Data Stories for/from All: Why *Data Feminism* is for Everyone

Yasamin Rezai <yxr178_at_miami_dot_edu>, University of Miami 🆔 https://orcid.org/0000-0002-3314-6677

Abstract

Looking through the intersectional feminist lens, Catherine D'Ignazio and Lauren Klein introduce data as a tool of power in the past and present world in their book *Data Feminism* (MIT Press, 2020) and reveal how authorities have used data as a weapon to maintain a hierarchy of power in favor of their position in the unequal status quo. By calling *data* a part of the problem, they also conceive of it as a part of the solution by analyzing how "data justice" can and ought to be redeployed to challenge power.

Data Feminism brings an intersectional feminist approach to any task having to do with data. Collecting, analyzing, creating datasets and later algorithms – it rightfully questions them all. The introduction provides readers with the logic behind why we need an intersectional feminist approach to data science and what the term offered by the authors, *data feminism,* means. Immediately explaining their approach in the Introduction, *Data Feminism* ""is a book about power in data science. Because feminism, ultimately, is about power too"" [D'Ignazio and Klein 2019]. The term *feminism* in the title is demystified upfront as well: the authors believe that *Data Feminism* "isn't only about women … it isn't only for women … it isn't only about gender either. Feminism is about power – who has it and who doesn't." [D'Ignazio and Klein 2019, 14]. *Data Feminism* uses feminism as a tool and critical thinking method to look at communities who have long benefited from data and those who were pushed out to margins by it.

[1]

The book offers a list of seven comprehensive acts to prompt data activism. To make Data Feminism possible, the authors provide seven main principles – examining the power, challenging it, elevating emotion and embodiment, rethinking binaries and hierarchies, embracing pluralism, considering context, and eventually making labor visible – and go through them in each of the seven chapters. *Data Feminism*, both as a book and now as a hopefully known term, is the best example of bridging the theory-work produced in academia with the outside world in a real-life setting. It calls for tangible action and not merely reflection from our minds and classrooms to policymakers and data policies. The book is neither an answer to data injustices nor an explanation of data ethics; it is a step-by-step guide to employing the reader's logic for forming better questions while moving ahead for a more just society because "None of us are free if some of us are not" [Forrest 2021] as the authors put it well.

[2]

Seven Principles

The first chapter's core idea is the exclusion of some communities caused by the "privilege hazard", meaning ignorance of being on top. While the chapter is grounded around the concept of the "matrix of domination" originally offered by Patricia Collins, with the main aim to identify and **examine the power**, it explains how data are extracted by dominant groups of people and mostly extracted from others. The authors find it crucial to understand how systems of oppression work with data as one main ingredient before taking the next step, **challenging the power**.

[3]

In the second chapter, the language employed to address the questions of ethics and values when discussing data and supporting algorithms is analyzed to reform those discussions. "Imagined objectivity" and concepts that uphold it are compared and contradicted with "real objectivity" and intersectional feminist concepts that strengthen it. The authors argue that the first "locates the source of the problem in individuals or technical systems" while the latter "acknowledges

[4]

structural power differentials and works toward dismantling them" [D'Ignazio and Klein 2020, 60]. For both, a terms table that distinguishes the two objectivities: *ethics* vs. *justice*, *bias* vs. *oppression*, *fairness* vs. *equity*, *accountability* vs. *co-liberation*, and *transparency* vs. *reflexivity* is offered to equip their readers with better terms.

The third chapter, a favorite, challenges the possibility of pure objectivity and neutrality of data visualization in order to **elevate emotion and embodiment**. It attempts to remind readers of the partiality of knowledge, drawing from Sandra Harding and Donna Haraway's "god trick"[Haraway 1988]. It argues that demonstrating uncertainty in data visualizations is feasible but still hard for people to understand. Therefore, regardless of how hard data graphics try to stay neutral and rational, individuals recognize them differently. One excellent and straightforward example is shown in weather forecasts and how the phrase "There is a 30% chance of rain tomorrow" can be interpreted in various ways by the public to mean "It will rain 30% of the time" or "It will rain in 30% of my area" [D'Ignazio and Klein 2020, 267]. On the other hand, it is discussed that devaluing emotion and affect and centering rationality as the authoritative mode of communication in reading data visualization might tell us about "who and what the system is trying to exclude" [D'Ignazio and Klein 2020, 95]. By assuming that rationality is assigned to male bodies and emotion to female bodies, the chapter criticizes the gendered binary of normative / derivative assigned to male bodies / female bodies.

As the fourth chapter quotes from Joni Seager, "What gets counted counts," and it stands for **rethinking binaries and hierarchies**. Classification of knowledge and constructed social categories used in science are introduced as other tools of power in data science. The Seager principle explains how binaries and constructed categories are the product of cultural and historical sets of values and biases of societies, and are therefore not only unreliable but also different from one society or community to another. Such binaries and classification systems are not extensible and need to be questioned and reconsidered. The peril lies where binaries are translated to codes and then encoded to technical systems. Such technical systems use data and datasets gathered and processed in the mentioned categories as input, process them into deliverable social policies or decisions, and perform them on human bodies and individuals. The authors invite readers to think of categories that, if left unquestioned, might be reductive, incomplete, or exclusive, and so too the policies and their deliverables.

Seeking multiple perspectives, the fifth principle and chapter, **embracing pluralism**, takes up the fourth chapter's argument with the same concern of the possible exclusion of local, Indigenous, or experiential ways of knowing. Through stories of projects such as the Anti-Eviction Mapping Project and the EJ Atlas, the chapter shows how not leaving the data work entirely to the ones who are aliens to the subject of a project – assuming that their neutrality in relation to the subject of the project is the only necessity – can be beneficial. Employing the communities, including marginalized ways of thoughts, and making the process more pluralistic by "synthesizing multiple perspectives" [D'Ignazio and Klein 2020, 125] can direct data projects toward more co-liberated models of data. Intending to make data models that promote *co-liberation* rather than *accountability* (terms that were counterposed with each other and discussed in the second chapter) is another way of feminist-thinking-through-data-work that makes Data Feminism possible.

The sixth chapter discusses one of the most critical and challenging principles: **considering context**. The authors discuss how numbers, or raw data, are alone not enough to show results in works of research that have to deal with humans. They offer that "raw data" are, in fact, already "cooked" in previous systems of bias or exclusion before becoming a number or data because they are obtained from people, regardless of the context in which the datasets were produced. Considering, acknowledging, and naming the systems of oppression at work is essential when talking about numbers, as "they do not speak for themselves" [D'Ignazio and Klein 2020, 159]. The authors show through various examples how a simple bar label and caption can point out context and make a difference in how it will be comprehended. Although tricky, a feminist approach can be taken to interpret or, in other words, speak for the numbers in different projects. "Letting the numbers speak for themselves" is viewed as unethical and undemocratic, with the potential to do more harm than good by reinforcing the unjust status quo [D'Ignazio and Klein 2020, 159].

Reminiscently, if nothing else, Data Feminism is a valuable project that by itself starts practicing the seventh principle of **making the labor visible**. If read between the lines, the authors leave to the audience two valuable lists as proof: first, a list of activists and scholars who are members of previously colonized nations, LGBTQIA+ and Indigenous

communities and people of color, and second, a list of corporations and projects led by people of color. "Nearly two-thirds of their citations are from women or non-binary people; almost every chapter has a project from the Global South; a third of all citations are from people of color, and nearly half of all projects mentioned in the book are led by people of color" [Shukla 2020].

Data for Everyone

There is no doubt that the consistency of their chain of thoughts and the logic that accompanies it are best presented in the threads of real-world well-researched stories. D'Ignazio and Klein master the art of storytelling in their scholarship by sublimely taking the readers' hands and walking them through one question to another. Each question begins with a story and slowly leads the readers to the following question, then the story leaves the readers for another story. Questions and stories, with a touch of the desire to discover a part of the truth behind data industry and its history, take the reader line by line and chapter by chapter with an excellent suspension element to find the complete picture of the main idea. D'Ignazio and Klein are not the first people who point out intersectionality, data ethics, or the history of colonialism, sexism, or racism, but they are some of the very few who treat them all as simple layers patiently and gradually added to a body of knowledge, forming an understandable array of new settings and concepts. As examples they define the "matrix of domination" in the first chapter and "auditing the algorithms" in the second chapter.

Among the readers' comments and reviews online from across the world, there is one question everybody asks, and many try to answer: who is this book *for*? A variety of answers have been given: teachers, sociology classes, or feminists in the digital age who work with data [Garbee 2020]. I firmly believe that this book is for *everyone*, and maybe that is why it offers such a consistent, easy-to-read, and easy-to-follow language. For scholars, teachers, or students in the humanities or social sciences, its main ideas raised on the intersectionality of race and sex might not be new. However, actively engaging with the book and considering teaching it in undergraduate or master's courses is suggested because of its valuable and essential intersection of theory with the most powerful asset of today: data.

For data scientists, engineers, managers, or anyone *in* the tech world or directly working with data, *Data Feminism* is an essential read. Not requiring the ability to read wordy, slowly-nuanced essays typical to the humanities, this book will neither disappoint experts in corporations nor confuse them. The book is full of straightforward anecdotes about tech companies and projects, both governmental and voluntary, big and small, that are good or bad examples of practicing what the authors call *data justice*. Important thematic and primary questions are raised by the authors and formerly by intersectional feminists such as Kimberlé Crenshaw, Donna Haraway, and others. Indeed this book will make readers ask the basic *wh-* questions, "data science for whom?, data science by whom?," and "data science with whose interests and goals?" [D'Ignazio and Klein 2020, 26, 33, 39].

For the ones who neither work with data nor the humanities, it is still worth the read because of the very simple fact that data surrounds everybody, or it will sooner or later in life no matter where people live – if it hypothetically has not yet. It gives readers invaluable insight into the social and data structure of the present world: how an individual is, or sometimes is not, being counted, what individuals count and sometimes do not, how an individual is being [mis]presented, and how the world around is being modeled, displayed and shown to them. It makes the reader think twice or more critically about "god tricks"[Haraway 1988] or the "facts" one has already accepted as their lifetime "assumptions."

Out of the scope of this book, the authors seemingly continue the work while engaging other scholars. There is a workspace on Slack with more than 300 members from all over the world, on which anyone can share ideas, news, articles, tweets related to the data economy "to nurture the imagination of what an equitable data economy might look like" [Remake Data-Driven World 2021]. The battle to unlock the invisible black boxes of data and biases, and fight against their effects has begun and it calls for active contributions and a great deal of critical thinking for everyone.

Works Cited

D'Ignazio and Klein 2019 D'Ignazio, C. and Klein, L. *Data Feminism* community review site. https://mitpressonpubpub.mitpress.mit.edu/pub/dgv16l22/release/6

D'Ignazio and Klein 2020 D'Ignazio, C. and Klein, L. *Data Feminism*. MIT Press, Cambridge (2020).

Forrest 2021 Forrest, J. "None of Us are Free If Some of Us are Not: Catherine D'Ignazio on Data Feminism," *Nightingale* (2021) https://medium.com/nightingale/catherine-dignazio-on-data-feminism-ce3b3c65f04a (accessed 15 October 2021)

Garbee 2020 Garbee, E. "Review of Data Feminism by Catherine D'Ignazio and Lauren F. Klein," *Issues in Science and Technology* (2020) https://issues.org/doing-the-work-data-feminism-review/ (accessed 15 October 2021)

Haraway 1988 Haraway, D. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies*, 14, 3 (1988): 575-599.

Remake Data-Driven World 2021 Remake Data-Driven World. *Slack* workspace (2021) http://remakedatadrivenworld.slack.com/ (accessed 15 October 2021)

Shukla 2020 Shukla, P. "Book Review: Data Feminism by Catherine D'Ignazio and Lauren F. Klein," *LSE Impact Blog* (4 October 2020) https://blogs.lse.ac.uk/impactofsocialsciences/2020/10/04/book-review-data-feminism-by-catherine-dignazio-and-lauren-f-klein/ (accessed 15 October 2021)