-warren-and-the-detroit-geographic-expedition-and-institute-df9ee10e6ad2>

PROPUBLICA TEAM LED BY JULIA ANGWIN INVESTIGATES A WIDELY USED CRIMINAL BAIL ASSESSMENT ALGORITHM

Her team found that white defendants are more often mislabeled as low risk than Black defendants and, conversely, Black defendants are mislabeled as high risk more often than white defendants. This project successfully proved systemic racial bias. (53)

→ <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

OUR DATA BODIES (ODB) PROJECT

Led by a group of five women, this project is a community centered initiative focused on data collection efforts that disproportionately impact minoritized people. They recently released Digital Defence Playbook, which is a set of activities, tools, and tip sheets intended to be used by and for marginalized communities to understand how data driven technologies impact their lives. (64)

→ <https://www.odbproject.org/tools/>

LOCAL LOTTO PROJECT

Taught local high school students statistics and data analysis rooted in neighborhood and justice concerns. (68)

→ http://citydigits.org/utm_content=buffer0f105&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer

Chapter 3 - On Rational, Scientific, Objective Viewpoints From Mythical, Imaginary, Impossible Standpoints

Principle: Elevate Emotion and Embodiment

→ Data feminism teaches us to value multiple forms of knowledge, including the knowledge that comes from people as living, feeling bodies in the world.

Key Definitions:

- **Framing effects** → the impact data scientists have on how people interpret the graphics and what they take away from them (82)
- **Feminist objectivity** → a tool that can account for the situated nature of knowledge and bring together multiple partial perspectives (83)
- **Strong objectivity** → works toward more inclusive knowledge production by centering the perspectives or standpoints of groups that are otherwise theory (83)
- **Positionality** → concept that emphasizes how individuals come to knowledge making processes from multiple positions, each determined by culture and context (83)
- **Affect** → the term that academics use to refer to emotions and other subjective feelings (84)
- **Heuristics** → using mental shortcuts to make judgments (90)



Key Themes

- *God trick (76)*
 - Coined by Donna Harraway: data visualization is a trick because it makes the viewer believe that they can see everything, all at once, from an imaginary and impossible standpoint. But it's also a trick because what appears to be everything, and what appears to be neutral, is always what she terms a partial perspective.
- *Visualization as rhetoric (80)*
 - "When visualizing data, the only verifiable fact is that it's impossible to avoid interpretation" - Jonathan Stray
- *Context is queen (91)*
 - A design choice made in one context or for one audience does not translate to other audiences
- *Demarginalizing the margins (95)*
 - Feminist human-computer interaction (HCI) scholar Shaowen Bardzell explains that the people pushed to the margins in any particular design context demonstrate who and what the system is trying to exclude. Subsequent work in HCI insists that designers then work to "demarginalize the margins" by recognizing intersections that exist, and engaging solidarity to navigate towards equity and inclusion.

Data Feminism in Action

PERISCOPIC, "DO GOOD WITH DATA"

In 2012, twenty kindergarten children and six adults were shot and killed in Sandy Hook, Connecticut. In light of this unconscionable tragedy, Periscopic began a new project: to visualize all the gun deaths that took place in the United States over the course of a calendar year. (73)

→ <https://periscopic.com/#/>

DR. NIECA GOLDBERG, *WOMEN ARE NOT SMALL MEN*

in her book *Women Are Not Small Men*, she found that heart disease in women unfolds in a fundamentally different way than men. The vast majority of scientific studies - not just of heart disease, but of most medical conditions are conducted on men, with women viewed as varying from this “norm” only by their size. (83)

→ https://www.goodreads.com/book/show/1776520.Women_Are_Not_Small_Men

GUERRILLA GIRLS, AN ANONYMOUS COLLECTIVE OF WOMEN ARTISTS, PUBLISHED AN INFOGRAPHIC, *DO WOMEN HAVE TO BE NAKED TO GET INTO THE MET MUSEUM?*

The graphic makes a data-driven argument by comparing the gender statistics of artists collected by the Met to the gender statistics of the subjects and models in the artworks. The Met readily collects paintings in which women are the (naked) subjects but it collects very few artworks created by women artists themselves. (85)

→ <https://www.tate.org.uk/art/artworks/guerrilla-girls-do-women-have-to-be-naked-to-get-into-the-met-museum-p78793>

COMING HOME TO INDIGENOUS PLACE NAMES IN CANADA BY MARGARET PEARCE, A CARTOGRAPHER AND MEMBER OF THE CITIZEN POTAWATOMI NATION

This map depicts the land that is known in contemporary Anglo-Western context as Canada, but without any of the common colonial orientation points. Pearce leverages the authority of the god’s eye view to challenge the colonizer’s view to advocate for a “reseeing” of the land under terms of engagement that recognize Indigenous sovereignty and respect Indigenous homelands. (92)

→ <https://umaine.edu/canam/publications/coming-home-map/coming-home-indigenous-place-names-canada-pdf-download/>



Chapter 4 - What Gets Counted Counts

Principle: Rethink Binaries and Hierarchies

→ Data feminism requires us to challenge the gender binary, along with other systems of counting and classification that perpetuate oppression.

Key Definitions:

- **Paradox of exposure** → the double bind that places those who stand to significantly gain from being counted in the most danger from that same counting (or classifying) act (105)
- **Patriarchy** → a term that describes the combination of legal frameworks, social structures, and cultural values that contribute to the continued male domination of society (108)

Key Themes

- *Questioning classification systems (105)*
 - Lurking under the surface of so many classification systems are false binaries and implied hierarchies. Decades of feminist thinking have taught us to question why these distinctions have come about; what social, cultural, or political values they reflect; what hidden (or not so hidden) hierarchies they encode; and, crucially, whether they should exist in the first place.
- *Rethinking binaries in data visualization (111)*
 - By challenging the binary thinking that erases the experiences of certain groups while elevating others, we can work toward more just and equitable data practices and consequently toward a more just and equitable future.
- *Refusing data, recovering data (115)*
 - Questions about counting must be accompanied by questions about consent, as well as of personal safety, cultural dignity, and historical context.

Data Feminism in Action

BORN EQUAL. TREATED UNEQUALLY, AN INTERACTIVE FEATURE IN THE TELEGRAPH THAT EXAMINED THE GENDER GAP IN THE UNITED KINGDOM ALONG A NUMBER OF DIMENSION

Although the authors treated genders as a binary category, they used color to challenge the stereotypically man/woman color coding. An example of communicating clearly without reinforcing stereotypes. (111)

→ <https://www.telegraph.co.uk/women/business/women-mean-business-interactive/>

DOES THE NEW CONGRESS REFLECT YOU?, AN INTERACTIVE THAT APPEARED IN THE GUARDIAN

Users select their own demographic characteristics to see how many people like them are in the 2018 Congress. Clicking on “trans + nonbinary” leads to a blank map showing zero people in Congress like you. The absence of data becomes an important takeaway, as meaningful as the data themselves. (113/114)

→ <https://www.theguardian.com/us-news/ng-interactive/2018/nov/15/new-congress-us-house-of-representatives-senate>

BEYOND XX AND YY (2017), AMANDA MONTAÑEZ, A DESIGNER FOR SCIENTIFIC AMERICAN, CREATED AN INFOGRAPHIC TO ACCOMPANY AN ARTICLE ON THE EVOLVING SCIENCE OF GENDER AND SEX

Amanda discovered that insisting on binary categories of data collection - with respect to gender, to sex, to their relation, or to anything else - fails to acknowledge the value of what (or who) rests in between and outside. (113)

→ <https://www.scientificamerican.com/article/beyond-xx-and-xy-the-extraordinary-complexity-of-sex-determination/>

MAKE THE BREAST PUMP NOT SUCK, HACKATHON

An ongoing forum for sharing stories, hacking pumps, and reengineering the postpartum ecosystem that surrounds them. (121)

→ <https://makethebreastpumpnotsuck.com>

Chapter 5 - Unicorns, Janitors, Ninjas, Wizards, and Rockstars

Principle: **Embrace Pluralism**

→ Data feminism insists that the most complete knowledge comes from synthesizing multiple perspectives, with priority given to local, Indigenous, and experiential ways of knowing.

Key Definitions:

- **Data settings** → coined by Data studies scholar Yanni Loukissas, describes both the technical and the human processes that affect what information is captured in the data collection process and how the data are then structured (132)
- **Epistemic violence** → coined by Gayatri Spivak, the harm that dominant groups like colonial powers wreak by privileging their ways of knowing over local and Indigenous ways (133)

Key Themes

- *Disciplining data* (131)
 - Certain assumptions and anxieties remain consistent across different articulations about the need for tidiness, cleanliness, and order, and the qualities of the people who should be doing this work. They must be able to tame the chaos of information overload. They must “scrub” and “cleanse” dirty data. The underlying belief of these ideas, that data should always be clean and controlled, have tainted historical roots (eugenics). As data scientists, we cannot forget these roots, even as the ideas themselves have been tidied up over time.
- *The solidarity genius* (135)
 - Downplays the work of coalitions, communities, and movements that are - not coincidentally - often led primarily by women and people of color.
- *Multiplicity of voices* (136)
 - The idea that many voices, rather than one single loud or technical or magical one, results in a more complete picture of the issue at hand.

- *Co-liberation (141)*
 - Grounded in the belief that enduring and asymmetrical power relations among social groups serve as the root cause of many societal problems. Rather than framing acts of technical service as benevolence or charity, the goal of co-liberation requires that those technical workers acknowledge that they are engaged in a struggle for their own liberation as well, even and especially when they are members of dominant groups.
- *Embracing Pluralism*
 - Offers a way to work toward a model of data for co-liberation. This means transferring knowledge from experts to communities and explicitly cultivating community solidarity in data work.
- *Problem with the singular data scientist “wizard” (148)*
 - They will never defeat the matrix of domination alone, no matter how powerful their spells may be. But a well-designed, data driven, participatory process, one that centers the standpoints of those most marginalized, empowers project participants, and builds new relationships across lines of social difference - well that might just have a chance.

Data Feminism in Action

ANTI-EVICTION MAPPING PROJECT (AEMP)

Since 2003, San Francisco has had an escalating rate of evictions. Data is collected on the number of evictions, but no data is collected on how many of those people end up homeless, or which landlords are responsible for systematically evicting major blocks of the city. The AEMP is a self-described collective of “housing justice activists, researchers, data nerds, artists, and oral historians.” They are mapping eviction, and doing so through a collaborative, multimodal, and-yet-quite messy process. The main purpose is to document the effects of displacement and to resist it through critical and creative means. (125-130)

→ <https://antievictionmap.com/>



DESIGN JUSTICE NETWORK

The Design Justice Network is an international community of people and organizations who are committed to rethinking design processes so that they center people who are too often marginalized by design. They work according to a set of principles that were generated and collaboratively edited by their network.

→ <https://designjustice.org/>

DATA MURALS

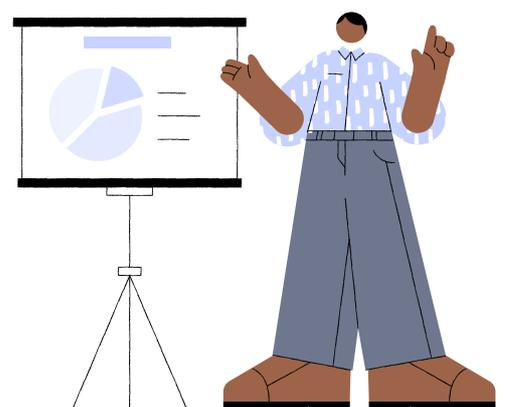
Rahul and Emily Bhargava partnered with community organizations to create data murals in public spaces. Each data mural originates from a need articulated by the community, rather than projected onto it by more powerful institutions. Each relies upon methods of data collection and processing. And each also incorporates an explicit process of knowledge transfer from external collaborators (consultants, academics, nonprofit specialists) to the community itself. (142-144).

→ <https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/3285>

THE GLOBAL ATLAS OF ENVIRONMENTAL JUSTICE

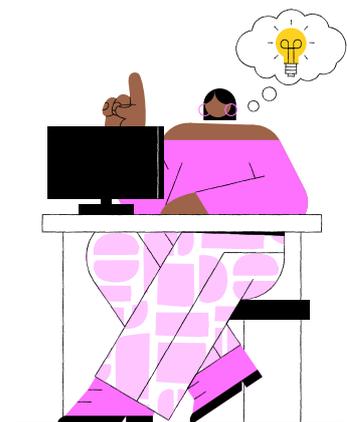
The Global Atlas of Environmental Justice works in partnership with activists, civil society organizations, and social movements to systematically document ecological conflicts around the globe. The EJ Atlas shows that scale is emphatically not incompatible with the feminist imperative to value multiple and local knowledges. (146-147)

→ <https://ejatlas.org/>



Discussion Questions

1. How might activating emotion-leveraging, rather than resisting, emotion in data visualization - help us learn, remember, and communicate with data? (77)
2. How did the field of data visualization arrive at a set of conventions that prioritize rationality, devalue emotion, and completely ignore the nonseeing organs in the human body? Who is excluded when only vision is included? (95)
3. Why is it important to question how our classification systems are constructed, what values or judgements might be encoded into them, or why they were thought up in the first place? (104)
4. What might be lost in the process of domination and disciplining data? Whose perspectives might be lost in that process? And conversely, whose perspectives might be additionally imposed? (131)
5. What might be gained if we not only recognized but also valued the fact that data work involves multiple voices and multiple types of expertise? What is producing new social relationships- increasing community solidarity and enhancing social cohesion- was valued (and funded) as much as acquiring data? (135)
6. What does data for co-liberation look like in action? Are there examples of feminist data science that value quantitative methods and pluralistic processes, data education, and community solidarity? (142)



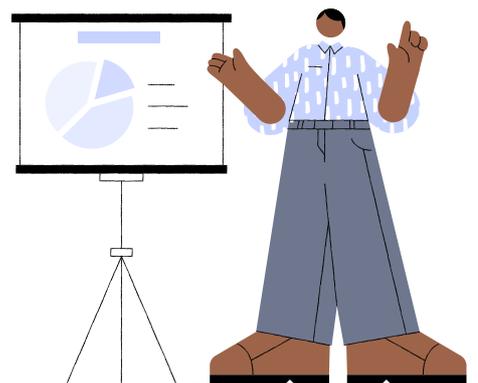
Chapter 6 - The Numbers Don't Speak for Themselves

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://eng318dataasrhetoric2017.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)

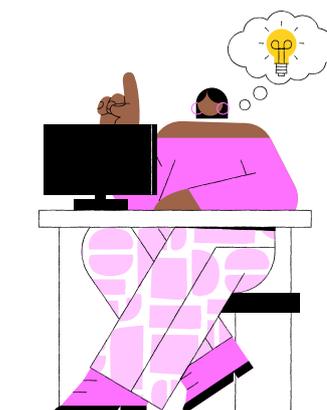
→ <https://www.womanstats.org/>

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

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 Question

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