

Transcript:

Databite No. 131: Data Feminism

Catherine D'Ignazio and Lauren F. Klein

Recorded: 05/13/2020

Sareeta Amrute:

Hello and welcome to Databite number 131. My name is Sareeta Amrute, Director of Research here at Data & Society. I will be your host for tonight supported by my team behind the curtain, CJ, Rigo and Eli.

For those of you who don't know us yet, Data & Society is an independent research institute studying the social implications of data and automation. We produce original research and convene multidisciplinary thinkers to challenge the power and purpose of technology in society. You can learn more about us through our website, datasociety.net. We will be spending the next hour together so let's get ourselves grounded. Throughout the hour, you may ask questions via the Q&A function at the bottom of your screen. We will archive all questions and try to address as many as we can during this event. Closed captioning is available by clicking on the icon at the bottom of your screen. As a reminder that this event will be recorded and shared afterwards, we will notify all registered attendees as soon as the link is available.

Spatially, Data & Society is located in what we now refer to as New York City; a network of rivers and islands in the Atlantic Northeast, home to the ancestral unceded territory of the Leni Lenape people. It is with their permission that I want to open up this space. Land acknowledgements are acts of truth-telling that recognize the struggle of the dispossessed but often fail to name the mechanisms by which indigenous lands were legally seated. These were deliberate design-based decisions taken under the logic of white settler colonial expansion and because Northern European settlers kept pretty good records, we have their receipts. Right now I'm in Brooklyn, the ancestral home of the Carnarsie people. When I think about the current map of Brooklyn, I think about the many pathways that native people used that are now our main thoroughfares and roads like Atlantic Avenue and

Flatbush. Catherine D'Ignazio is joining us from the land of the Wampanoag Nation and Lauren F. Klein is on the land of the Muscogee Creek people.

If you'd like, take a moment to post in the Q&A where you are sitting and provide a land acknowledgement of your own while I turn things over to our speakers to learn more about *Data Feminism*.

Catherine D'Ignazio:

Hi everyone, my name is Catherine D'Ignazio. Tonight was to be our very celebratory New York book launch party because both of us are major fans of Data & Society's work. Sadly we can't be here in person, but we're really excited to still be included as a data byte and to participate with all the folks who have joined us tonight to help launch this book into the world for New York people and many others, as I'm seeing that people are from Australia and Oakland, California- Ohlone land- and Manchester, UK and so on. Those are folks who probably wouldn't have been with us otherwise so we're really excited to share it with you. I'm an assistant professor of Urban Science and Planning at MIT and I'm the director of something called The Data + Feminism Lab there. I'll turn it to Lauren.

Lauren F. Klein:

Hi everyone. Thank you so much for joining us. As Catherine mentioned, we're sad that we couldn't do this event in person, especially as I'm from New York. It would have been great to be there with you all but the virtual event means that more people can attend and so we're really excited to be able to share this work with you. In my professional life, which seems hard to remember, I'm an Associate Professor of English, and Quantitative Theory and Methods at Emory University, where I also direct the Digital Humanities Lab. I think that's all for the intro so I'll pass it back to Catherine.

Catherine D'Ignazio:

Great. We see *Data Feminism* as a growing body of work that's holding corporate and

government actors accountable for their racist, sexist, and classist data products. Things like- and I think Data & Society folks have heard about these in the past already- fixing face detection systems that cannot see women of color, hiring algorithms that demote applicants that went to all-women schools, search algorithms that circulate negative stereotypes about black girls, child abuse detection algorithms that punish poor parents, data visualizations that reinforce the gender binary, all of these things and more. We've put down at the bottom of this current slide some of the inspirations and work that we draw on as part of our own work.

We sort of situate ourselves in contrast to this sort of techno-hype message, this idea that data is the new oil. This is something folks have certainly heard as it's been a kind of meme in the data and technology world since the Economist Magazine said it. It's now been at least eight or ten years and it's been repeated over and over again, meaning data will yield big extraction extractive profit. But there's been this really incredible and inspiring pushback, even since we've been writing this book, coming from women of color, white women, indigenous people on immigrant communities, LGBTQ folks, and more, pushing back and saying, "Actually there isn't really anything all that new, data is the same old oppression that we have been seeing for a long time." So that leads us to what we bring to this conversation which is a focus on feminism, and a focus on intersectional feminism in particular.

Before we get to the main argument of the book, about why data science needs feminism kind of desperately, we thought we'd do some level-setting about feminism. So, thinking about what is feminism in the first place? Following Beyoncé's definition, a feminist is about a belief. A feminist is a person who believes in equal rights for men and for women and for non-binary people. But at the same time, feminism is not only a belief but is also organized activity on behalf of women's and non-binary people's rights and interests. So it's belief and it's action. And here I will turn it to Lauren.

Lauren F. Klein:

Oh great. So as Catherine said, feminism is a belief, it's action, organized action, and

feminism can also mean a set of theories and ideas. These theories begin by thinking through issues of inequality and respect to sex and gender. The past 40 years of scholarship and the current political reality have brought many more dimensions of inequality into the conversation and these include race, class, ability, sexuality, and more. And this leads to the most important takeaway of this very brief intro and overview, which is that feminism in the year 2020 must be understood as intersectional. Many of you may be familiar with this term but if not, it was a term coined by the legal scholar Kimberlé Crenshaw, which she uses to explain how social inequality cannot be explained by only one dimension of difference, such as gender. So when we talk about inequality or oppression we must be talking about the intersection of the many factors and the forces that produce it, such as racism, classism, colonialism, and so on.

The key thing to understand about intersectionality, and it's a thing that's often overlooked, is that intersectionality doesn't just describe markers of individual identity and their [inaudible] it's the structural forces of power and their intersection that produce those effects. It's the work of women of color feminists and black feminists in particular that have really done the work of foregrounding this conversation about structural forces. One final note about intersectionality. While Kimberlé Crenshaw coined the term, the idea was described by many others before her, probably most famously by the Combahee River Collective, who in the late 1970s described systems of oppression as "interlocking". And even before that, in the 19th century, there were black women scholars and activists like Anna Julia Cooper, Frances Harper, Sojourner Truth even, who described intersectionality in practice, if not by name.

To sort of sum up, intersectional feminism, which provides the underlying framework for our book, is not just about women and gender. It's about power. It's about who has it and who doesn't. And in today's world, as you can see, data is power. And so our gambit is, intersectional feminism, when applied to data science, can help that power be challenged and changed. And so our argument is really that data science needs feminism, and intersectional feminism in particular, if we ever hope to overturn the power imbalances that we see in our data sets, in our data systems, and how data affects our lives in the world.

Catherine D'Ignazio:

So here's the sort of road map of the book. When Lauren and I were sitting down and starting to draft the book, we looked across a wide variety of literature. Feminism is a really interesting thing because there's feminist...you know, one of the things that Lauren was saying about how feminism is a theory or set of ideas is one of the most exciting aspects of it to me because it's a kind of intellectual heritage that we inherit and then kind of operationalize and mobilize to adjust new circumstances. We looked across a lot of work that's happened across a lot of different fields. We also asked ourselves what we have learned from our own schooling and feminism, our experience in different activist communities, and we came up with seven principles that to us encapsulate the most important aspects of intersectional feminism as they relate to data. And so this is also how the book is structured. We have these principles, and then each chapter is about one of these principles. So we have examine power, challenge power, rethink binaries and hierarchies, and so on, and we treat those at length in the chapter. So the goal here was not to make new feminism, not to make new feminist theory, but to operationalize existing ideas for data science. To really think about models that might guide the work of people who are working with data already, people who want to work with data, or people who want to refuse and step away and push back on working with data. For the rest of this talk, we don't have time to go into all of these things in depth, all the principles, so we're just going to show you three different examples that illustrate one of the principles in the book.

Lauren F. Klein:

Great, so we'll start with this one. This is principle of examining power. Examining power is obviously central to the feminist project because of how gender inequality, as I've already said several times, is at root a question of power. And one of the contributions of feminist organizers and activists, and also of theorists, is to give us models that show how power operates in the world. So for example, the famous black feminist sociologist Patricia Hill Collins describes power as what she terms the "matrix of domination" by what she means that power operates not just from the top down, like the government saying women can't vote, but across many layers of society.

We go into this in more detail in the book but the key point here is that once we have a model for how power works in the world, we can start to understand how it operates, and then how to change it. And in the book, in the chapter on power, we tell the story of Mimi Onuoha's efforts to collect which she calls "missing data sets." These are data sets that a reasonable person might expect to exist because they address issues of really pressing social need but because of various reasons they don't actually exist in real life. So data sets like trans people killed or injured, and instances of hate crime. There's no comprehensive database on this. People excluded from public housing because of criminal records, this is a thing that happens to people but because the people in power don't recognize it as such we don't have data on this. Or for a very timely example, like a gender [inaudible] in the United States, we just don't have this [inaudible]. So one version of Onuoha's project is a GitHub repository. You can actually see this on the right. It just lists these missing data sets. But another instantiation of this is a physical artwork. It's what you see on the left. It's a file cabinet, each with folders inside and each one has a label with one of the data sets. And the idea is that you tab through the folders and look at the labels but when you open up the folder the data set is missing. It's just not there. And as Onuoha explains, and you can see this actually in the GitHub repository which contains her artist statement, these missing data sets "reveal our hidden social biases and indifferences." By calling attention to these data sets as missing, she also calls attention to why these data sets are missing. They're missing because of a lack of personal, social, political, or governmental will, or some combination of all of those. Or in the case of the Coronavirus tests, they're missing precisely because of political and governmental will. But in either case, since the data are missing, we can't move forward with our goal of working towards greater justice in the world.

Catherine D'Ignazio:

A second example that builds off of this, and this comes from chapter two which is about challenging power, is the example of femicides in Mexico, and also in basically every other country, but we talked specifically about Mexico. This is another case of missing data sets. In the book we tell the story of María Salguero, who resolved basically to head straight to the problem and collect the missing data herself. So just to explain a little bit, femicides are

gender related killings of women and girls, they include cis and trans women. They are legally defined as crimes in a handful of countries including Mexico, so there's actually a legal framework that characterizes this as a crime, but the state is not systematically collecting data on femicides. And so they're the subject of emerging public anger in Latin America. You can see the hashtag here, #NiUnaMenos, which means "not one less woman."

So the state is sort of neglecting to fully implement its own laws and provisions and to actually measure the scope and the scale of the problem. Salguero was a kind of individual citizen, she was frustrated by the lack of action and so in 2015 she started single-handedly compiling a database. She's now been doing it for five years and at this point she's amassed the largest public database of femicides for the entire country. She spends two to four hours a day logging these deaths on a Google Map that she culls from media reports. She has helped families locate their loved ones, she's shared her data with journalists and activist organizations, and she's even testified in front of Mexico's congress multiple times about the issue. And so in the book we characterize this as an example of feminist counter data.

A way to do activist data collection that steps in when the state and other institutions have systematically found to ensure the basic safety of their populations. So it's one way to use data to challenge power but it comes with an important caveat which is that not all problems are problems with missing data in the first place. And also that not all problems can be addressed by collecting counter data. We characterize this kind of, we say this repeatedly in the book, but data is really a double-edged sword. So more data is not always better, right? Because sometimes more data puts vulnerable people in more paths of harm. So other strategies to challenge power also include things like auditing algorithms, teaching data science like an intersectional feminist would, and centering equity and justice instead of, or in addition to, ethics and data science.

Lauren F. Klein:

So another thing that feminism can do is not just help us identify issues to address but also to inform the process of data science work. This example comes from a sort of a combination of principles. It comes from the chapter on embracing pluralism. The idea

derives from Donna Haraway's idea of situated knowledge and her view that the most complete knowledge comes from bringing together multiple perspectives. So in this model, knowledge is not top-down but is created through dialogue and exchange. And ultimately, because you're bringing together all of these different perspectives, it results in a more complete picture of the problem at hand. And we see this in the example of [inaudible] the large image on the left, also known as the AEMP. They are a self-described collective of "housing justice activists, researchers, data nerds, artists, and oral historians." And since 2013, the AEMP has worked to quantify and organize around the housing crisis in San Francisco and the greater Bay Area. They work in collaboration with tenants rights organizations and community groups and they also create oral histories which is what you see here in this Narratives of Resistance and Displacement Map. Each blue dot on the map leads to a video story from a single person or a family who is facing displacement from their home.

So in the book we contrast this with the Eviction Lab which you can see in the smaller image on the right, which is based at Princeton University. This lab's goal is to present a national picture of the eviction crisis. And we should say that this is a really worthy goal and it's a valuable project, but it's really widely different in terms of process, right? The Eviction Lab's map derived from seemingly bigger data and the map presents a seemingly more comprehensive picture of the problem of eviction in the United States. But the AEMP has actually shown that national real estate databases, like the ones that the Eviction Lab uses, significantly undercount evictions because there's a sort of litmus test as to what qualifies as an eviction and as many people have firsthand experience with, you can be pushed out of your home in a lot of different ways. Working instead with local tenants rights organizations, the AEMP has gathered probably messier, but actually much more comprehensive and more contextualized data, that documents a greater extent of the problem at hand.

Catherine D'Ignazio:

So this brings us to the last point I want to make in the presentation, which is that one of the things we make, and now that I'm thinking about it I'm not sure we even made it

strongly enough, but I think that data feminism requires an expanded definition of data science. So our argument is that data science is not about the size of the data. It's not about the sophistication of the analysis methods. It's also not about the technical credentials of the people undertaking the work or the places that they're affiliated with because these dimensions are always continuously, repeatedly, still today used to exclude women and people of color from the field, as well as to exclude work whose contribution, whose innovation, is socio-technical rather than just purely technical. By expanding the book, and that's what we try to do throughout the book in terms of whose examples, from where do our examples come.

We see that some of the most exciting work in data science today is actually being undertaken by artists, by journalists, by humanists, by community organizers, and activists. So here, to tell you whose work we're showing on the slide, we want to give a shout out to Margaret Mitchell and her team for her research on bias in natural language processing. Artist Stephanie Dinkins is pushing the boundaries and the scale of data with her interactive, talking sculpture that was trained on an intergenerational dialogue between black women in her own family. On the right, you can see The Pudding invented a fun and super interesting data journalism which is exposing gender bias in Hollywood screenplays.

And then finally at the bottom is a data mural by the group Data Therapy, where they work with community-based organizations to create data murals in situ with people from the community. One of the reasons we're saying this here too, is because in the open review process for our book, we actually got a number of comments like, "Oh, these seem like nice little projects that you're including but they're not real data science." So we realized that we really need to be very clear that we consider these works data science. So that, I think, is one of our arguments, and I think a way to enlarge the field and the playing field and the table and who's at the table, of what do we actually mean by data science.

Lauren F. Klein:

Here's just a little bit of review. So data feminism is data science that exposes it and challenges power. It's led by, and ideally centers, minoritized people. It can be a counter

data science about the injustices created by mainstream data science. It's unfortunate that we were in this situation but that is a reality of life today. It looks at many axes of inequality including but not limited to gender. [inaudible] It considers process, how inequality permeates all stages of a data science project from funding and choice of research to undertake, outside deployment and circulation, of the product. And then it credits labor. It acknowledges how data science is the work of many hands. So I think that's all we have for our formal remarks today. Thank you so much for listening and now eager to hear your questions.

Sareeta Amrute:

Thank you so much for your talk Catherine and Lauren, it was really beautiful. I'd love to learn more. Everyone who's tuning in, please post your questions to the Q&A and I will try to ask as many of them as I can before we run out of time. But as the moderator, I'm going to take a little bit of my moderator's prerogative to ask you a few questions to start out. My first question is about the work of Patricia Hill Collins. You both used Patricia Hill Collin's scholarship in your work. I've been seeing her popping up quite a bit lately and it's really gratifying. Can you tell us a little bit about when you were first introduced to her scholarship and why her matrix of domination plays such an important role in your own?

Catherine D'Ignazio:

Sure. I'll take that first answer there. For me, I ran a big sort of research project/feminist hackathon called Make the Breast Pump Not Suck hackathon. This is now a couple of years ago, so back in 2017, we were planning that. That for me was actually the first time I had come across her work because we were doing a lot of work, it was a racial equity and socioeconomic equity focused hackathon in terms of who we were working with and who were the innovators at the table. And so for me it was through a webinar with the Black Mamas Matter Alliance, which is a maternal and birth justice group here, where they were talking about uplifting the voices of black leaders in the maternal health crisis in the United States. That's where I first learned about it, and it was in the framework of birth justice. But then, I also have written really closely following Sasha Costanza-Chock's work on design

justice and they also draw on the matrix of domination. That was also really inspiring to see so it feels very current, even though the book was written many years ago. I also am enjoying the fact that it is being brought back in a number of different ways and feels really relevant for this present moment with data and AI.

Lauren F. Klein:

Yeah, I mean, I guess I'll just say that I am and have always been a big theory-head and so I can get down on a lot of theories of power. And I think one of the things that's really important about Patricia Hill Collins' work, and actually it's interesting, a formal philosopher read our book on Twitter and then was like, "This is undertheorized. It has good examples but it's undertheorized." And it's like, "No, actually this is a quite thorough theory of power." I think what makes it really useful for our book and also useful as a feminist theory is that it takes the personal into account, right? Like you have a lot of ideas about how power works in the world, you can think of Michel Foucault, the panopticon, you can think of like Louis Althusser and global triangle, right? But very few of those theories name how power operates at the level of the individual and the community and the group. And so again, getting back to what we said at the beginning, oftentimes as soon as you see it spelled out for you, you're like, "Oh right, this is how it works." But until you have someone providing you with a conceptual apparatus that sort of shows you how it works, it's sometimes hard to pinpoint from the undifferentiated mass of forces in the world on how things are working. That's what I really value about the matrix of domination in particular.

Sareeta Amrute:

Thank you. That's extremely helpful. I like the way that you're pulling on a theory of power that both gets at all the intersectional frames through which power operates but also works at different scales, as you were saying Lauren. That's really helpful. My second question is related to the slides you showed and I'm just going to hold up my copy of the book. The book is absolutely gorgeous and I'm opening it up to an argument that you were making, that as soon as I read it, I went and showed it to several people in my household. And so I wanted to hear a little bit more from you and I think it relates to the point you were making

at the end about artists, journalists, and activists being at the cutting edge of data science. I wanted for you to talk us through a little bit on your decisions on the aesthetics of the book. It's a very beautiful book. It focuses quite a bit on data visualization and maybe backgrounds a little bit data collection. I was wondering if that was an intentional choice and if you could talk us through some of the thinking behind the focus.

Lauren F. Klein:

It's so perceptive of you to have noticed that because there's sort of the physical book, but then there's also what led us to that point. Catherine and my first work together was on the topic of visualization, and the book actually, or at least the project, started as thinking through the idea of feminist data visualization. Because both of us, in different ways, have prior work in that area. But what happened as we started to draft the proposal for the book, is that we started to realize that you couldn't talk about the end product without the process that leads up to it. And that was what sort of took us back. It was almost sort of a retrospective, like going back to the beginning and saying, "Okay, well how do we get to the point where we end up with these beautiful and compelling and evocative images?" The other thing that I'll say about the book itself, and Catherine mentioned this earlier in the talk, we really see a lot of what we can do is bring together all of this amazing work that is already out there. We don't think we're reinventing the wheel with this book. There is just so much work to draw on. Not by us personally but by so many people. And we wanted to make that work as compelling as it could be to our readers when they encounter the book. So we did, and we sort of deliberately marshalled our resources towards paying for color photos and things like that because we wanted to be able to represent the material as best we could. I don't know Catherine if you have other things to add.

Catherine D'Ignazio:

I think the only thing to add is thinking through how there's been so much work that's going on in a variety of different fields like critical data studies, law and policy, the FAT Conference, and related technical work. Those sort of interrogating questions of power and structural oppression and data. To me, and I think to us, there was little work that was

actually addressing data communication and how questions of structural inequality are also showing up in our objects of communication, whether those are visualizations, or "scrollytelling" things by journalists, or so on. And so it felt really worthy to introduce that aspect as well since often that is the broader public's interface to these questions. It felt really worthy to bring our attention also there, in addition to these other stages of the pipeline.

Sareeta Amrute:

Yeah, that's great. I'm going to show another page from the book that probably won't be too visible to people at home. But my third question relates to something that Catherine and Lauren do at the end, which is they make this very courageous move at the end of the book to do an audit of how the book did according to the principles of data feminist and the book outline. So you can see that product here; it has it has various categories like racism, patriarchy, heteronormativity, and then they outline what their aspiration was and how the book did. I was hoping you could talk to us about what you discovered doing the audit. Do you find that this kind of audit contradicts or problematizes in some way, some of your data feminism principles? For instance, the principle of centering embodiment.

Catherine D'Ignazio:

Thanks for question. Yeah. The background on this is I think this in a way also relates to this hackathon project I was working on where we're really focusing on equity. And as part of that, I want to do a shoutout to Jen Roberts who was part of our leadership team. She actually, for that particular event, helped to design these metrics for inclusion. For who we wanted at the table. And mainly that was around people of color and socioeconomic diversity, and thinking through what makes for an inclusive space. When we started working on this we were talking a lot. We started having conversations about exactly these citational politics, of like, "Who do you cite?" because as you cite somebody, it's a really meaningful thing. You're sort of bringing that person's voice into the room. And yet so much of the work, so much on data science, is very male, very white, very elite. So really thinking through, "How do we give ourselves a challenge on this to consider all these

different dimensions of oppression, and bring into play other voices as we set these metrics?" And we're very explicit. We want to cite 70% people of color, for example. We did that in the first draft. We posted an open source draft back in fall of 2018 and we did an audit of that draft. Then we revised the book. Then we reaudited.

The interesting and somewhat disheartening finding, I think for us, was that in the process of taking in all the comments that we received between the open source version and the final version, and we write about this at the end, we reflected on how we made the book "more academic" because people are like, "Who are you citing? You know, support this thing you said here. Support the thing you said there." And so we did that. We purposely, were not trying to gain the metrics either. We cited the thing and did the thing. We weren't really doing any kind of real time tracking. We reaudit and then we noticed that we did sort of worse on the metrics. So I think for us this was a sort of experiment. I think there are ways to look at it, we got criticized for it too, we had some interesting conversations about it with people because they're saying, "You're reducing intersectionality to just one dimension. Or if you're counting a trans author, you're imagining that they're speaking about that oppression, which they may or may not be." But I think for us it was a challenge to ourselves to try to walk the talk. And then also realize how when you want to walk the talk, we're still failing because of these various kinds of structural forces, one of which I would say is the pressure to create an academic book that cites the people that people are like, "Oh, you have to cite this person." I would say I stand by it as an experiment. I would challenge others to do it and challenge ourselves to do better next time. I don't know how you feel. And Lauren, we haven't really talked about it since we wrote that last statement within the book.

Lauren F. Klein:

Yeah. I think that it was a super interesting process for me. Catherine was bringing this from the Breast Pump hackathon and it seemed interesting to me. In some of my other historical research, I work with a lot of data about people and I think a lot about what it means to sort of reduce lived experience to a check box and we use a spreadsheet. We should mention Issy Carter, our research assistant, actually did all of the work of identifying

each of the people and scholars and projects and people who we cite in the book. They write a little bit about their process in the book as well. I spent a lot of time thinking about that act of reducing life to a data point. Especially the legacies of slavery and colonialism and really the violence that sort of inheres in that act. But I also, and this is something that Catherine and I did talk through, and I think we both believe, you can sort of hold two things in your hands at the same time, right? You can understand how you're not representing the people who were citing and the richness of their work and their lived experience, but it also sort of shows you a way forward. To make this less abstract and more concrete, one of the really interesting things in the revision process, I would say also because we did it furiously. I don't know who read the draft and then read the final version but it's like 50% longer. There weren't footnotes, now there are 600 or something.

Catherine D'Ignazio:

And they're good, read the footnotes.

Lauren F. Klein:

The notes are good. But in the first draft it was really interesting because we did have these metrics and we also came up with our statements of our values, and what we wanted to inhabit as we were writing the book and whose voices we wanted to center. And we initially realized that all of our chapters began with sort of a bad object. It'd be like Edward Tufte, or Francis Galton. Some of these old white men who tend to be cited when talking about data practices. And one of the things that the knowledge of the metrics in the back of our mind made us do, in most but not all cases, is remove the bad object. Why do you need to be oppositional from the start? Why do you need to reify those people by giving them the opening salvo? We can just start with who we want to cite. And can we find someone or can we just think? Usually it was like 30 seconds where you're like, "Oh, I don't need Francis Galton. I need Jessica Marie Johnson who's a [inaudible] scholar, writing about the history of data." You can start your conversation elsewhere. And I think making it explicit whether or not, and again, we didn't have the data on our choices until after we made the book. But having it lurking in the back of our mind, at least for me, made that more of a part of the

routine of the process of writing. It's definitely something that I've thought more about moving forward. I always cite the people I like, but it's made me think more about removing the people I don't like in some of my other work.

Sareeta Amrute:

I think that's such an astute answer because to me what's showing is the queries that you're making of the book as data, actually help you conduct a counter narrative. A counter data around the project, which is wonderful. I'm going to turn now to some of the amazing questions budding up, building up in the Q&A chat. And I'll read some of them out. There's a lovely question that kind of leads right from what we were talking about just now. The question is, "I work in transparency and open data. A collective of women and non-binary folks in this space have founded Open Heroines to raise the issue of gender in our sector. One of the things we often come across is people saying, 'The problem is not with gender. It's with inequalities.' Have you come across this in your work?" And I think the followup is, "How do you counter that kind of an argument?"

Catherine D'Ignazio:

Thanks. I'm super interested to learn more about Open Heroines. I'll go look them up, or if somebody could post the URL in the chat, that'd be great. What I hear in that response is that they're downplaying that gender might count as one of those inequalities, potentially. That to me is already a kind of flag. It's sort of like saying that gender is one of those equalities and there are these other inequalities, and we are concerned about all of them, but we are taking a specifically gendered lens on them. There's increasing precedent and value and recognized worth in taking a "gendered lens" when we look at things like the United Nations sustainable development goals. There's various urban planning researchers right now, like Inés Sánchez de Madariaga, who are advocating for a gender aware approach to basically every kind of city service design, every kind of built environment design. I think there's a broad range of arguments for why gender specifically really matters. I don't know what kind of domain you all work in but I think my response back would be, "Yes, it's about all of those inequalities and we're working specifically on gender,

which is a huge aspect of those inequalities." I don't know Lauren if you have something to add there.

Lauren F. Klein:

No, that was terrific. I guess the only thing that I will add to that, is that what usually ends up in those moves is sort of a displacement of responsibility. And one of the points that we make in the book is whether you're talking about inequality more generally, or gender in particular, it shouldn't always be the responsibility of the people in the minoritized gender position to fix the problem. Right? It actually is more of the problem of the people in the majoritized position to do the work of making the space or the data set or whatever it may be, institution open to people of all genders. And so again to me, and I'm sure that the person who asked the question is familiar with this firsthand, but to really be where the deflection of responsibility, and in the response to not just reclaim gender but also to reclaim collective responsibility, regardless of what your personal gender is.

Sareeta Amrute:

Thank you. We have a question about studies and training. Or a few questions, and I'll try to collate them. "Thank you for your fantastic talk. I'm an undergraduate data science student." Another one is a recent grad with a mathematical background currently working in data science. They both are interested in data feminism and issues of social justice. Do you have any advice for budding data scientists trying to find their place in this world, in this research space? Or as a volunteer?

Catherine D'Ignazio:

Yeah. Thanks so much. So undergraduate, I think Lauren should talk about her lab. Just because I think your department is super interesting. They're creating a really interesting connected space between humanities and technical fields.

Lauren F. Klein:

Yeah. I mean, I will say just sort of in general that it seems to me like right now is a really good time to be having these interests. I think this is one of the things that brought Catherine and me together, I think across territories and walks of life, was that we both had this experience of being interested in both of these issues and not really finding a path in formal academic training in order to do this. My response was to sort of toggle between two. What I thought was sort of a binary choice, one was sort of humanities or computer science. And Catherine left academia for a long time just thinking that wasn't the place for her and then came back later. I think that that's changing in certain places and if you pay enough attention you can see spots like at Emory where I am, we're trying to create this new department. It is a department as of this year that's bringing together formal statistical training with traditional liberal arts disciplines. There's places like Bucknell that have a really interesting Ethical CS department or curriculum. They're trying to reshape what it looks like to study computer science, foregrounding issues of ethics and justice. Which is a long way of saying that I think even though there aren't these like ossified institutions that will guide you in your paths, I think that there are individuals who you can seek out. And by all means, person asking the question, please email me. I think that many academic communities are open to this kind of synthesis in a way that I don't think that they were as much before.

In terms of finding jobs, there's sort of practical, there's places you can look, but maybe Catherine can speak a little bit more to that as well. The last thing I'll say before I'll pass it back to Catherine is that you can always make choices from the place where you sit, or stand, regardless of the context in which you work. Right? And so there are always small choices that you can make, even if external constraints have you in a place where people who control the work that you do are less invested in issues of justice. Obviously it'd be amazing if everyone could choose and find a job that aligned perfectly with their interests but we recognize that's not always the case. But I think it's important to recognize that just because you're not in the perfect situation, it doesn't mean that you can't make little attempts to change things from where you are. I'll pass back to you, Catherine.

Catherine D'Ignazio:

I think the last thing I'll say is, looking at ways to get involved through both work with nonprofits and also in community-based organizations and movements who are working with data. One interesting organization to look at is DataKind, who's starting a kind of service corps for data scientists. That's kind of an interesting model. Kind of like volunteer lawyers where data scientists work with nonprofits and community-based organizations in a kind of structured way. So that's one to look at. And then there's other groups who have sprung up. The one that comes immediately to mind is Data for Black Lives. Or others like Anti-Eviction Mapping Project, for whom data is really central to the movement building work that they do. Of course that's probably, you can't get maybe a full-time job because they're still kind of small efforts, but I think those are ways to connect with folks, to connect with like-minded folks, and to further pursue that pathway. One of the things that I always say, like Lauren I followed a path where I have very disparate interests and for a long time I thought they had to be separate but then I kept believing that they could come together and eventually I was able to find a way. If there are things that you are pursuing in conjunction with each other I think you can find a way to bring them together. And I would encourage you not to give that up even when the world tells you that those two things don't belong together.

Sareeta Amrute:

Okay. I'm going to try to squeeze in at least two more questions from the Q&A. One, building on what we've just been talking about, is "Many development organizations are working with big data corporations like Facebook in order to use their data to paint clearer pictures on issues like femicide, and in turn look towards solutions. What do you think about the ethics behind these collaborations with big tech, especially when interrogating how these corporations gather and sell their data?"

Catherine D'Ignazio:

Yeah, I think it's a huge thing to consider and to look at. Not only because of the corporations participating, but also honestly because of the international development

organizations also participating in those kinds of collaborations. One of the things that we say in Data Feminism, or one of the things that an intersectional feminist lens can bring, is this idea of asking "Who questions?" So like, "Who is doing the work? Who are the data about? Who is being harmed and who is being benefited?" And really thinking about that from a systematic perspective. So when we're talking about elite international development organizations collaborating with elite, super rich corporations, and then gathering data and using that data to make decisions from afar, about people in the global south most likely, without their participation. I don't necessarily see how that works out. I think what we would advocate for from a data feminist perspective is really much more about participation and co-creation. A set of much more participatory methods which I think don't preclude the idea of a big picture. It's thinking about "How do you acquire the big picture and for who is the big picture?" Right? That's my response, I guess. Lauren, what you say?

Lauren F. Klein:

That's a great response. The only thing I think I would add to that, is that even though you're getting a bigger picture, it's not the whole picture. You will never have that. Right? And in the book, we talk about this from a lot of different angles, but we have a chapter on context and we talk about, for instance, if you had Noble's work in Algorithms of Oppression. She talks about Google data, which you know, is some of the biggest data that can be. And yet it still doesn't give you the complete picture of what's going on. So her example is if you search on black girls versus white girls on Google. If you search for white girls, you get wholesome stock photography. And if you search for black girls, you get taken to pornography. And this is coming from people who have these biases that are then clicking on the search results that they want to see, which in turn tell Google's algorithm that those are the search results that more people want to see. And there are a lot of search results. I mean there are a lot of user clicks on Google. I don't know if it's billions or quadrillions or what even number it is, it's huge. And yet, you still can't look at that picture without looking at the larger context in which these data are produced in order to understand what it is that you're looking at.

Sareeta Amrute:

I'm going to ask a question that follows on this point you're making, about for whom is the bigger picture and how was it acquired? The question is, "You've talked a little bit in your discussion about femicides in Mexico, but could you develop questions of privacy in those missing datasets? Minorities can be more easily identifiable. How are those data sets taking this into account?"

Catherine D'Ignazio:

Yeah, we very consciously profiled the position from below, the position from María Salguero's culling of news media reports in the work. I actually don't know of any really comprehensive effort on the part of governments to monitor femicides either. So I guess potentially we could have looked harder for one of those. But no, I think there's potentially huge issues of privacy. Also, especially when you're talking about something like the femicide, there's risk of media reporting that comes out before a person's family has actually been notified. So I think that has actually happened to María Salguero. Where the family ends up getting in touch with her because they're searching for one of their loved ones and then they find them in her database. It's culled from public reports but still she's keeping this comprehensive archive. So yeah, I think the issues there are huge and we can think about those also in relationship to when administration shifts. In the book we named this as the paradox of exposure of data. So it's sort of thinking about like, "When do you want to be counted? When is it a good thing to be counted yourself or to count a population? And when does it actually expose you to harm?" We can think about that in relationship, for example, to the DACA program. Where under the Obama administration, they were like, "Come and register yourself with us, all you young undocumented folks. And you're going to be okay. And we'll give you a path to citizenship." And then currently now, administration shifts. And now there's this great registry of undocumented folks that the administration can then pursue. And so I think these are really hard questions. Especially if you are the person who's collecting and holding that data. Because even though I'm sure under one administration they maybe had good intentions of doing that, maybe, we get into the situation now where that is a huge risk to these young people who have lived in the

United States their whole life. The only thing I point back to is that there is no one size fits all. We actually have a chapter that's about consider context. And I think this is where context makes all of the difference. Context meaning, asking those who questions. Like who's collecting the data? Who is it for? Who's harmed? Who has benefited? And not even just right now, but way far in the future from now. You have to think at different time scales when we're thinking about these things as well.

Sareeta Amrute:

Okay. Maybe one last question and then I'll hand it over to YouTube for any thoughts you want to wrap us up with. Two people in the Q&A want to know if you have any examples of "converting data scientists" to a data feminism point of view. And a related question was, the questioner said, "I understand how the projects you talked about, the art projects and the social justice projects, were data," but they didn't quite understand how they were "science", in the data science.

Lauren F. Klein:

Those are good questions. I feel like we need a 12 step program. I haven't converted anyone, I think. Other than ultimately to show how asking these questions, and thinking about all of these additional questions surrounding the data ultimately leads to better data science. Right? Like if you're working with missing data or data that has missing values you are more aware of in what cases your results can be applied, or are predictive, and in what cases they are not. If there's constraints surrounding the collection environment you are more aware of those constraints by asking these questions first. And so I think to back up and say ideally everyone would become a data feminist because they would see that this ultimately leads to better, more accurate, and more valuable data science work. It's good thinking to have done. In terms of projects not being data science, it's super interesting, the history of science. That's a whole discipline. But you know, what counts as science is incredibly contextual and politicized and it makes sense in a certain time and place. If you've ever read 18th century science, it's totally nutty. People are performing experiments in open air and then inviting spectators and writing down, sort of conjuring the scene, about what

happened. And that was science. I think one of the things that we would sort of push back against and say, "Does this mean that you shouldn't be calculating P values and choosing this regression model versus another?" By all means you should be doing that type of work and I teach my students how to do that. But we also want to be really cautious about policing what counts as science and what doesn't. Because traditionally, again, it's women, it's people of color, it's other minoritized groups, who are told usually after doing the real work of creating that field, that their work is no longer science anymore. That they should be paid less. That they should be replaced by people who have technical credentials. We're running out of time, so I'll get off my soap box. I'll end there. Catherine, take over.

Catherine D'Ignazio:

This is what's great about working with a historian. Everything has a long history and it's important to know that history. For more recent history, I would also point to the fact that data scientists were not data scientists until very recently. It was somewhere in the mid 2000s, data analysts, lowly, number crunching data analysts, get rebranded as data scientists. That was a move to elevate that position, to ally that position with science as a kind of unique validator of all things technical and positivist and so on. And so I think again, sort of questioning where these terms come from, and who gets encompassed by those terms. And then what happens when data analysts get rebranded as data scientists? Lo and behold, all the women get pushed out of the field and the men get paid more. So yeah, I think the idea for having a more inclusive definition is to be paying attention to the innovation that actually is happening, and it's just happening at the margins. We can learn a whole lot from that work.

Sareeta Amrute:

Thank you. We are out of time. I just wanted to say to everyone, thank you so much for joining us tonight. Thank you again to Catherine D'Ignazio and Lauren F. Klein. Their new book, *Data Feminism*, is available digitally and via delivery. We shared links in the chat, as well as hashtags and handles to continue the conversation online. We welcome your

feedback on this event and suggestions for future programming. Check out our website and sign up for the Data & Society events list to stay informed. Thank you. Good night.