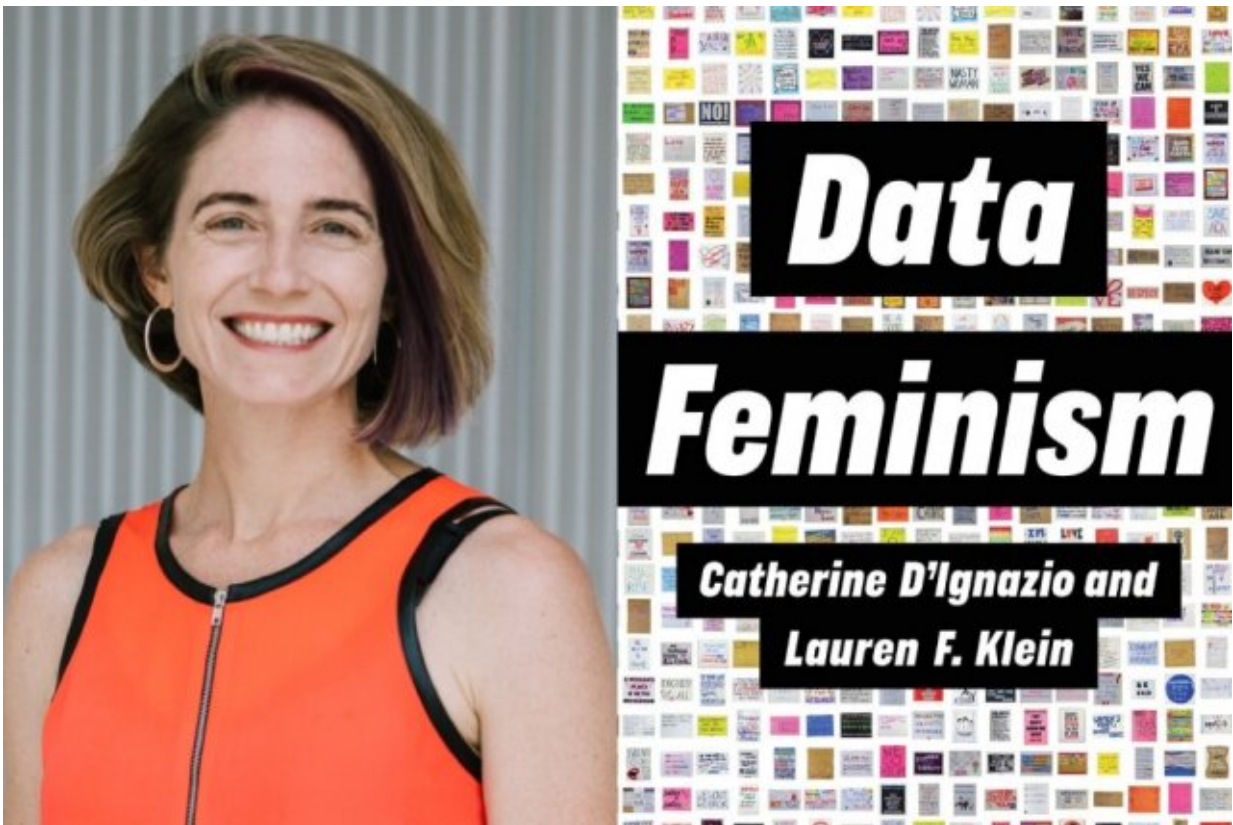# 'Data feminism' examines problems of bias and power that beset modern information

March 9 2020, by Peter Dizikes

Credit: Diana Levine and MIT Press

Suppose you would like to know mortality rates for women during childbirth, by country, around the world. Where would you look? One option is the WomanStats Project, the website of an academic research

effort investigating the links between the security and activities of nation-states, and the security of the women who live in them.

The project, founded in 2001, meets a need by patching together data from around the world. Many countries are indifferent to collecting statistics about women's lives. But even where countries try harder to gather data, there are clear challenges to arriving at useful numbers—whether it comes to women's physical security, property rights, and government participation, among many other issues.

For instance: In some countries, violations of women's rights may be reported more regularly than in other places. That means a more responsive legal system may create the appearance of greater problems, when it provides relatively more support for women. The WomanStats Project notes many such complications.

Thus the WomanStats Project offers some answers—for example, Australia, Canada, and much of Western Europe have low childbirth mortality rates—while also showing what the challenges are to taking numbers at face value. This, according to MIT professor Catherine D'Ignazio, makes the site unusual, and valuable.

"The data never speak for themselves," says D'Ignazio, referring to the general problem of finding reliable numbers about women's lives. "There are always humans and institutions speaking for the data, and different people have their own agendas. The data are never innocent."

Now D'Ignazio, an assistant professor in MIT's Department of Urban Studies and Planning, has taken a deeper look at this issue in a new book, co-authored with Lauren Klein, an associate professor of English and quantitative theory and methods at Emory University. In the book, "Data Feminism," published this month by the MIT Press, the authors use the lens of intersectional feminism to scrutinize how data science reflects

the social structures it emerges from.

"Intersectional feminism examines unequal power," write D'Ignazio and Klein, in the book's introduction. "And in our contemporary world, data is power too. Because the power of data is wielded unjustly, it must be challenged and changed."

The 4 percent problem

To see a clear case of power relations generating biased data, D'Ignazio and Klein note, consider research led by MIT's own Joy Buolamwini, who as a graduate student in a class studying facial-recognition programs, observed that the software in question could not "see" her face. Buolamwini found that for the facial-recognition system in question, the software was based on a set of faces which were 78 percent male and 84 percent white; only 4 percent were female and dark-skinned, like herself.

Subsequent media coverage of Buolamwini's work, D'Ignazio and Klein write, contained "a hint of shock." But the results were probably less surprising to those who are not white males, they think.

"If the past is racist, oppressive, sexist, and biased, and that's your training data, that is what you are tuning for," D'Ignazio says.

Or consider another example, from tech giant Amazon, which tested an automated system that used AI to sort through promising CVs sent in by job applicants. One problem: Because a high percentage of company employees were men, the algorithm favored men's names, other things being equal.

"They thought this would help [the] process, but of course what it does is train the AI [system] to be biased toward women, because they

themselves have not hired that many women," D'Ignazio observes.

To Amazon's credit, it did recognize the problem. Moreover, D'Ignazio notes, this kind of issue is a problem that can be addressed. "Some of the technologies can be reformed with a more participatory process, or better training data. … If we agree that's a good goal, one path forward is to adjust your training set and include more people of color, more women."

"Who's on the team? Who had the idea? Who's benefiting?"

Still, the question of who participates in data science is, as the authors write, "the elephant in the server room." As of 2011, only 26 percent of all undergraduates receiving computer science degrees in the U.S. were women. That is not only a low figure, but actually a decline from past levels: In 1985, 37 percent of computer science graduates were women, the highest mark on record.

As a result of the lack of diversity in the field, D'Ignazio and Klein believe, many data projects are radically limited in their ability to see all facets of the complex social situations they purport to measure.

"We want to try to tune people in to these kinds of power relationships and why they matter deeply," D'Ignazio says. "Who's on the team? Who had the idea? Who's benefiting from the project? Who's potentially harmed by the project?"

In all, D'Ignazio and Klein outline seven principles of data feminism, from examining and challenging power, to rethinking binary systems and hierarchies, and embracing pluralism. (Those statistics about gender and computer science graduates are limited, they note, by only using the

"male" and "female" categories, thus excluding people who identify in different terms.)

People interested in data feminism, the authors state, should also "value multiple forms of knowledge," including firsthand knowledge that may lead us to question seemingly official data. Also, they should always consider the context in which data are generated, and "make labor visible" when it comes to data science. This last principle, the researchers note, speaks to the problem that even when women and other excluded people contribute to data projects, they often receive less credit for their work.

For all the book's critique of existing systems, programs, and practices, D'Ignazio and Klein are also careful to include examples of positive, successful efforts, such as the WomanStats project, which has grown and thrived over two decades.

"For people who are data people but are new to feminism, we want to provide them with a very accessible introduction, and give them concepts and tools they can use in their practice," D'Ignazio says. "We're not imagining that people already have feminism in their toolkit. On the other hand, we are trying to speak to folks who are very tuned in to feminism or social justice principles, and highlight for them the ways data science is both problematic, but can be marshalled in the service of justice."

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: 'Data feminism' examines problems of bias and power that beset modern information