

# Data visualization from a feminist perspective

## Interview with Catherine D'Ignazio

BY NANNA THYLSTRUP AND KRISTIN VEEL

---

Catherine D'Ignazio is a scholar, artist/designer and software developer who focuses on data literacy, feminist technology and civic art. She has run breastpump hackathons, created award-winning water quality sculptures that talk and tweet, and led walking data visualizations to envision the future of sea level rise. Her research at the intersection of gender, technology and the humanities has been published in the *Journal of Peer Production*, the *Journal of Community Informatics*, and the proceedings of *Human Factors in Computing Systems (ACM SIGCHI)*. D'Ignazio is an Assistant Professor of Civic Media and Data Visualization at Emerson College, a faculty director of the Engagement Lab and a research affiliate at the MIT Center for Civic Media.

Nanna Thylstrup is a postdoc at the Department of Arts and Cultural Studies, University of Copenhagen. Her work is concerned with infrastructures, archives, technological imaginaries and mass digitization, and she is currently working on the critical data studies project *Uncertain Archives*.

Kristin Veel is associate professor at the Department of Arts and Cultural Studies, University of Copenhagen. She has published extensively on surveillance and archival technologies, information overload and in/visibility, and is currently PI of the critical data studies project *Uncertain Archives*.

**W**hat are the current political implications of the collection, analysis and spread of data visualization?

Data has become a currency of power. The most successful internet businesses make their money by aggregating data. Decisions of public import, ranging from which products to market, to which prisoners to parole, to which city buildings to inspect, are increasingly being made by automated systems sifting through large amounts of data. As a result, knowing how to collect, find, analyze, and communicate with data is of increasing importance in society. And yet ownership of data is largely centralized, mostly collected and stored by corporations and governments. Critically, the technical knowledge of how to work effectively with data is in the hands of a small class of specialists (who, as Kate Crawford points out, are mostly male and white). People are far more likely to be discriminated against with data or surveilled with data than they are to use data for their own civic ends. Because data in contemporary society is so intimately intertwined with power and inequality, this makes data, and its products in the form of visualizations, a timely and important object of analysis for feminist theory.

Data visualization, specifically, is emerging as a mainstream form of public communication that we see occurring with frequency in journalism, policy, advocacy, the arts, and other domains. But data visualizations wield a tremendous amount of rhetorical power. They seem to be generalized, scientific and to present an expert, neutral point of view. I should note that this is particularly true for people who do not make them. As visualizations travel out into the world from more specialized contexts, we may need to rethink what we visualize, how, when and for whom.

*What are the implications and potentials of exploring questions of gender and diversity through and in interplay with data visualization?*

There are many reasons that data visualization needs feminism right now. In my lectures, I frame these as ‘missing body problems’ in that the current data practices are leaving out certain bodies at every stage of the process, from collection and analysis to production and reception. While visualization has been heralded as a form of neutral analytic reasoning, these inequities, which have to do with the differential powers and privileges of human bodies, are obscured by that very same neutrality. Visualizations foreground apparent facts and obscure the bodies that contributed to their collection, creation and interpretation.

*To what extent is data visualization, as a form of quantitative data science, also a question of design – and how do these design questions relate to the question of gender?*

A data visualization is a designed artifact in the same way that a piece of furniture has been designed or that a marketing brochure has been designed. Just because there are numbers and some amount of informational complexity to visualizations doesn’t excuse them from being rhetorical objects. This is not to detract from their truthfulness, just to put their communication goals into the context of other objects that communicate.

In regards to how this relates to gender it relates back to the missing body problem. In the rush to work with ‘big data’ we are literally missing the bodies who, in various ways, are impacted by data. I have identified four *missing body problems* to begin with which are as follows:

### 1) Bodies are extracted.

There is profound inequity and asymmetry in data practices. It is states and corporations (increasingly state-like) who have the resources to collect, store, maintain, analyze and derive insight from large amounts of data. What this means is that bodies are extracted and ‘mined’ by states, corporations and institutions.

### 2) Bodies are absent.

Women and people of color are underrepresented in data science just as they are in STEM (Science, Technology, Engineering, and Mathematics) fields as a whole. Kate Crawford has characterized this as “Artificial Intelligence’s White Guy Problem” in which the pressing challenge to humans working with smart, data-driven systems is not that computers may ultimately outsmart us but that humans might make computers dumber by encoding our age-old biases and structural inequalities into the system.

### 3) Bodies go uncounted.

There is differential counting in the data collection and analysis process. First, there is the issue of what people in power decide is worthy of allocating scarce resources towards quantifying. This is why we have detailed data sets on gross domestic product and erectile dysfunction but poor data on hate crimes and the composition of breastmilk. In some cases where reliable data is collected, it may not be disaggregated into proper categories to make gendered, racial or other patterns apparent. For example, charting trends such as cell phone penetration in rural Africa may show an upward pattern for men and a different pattern for women.

### 4) Bodies are rendered invisible.

Data visualizations wield a tremendous amount of rhetorical power, particularly for people who do not make them and who are not part of an expert community in which

they circulate. Donna Haraway characterizes it eloquently as the “the god trick of seeing everything from nowhere”. The trick, in this case, is precisely that the bodies involved in rendering an abstract, aggregated, generalized view of some set of information have been rendered invisible. There are no bodies. There is no perspective. There is only the data.

*Which values and principles might a feminist data visualization practice include and adhere to?*

Why, thank you for asking! In a short paper for IEEE Visualization in 2016, Lauren Klein and I outlined six design principles for Feminist Data Visualization. It’s important to note that these principles apply to *content*, *form* and/or *process*. It’s probably worth explaining what this means.

A data visualization could be feminist in *content* if it is topically about women, other marginalized groups, and/or issues of power and inequality. It could be feminist in *form* if the resulting artifact incorporates multiple voices and authors, values emotion/affect/embodiment, or is self-reflexive in regards to the standpoint of its authors. A visualization could be feminist in *process* if the way it was produced was participatory, pluralistic, self-reflexive, considers context, and incorporates marginalized voices.

That said, here are the six principles in brief:

1. Rethink binaries: Feminist theory disavows binary distinctions. This doesn’t mean NEVER using a binary way of capturing data and variables, but to double check that binaries (like male/female) are the most appropriate way of capturing the data that you need. Could some variables like gender be represented as continuous and multidimensional rather than as binary distinctions?

2. Embrace pluralism: Feminist theory would assert that different bodies discover different truths in the world. By embracing pluralism, a data visualization can include more voices in the process and the form. This can mean being self-reflexive about the designer's own perspective. Or it could also mean deliberately incorporating the voices of others in the process. Or it could mean helping the end user discover their own truths through engaging with the visualization. For example, Rahul and Emily Bhargava work with community organizations to build their capacity to analyze data and then tell a public story about their work. This participatory process results in 'data murals' which are large-scale public paintings, conceived and painted by many different people. More info here: <https://datatherapy.org/data-mural-gallery/>

3. Examine power and aspire to empowerment: A feminist approach interrogates power in various forms (including her own design team) and strives for inclusion of marginalized perspectives. This one can be difficult because often an agency commissions a data visualization to serve a celebratory or PR function for their institution. Creating visualizations that examine power structures may take courage to challenge norms and institutions.

4. Consider context: All knowledge is *situated*. Acknowledging this basic fact means understanding when data visualization (and possibly even data collection) is not the right thing to undertake. Who is being collected for whom? Considering context also means determining, through consultation and deliberation, what output form makes sense for a particular audience. A data visualization for community gardeners may look different than a data visualization for middle school students may look different from a data visualization for newspaper readers.

5. Legitimize affect and embodiment: Experiences that derive from sensation and emotion

are genuine ways of knowing and understanding the world. Artists, designers and journalists have begun to undertake more vibrant experiments with the affective dimensions of data visualizations, creating quilts, murals, sculptures, walks and installations. For example, the Elevator Repair Service Theater collaborated with the Office of Creative Research to create a performance of the metadata of the Museum of Modern Art. They grouped artists by first name and ordered them by the number of works in the collection. What this means is that "John" came first, followed by "Steven", "Matthew", "James" and so on. The performers read these names as they appeared on a computer tablet and it takes 3 minutes to arrive at the first female name - "Mary". Using performance, the audience feels the gender imbalance of the museum's collection rather than seeing it all at once in a bar chart.

6. Make labor visible: So many bodies and so much time is involved in the collection, cleaning, storing and cataloging of data. Who are the actors – both institutional and individual – who have labored to generate a particular data set? Can we visualize some of the human and material infrastructure that makes the data possible? For example, the citizen science group Public Lab has created a low-cost mapping technique in which cameras hang from kites and balloons to capture aerial imagery. One of the most interesting by-products of this way of mapping is that the resulting artifacts often capture the bodies of the people doing the mapping. (See this image from a Public Lab research note by Eymund Diegal about mapping sewage flows in the Gowanus Canal. Note the people on boats doing the mapping and the balloon tether that links the camera and image back to their bodies.)

*Which design strategies exist for feminist data visualizers? And which need to be invented?*

Some of the design principles illuminated

above are starting points. And, in general, I think frameworks like design justice, co-design and participatory design are great starting points for existing ethical design strategies that can be better integrated into data visualization design theory and practice.

However, there are certain things that are specific to visualization that we need to invent. The first is to invent new ways to represent uncertainty, outsides, missing data, and flawed methods. While visualizations – particularly popular, public ones – are great at presenting wholly contained worlds, they are not so good at visually representing their limitations. Where are the places that the visualization does not go and cannot go? Can we put those in? How do we represent the data that is missing? This starts to point to better, more visual ways to represent the *provenance* of the data. Another form I have been experimenting with recently in classes is to have students write *data biographies*. Instead of taking a data set and analyzing it as-is, I ask students to track down the history of the data, why it was collected, who does the collection, how it is stored, and who does it impact. By understanding how human (and how fallible and prone to error) these processes are, learners start to wrap their heads around what stories can and cannot be told with the data at hand.

Another design strategy we may need to invent has to do with how we might make dissent possible. While there are plenty of ‘interactive’ data visualizations what we currently mean by this is limited to selecting some filters, sliding some sliders, and viewing how the picture shifts and changes from one stable image to another stable image as a result. How can we devise ways to talk back to the data? To question the facts? To present alternative views and realities? To contest and undermine even the basic tenets of the data’s existence and collection? A visualization is often delivered from on high. An expert designer or team with specialized knowledge finds some data, does some wizardry and presents their artifact to the world with some highly prescribed ways to view it. Can we imagine an alternate way to include more voices in the conversation?

## REFERENCE

- Crawford, K. (2016): Artificial Intelligence’s White Guy Problem. *The New York Times*. Retrieved from: <http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>