Not too many books about oppression have this exciting a cover. Credit: Amy J. Ko

Decoding Data Feminism

Amy J. Ko · Follow

Published in Bits and Behavior · 7 min read · Mar 15, 2021

👏 44   💬 1   🔖⁺   ▶   ⬆

When I first learned to code in middle school in the 1990's, *data* was the last thing on my mind. Code, for me, was about experiences: games, animations, and interactive stories, guided by hidden logic, to entertain my friends and I between classes or on the bus ride home. While all of these things had data, whether game art my friends had made in Microsoft Paint, or illustrations I'd painstakingly plotted pixel by pixel on my TI-82, or the dialog of my fan fiction text adventures, none of them felt like data. They were *content* that I'd created to express my ideas. Code, therefore, had little to do with data or information: it was about creative expression, loosely structured by variables and conditionals.

But while I was making, a revolution was afoot. Just a decade earlier, Visicalc—the world's first digital spreadsheet—had shown the world that computers could be more than just fast calculators. They were beginning to store vast amounts of data, previously encoded on paper in file cabinets, and beginning to become tools for processing that data to answer critical questions in science, business, government, and even health. Spreadsheets, databases, and programming languages emerged to support data analysis, all promising a future in which all the information we might need or want might be stored in a table, queried, analyzed, and reported instantaneously, giving everyone the power to make the most of exponentially growing data troves that modeled the world.

My story and the story of data science collided in 2001, when I was a senior in college. I was finishing an undergraduate thesis that investigated the challenges that scientists faced when using statistics programming environments like SPSS and Stata. I was fascinated by the many problems that scientists faced in making sense of the data they gathered, not only statistically, but computationally, as they tried to create reproducible analyses to answer their questions. My thesis work had found that code was

far from a panacea the 1990's had promised: APIs, programming languages, data wrangling, and myriad other challenges seemed to make analysis highly prone to error, largely because of the learning barriers they faced. I wanted to invent a programming language and environment that would make data analysis easier to learn and less error prone.

In my first few weeks of graduate school, I presented this project idea to my new doctoral advisor. I imagined an IDE with scripts, embedded data visualizations, verification to guide analysts toward sound analysis choices, exploratory data analysis affordances, and tools for manipulating data structure to support analysis. He dismissed it, suggesting that tools for doing statistics seemed niche, and probably not that interesting. I trusted him, and set aside the ideas bouncing around in my head. Several years later, R Studio, Jupyter Notebooks, and more all emerged, realizing many of those ideas. And now my advisor's research is all about interactive tools for data scientists. (No hard feelings Brad! Such is life in academia.)

Of course, none of these experiences remotely concerned questions about data and society. My ideas, and the ideas behind data science tools, were all fundamentally about *productivity* and *validity,* helping data scientists work faster and more rigorously to advance science and industry. And they still are. In dominant discourse on data science, there has rarely been any question about the data itself, where it came from, what it represented, what it was being used for, or what impact that might have on the world. My computer science education had taught me, after all, that data is inert, static, abstract, binary: how could it possibly be anything more?

D'Ignazio' and Klein's 2020 book, *Data Feminism,* answers that question clearly and compellingly, and shattered any notion in my mind that better tools would have any real impact on how data is used and abused in the

world. The book begins by explaining its use of the word *feminism*, which is often assumed to only concern women. The book's feminism is a broad one, invoking the much larger missions that have emerged through various waves of feminism to examine systems of power and oppression and work towards co-liberation. Data feminism, then, is about examining the uses and limits of data from the perspective of power, with a commitment to action.

While the book is not the first to examine these ideas in the context of data—and it readily admits this through the explicit and abundant description of prior works that have developed these ideas—it does neatly synthesize them into seven principles, which promise to powerfully organize and challenge the conventional activities of data science in academia and industry. They are:

- **Examine power.** *Data feminism begins by analyzing how power operates in the world.* The book is smart to begin here, as it frames data, data analysis, and data science practices as fundamentally about serving systems of power in the world. Power shapes who does the data science, and who the data science is done for (and to). The emphasizes the importance of asking whose interests and goals are served by gathering and analyzing data.

- **Challenge power.** *Data feminism commits to challenging unequal power structures and working toward justice.* Whereas the first chapter concerns the rational consideration of distributions of power in society, and how data is harnessed for power, the second chapter invokes a particular moral stance on power: that it should be distributed more equally, and used for more fair and just outcomes. The chapter notes the many ways in which data practices is often not only examined, but left unchallenged, leaving systems of power and oppression intact.

- **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.* This chapter primarily examines visualization, observing that current practices around making sense of data can make data feel clinical, sterile, and abstract, when in fact it is always visceral, usually representing some aspect of people's identities, experiences, and lives. It challenges data scientists to reify these visceral realities in representations of data and data analysis, so that consumers of data analysis aren't allowed to forget that data are people.

- **Rethink binaries and hierarchies.** *Data feminism requires us to challenge the gender binary, along with other systems of counting and classification that perpetuate oppression.* In the same way that visualizations can be reductive, the underlying mechanisms of measurement—counting and classification—can be as well. The chapter challenges data scientists to critically examine the consequences of these reductions, and if necessary, refuse reduction if those measurements do not serve just ends.

- **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.* This chapter asserts that the dominant ways of knowing in data science—namely computer science and statistics—are not only insufficient to achieve justice, but even unqualified. It observes that somehow these two disciplines were given immense power to shape how data is used, while none of the communities or expertise for which they are used were given any power. It challenges data scientists to not only embrace these other ways of knowing, but also give back power to communities, positioning data science experts instead as co-liberators, much like Costanza-Chock does in Design Justice.
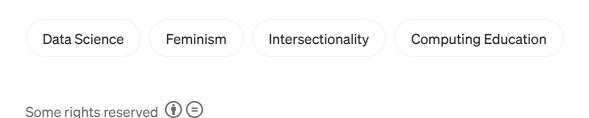
- **Consider context.** *Data feminism asserts that data are not neutral or objective. They are the products of unequal social relations, and this context is essential for conducting accurate, ethical analysis.* This chapter challenges objectivity, reasserting the basic truth that all good scientists know: rarely is data is objective. Rather, all data has a context, and that context is socially constructed, and because society is unequal, the resulting data mirrors those unequal structures. Any data scientist that ignores context is does not know what their data means and therefore cannot use it to meaningfully answer questions.

- **Make labor visible.** *The work of data science, like all work nt he world, is the work of many hands. Data feminism makes this labor visible so that it can be recognized and valued.* This chapter challenges the idea that data is somehow free; many data scientists view it this way because data sets are simply found, as if they were always there to discover. But the chapter observes that all data requires labor to produce, and that there are larger ecologies at play in this production that must not only be understood to interpret data and data analysis, but also credited when data is used to provide other value.

Together, these principles represent a weaving together of the activities of data science with feminist ideas, as well as those from critical race theory, providing a set of operating principles for data scientists to consider the nature of their work.

They also provide a strong foundation for how to frame data science education. For example, at the <u>University of Washington Information School</u>, we teach an introductory data science course (INFO 201) that covers data wrangling, R, and basic analysis and visualization. The course reaches over 1,200 students a year, many of whom go on to take more advanced data

science courses across campus. We've recently revised the course learning objectives to cover the seven principles above. We're not quite sure yet how to do this integration, but we have ideas. Some of the principles might appear as "warning labels" on technical content, helping students to understand the risks and responsibilities that come with something as mundane as computing a mean or excluding missing values. Other principles might come in the form of more substantial metadata on example data sets. In projects we might require students to critically examine the unintended consequences of their analyses. As we work through the pedagogy for these ideas, *Data Feminism* will be a helpful guide.

The book also connects to other work I'm leading in critical CS pedagogy, which seeks to prepare K-12 computer teachers to examine data and algorithms from the perspective of power. Much like *Data Feminism*, our work is developing weaving together ideas from Black feminist thought with concepts such as variables, control structures, data structures, algorithms, and programming languages. *Data Feminism* is complementary to this work, in its focus on data over algorithms; we can only hope that our own forthcoming writing can provide such a clear picture for how to reimagine our use of data in the world.

Data Science   Feminism   Intersectionality   Computing Education
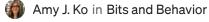
Written by Amy J. Ko

Follow

8.4K Followers · Editor for Bits and Behavior

Professor, University of Washington iSchool (she/her). Code, learning, design, justice. Trans, queer, parent, and lover of learning.

More from Amy J. Ko and Bits and Behavior

Amy J. Ko in Bits and Behavior

Amy J. Ko in Bits and Behavior

Paulo Freire's "Pedagogy of the Oppressed": a manifesto on…

Most Ph.D.s aren't professors

Critical consciousness through dialog

Everything one can do other than be professor.

11 min read · May 7, 2020

9 min read · Sep 1, 2017

500   7

412   1