Introduction: Why Data Science Needs Feminism

Christine Mann Darden first passed through the gates of NASA's Langley Research Center in Hampton, Virginia, in the summer of 1967. Her newly minted master's degree in applied math had earned her a position as a data analyst there. In the city of Hampton, and across the United States, tensions were running high. In Los Angeles, a massive protest against the Vietnam War ended only when more than a thousand armed police officers attacked the peaceful protestors. One month later, even more violence engulfed the city of Detroit after a police raid spiraled out of control. The 1967 Detroit Riot, or the 1967 Detroit Rebellion, as it is increasingly known, resulted in over forty deaths and one thousand injuries.

The gates of Langley might have shielded Darden from those physical confrontations, but her work there was no less removed from the national stage. By 1967, the space race was well underway, and the United States was losing. The Soviet Union had already sent a man into space and a rocket to the moon. The only thing standing in the way of a Soviet victory was to put those two pieces together. Meanwhile, the United States had suffered a series of defeats—and, in January of that year, an outright disaster, when a sudden fire during a launch test of the *Apollo 1* spacecraft killed all three astronauts on board.

While the nation mourned, everyone at NASA threw themselves back into their work—including Darden, who began her data analyst job in the immediate aftermath of the *Apollo 1* disaster. Two years later, it would be her precise analysis of the physics of rocket reentry that would help to ensure the successful return of the *Apollo 11* mission from the moon, effectively winning the space race for the United States. But it would be Darden herself, as a Black woman with technical expertise, working at a federal agency in which sexism and racism openly prevailed, who demonstrated that the ideological mission of the United States—as a land based on the ideals of liberty, equality, and opportunity for all—was far from accomplished (figure 0.1).

Figure 0.1

Christine Darden in the control room of the Unitary Plan Wind Tunnel at NASA's Langley Research Center in 1975. Courtesy of NASA.

The 1960s, after all, were years of social protest and transformation as well as exploration into outer space. Darden herself had participated in several lunch-counter sit-ins at the Hampton Institute, the historically Black college that she attended for her undergraduate studies.[1] By the time that Darden joined NASA, its Virginia facility had been officially desegregated for several years. But it had yet to reckon with the unofficial segregation of both race and gender that remained in place—particularly among its women employees known as *computers*.

Darden's arrival at Langley coincided with the early days of digital computing. Although Langley could claim one of the most advanced computing systems of the time—an IBM 704, the first computer to support floating-point math—its resources were still limited. For most data analysis tasks, Langley's Advanced Computing Division

relied upon human computers like Darden herself. These computers were all women, trained in math or a related field, and tasked with performing the calculations that determined everything from the best wing shape for an airplane, to the best flight path to the moon. But despite the crucial roles they played in advancing this and other NASA research, they were treated like unskilled temporary workers. They were brought into research groups on a project-by-project basis, often without even being told anything about the source of the data they were asked to analyze. Most of the engineers, who were predominantly men, never even bothered to learn the computers' names.

These women computers have only recently begun to receive credit for their crucial work, thanks to scholars of the history of computing—and to journalists like Margot Lee Shetterly, whose book, *Hidden Figures: The American Dream and the Untold Story of the Black Women Who Helped Win the Space Race*, along with its film adaptation, is responsible for bringing Christine Darden's story into the public eye.[2] Her story, like those of her colleagues, is one of hard work under discriminatory conditions. Each of these women computers was required to advocate for herself—and some, like Darden, chose also to advocate for others. It is because of both her contributions to data science and her advocacy for women that we have chosen to begin our book, *Data Feminism*, with Darden's story. For *feminism* begins with a belief in the "political, social, and economic equality of the sexes," as the *Merriam-Webster Dictionary* defines the term—as does, for the record, Beyoncé.[3] And any definition of feminism also necessarily includes the activist work that is required to turn that belief into reality. In *Data Feminism*, we bring these two aspects of feminism together, demonstrating a way of thinking about data, their analysis, and their display, that is informed by this tradition of feminist activism as well as the legacy of feminist critical thought.

As for Darden, she did not only apply her skills of data analysis to spaceflight trajectories; she also applied them to her own career path. After working at Langley for a number of years, she began to notice two distinct patterns in her workplace: men with math credentials were placed in engineering positions, where they could be promoted through the ranks of the civil service, while women with the same degrees were sent to the computing pools, where they languished until they retired or quit. She did not want to become one of those women, nor did she want others to experience the same fate. So she gathered up her courage and decided to approach the chief of her division to ask him why. As Darden, now seventy-five, told Shetterly in an interview for *Hidden Figures*, his response was sobering: "Well, nobody's ever complained," he told Darden. "The women seem to be happy doing that, so that's just what they do."

In today's world, Darden might have gotten her boss fired—or at least served with an Equal Employment Opportunity Commission complaint. But at the time that

Darden posed her question, stereotypical remarks about "what women do" were par for the course. In fact, challenging assumptions about what women could or couldn't do—especially in the workplace—was the central subject of Betty Friedan's best-selling book, *The Feminine Mystique*. Published in 1963, *The Feminine Mystique* is often credited with starting feminism's so-called second wave.[4] Fed up with the enforced return to domesticity following the end of World War II, and inspired by the national conversation about equality of opportunity prompted by the civil rights movement, women across the United States began to organize around a wide range of issues, including reproductive rights and domestic violence, as well as the workplace inequality and restrictive gender roles that Darden faced at Langley.

That said, Darden's specific experience as a Black woman with a full-time job was quite different than that of a white suburban housewife—the central focus of *The Feminine Mystique*. And when critics rightly called out Friedan for failing to acknowledge the range of experiences of women in the United States (and abroad), it was women like Darden, among many others, whom they had in mind. In *Feminist Theory: From Margin to Center*, another landmark feminist book published in 1984, bell hooks puts it plainly: "[Friedan] did not discuss who would be called in to take care of the children and maintain the home if more women like herself were freed from their house labor and given equal access with white men to the professions. She did not speak of the needs of women without men, without children, without homes. She ignored the existence of all non-white women and poor white women. She did not tell readers whether it was more fulfilling to be a maid, a babysitter, a factory worker, a clerk, or a prostitute than to be a leisure-class housewife."[5]

In other words, Friedan had failed to consider how those additional dimensions of individual and group identity—like race and class, not to mention sexuality, ability, age, religion, and geography, among many others—intersect with each other to determine one's experience in the world. Although this concept—*intersectionality*—did not have a name when hooks described it, the idea that these dimensions cannot be examined in isolation from each other has a much longer intellectual history.[6] Then, as now, key scholars and activists were deeply attuned to how the racism embedded in US culture, coupled with many other forms of oppression, made it impossible to claim a common experience—or a common movement—for all women everywhere. Instead, what was needed was "the development of integrated analysis and practice based upon the fact that the major systems of oppression are interlocking."[7] These words are from the Combahee River Collective Statement, written in 1978 by the famed Black feminist activist group out of Boston. In this book, we draw heavily from intersectionality and other concepts developed through the work of *Black feminist*

scholars and activists because they offer some of the best ways for negotiating this multidimensional terrain.

Indeed, feminism must be intersectional if it seeks to address the challenges of the present moment. We write as two straight, white women based in the United States, with four advanced degrees and five kids between us. We identify as middle-class and *cisgender*—meaning that our gender identity matches the sex that we were assigned at birth. We have experienced sexism in various ways at different points of our lives— being women in tech and academia, birthing and breastfeeding babies, and trying to advocate for ourselves and our bodies in a male-dominated health care system. But we haven't experienced sexism in ways that other women certainly have or that nonbinary people have, for there are many dimensions of our shared identity, as the authors of this book, that align with dominant group positions. This fact makes it impossible for us to speak *from experience* about some oppressive forces—racism, for example. But it doesn't make it impossible for us to educate ourselves and then speak about racism and the role that white people play in upholding it. Or to challenge ableism and the role that abled people play in upholding it. Or to speak about class and wealth inequalities and the role that well-educated, well-off people play in maintaining those. Or to believe in the logic of *co-liberation*. Or to advocate for justice through equity. Indeed, a central aim of this book is to describe a form of intersectional feminism that takes the inequities of the present moment as its starting point and begins its own work by asking: How can we use data to remake the world?[8]

This is a complex and weighty task, and it will necessarily remain unfinished. But its size and scope need not stop us—or you, the readers of this book—from taking additional steps toward justice. Consider Christine Darden, who, after speaking up to her division chief, heard nothing from him but radio silence. But then, two weeks later, she was indeed promoted and transferred to a group focused on sonic boom research. In her new position, Darden was able to begin directing her own research projects and collaborate with colleagues of all genders as a peer. Her self-advocacy serves as a model: a sustained attention to how systems of oppression intersect with each other, informed by the knowledge that comes from direct experience. It offers a guide for challenging power and working toward justice.

What Is Data Feminism?

Christine Darden would go on to conduct groundbreaking research on sonic boom minimization techniques, author more than sixty scientific papers in the field of computational fluid dynamics, and earn her PhD in mechanical engineering—all while

"juggling the duties of Girl Scout mom, Sunday school teacher, trips to music lessons, and homemaker," Shetterly reports. But even as she ascended the professional ranks, she could tell that her scientific accomplishments were still not being recognized as readily as those of her male counterparts; the men, it seemed, received promotions far more quickly.

Darden consulted with Langley's Equal Opportunity Office, where a white woman by the name of Gloria Champine had been compiling a set of statistics about gender and rank. The data confirmed Darden's direct experience: that women and men—even those with identical academic credentials, publication records, and performance reviews—were promoted at vastly different rates. Champine recognized that her data could support Darden in her pursuit of a promotion and, furthermore, that these data could help communicate the systemic nature of the problem at hand. Champine visualized the data in the form of a bar chart, and presented the chart to the director of Darden's division.[9] He was "shocked at the disparity," Shetterly reports, and Darden received the promotion she had long deserved.[10] Darden would advance to the top rank in the federal civil service, the first Black woman at Langley to do so. By the time that she retired from NASA, in 2007, Darden was a director herself.[11]

Although Darden's rise into the leadership ranks at NASA was largely the result of her own knowledge, experience, and grit, her story is one that we can only tell as a result of the past several decades of feminist activism and critical thought. It was a national feminist movement that brought women's issues to the forefront of US cultural politics, and the changes brought about by that movement were vast. They included both the shifting gender roles that pointed Darden in the direction of employment at NASA and the creation of reporting mechanisms like the one that enabled her to continue her professional rise. But Darden's success in the workplace was also, presumably, the result of many unnamed colleagues and friends who may or may not have considered themselves feminists. These were the people who provided her with community and support—and likely a not insignificant number of casserole dinners—as she ascended the government ranks. These types of collective efforts have been made increasingly legible, in turn, because of the feminist scholars and activists whose decades of work have enabled us to recognize that labor—emotional as much as physical—as such today.

As should already be apparent, feminism has been defined and used in many ways. Here and throughout the book, we employ the term *feminism* as a shorthand for the diverse and wide-ranging projects that name and challenge sexism and other forces of oppression, as well as those which seek to create more just, equitable, and livable futures. Because of this broadness, some scholars prefer to use the term *feminisms*,

which clearly signals the range of—and, at times, the incompatibilities among—these various strains of feminist activism and political thought. For reasons of readability, we choose to use the term *feminism* here, but our feminism is intended to be just as expansive. It includes the work of regular folks like Darden and Champine, public intellectuals like Betty Friedan and bell hooks, and organizing groups like the Combahee River Collective, which have taken direct action to achieve the equality of the sexes. It also includes the work of scholars and other cultural critics—like Kimberlé Crenshaw and Margot Lee Shetterly, among many more—who have used writing to explore the social, political, historical, and conceptual reasons behind the inequality of the sexes that we face today.

In the process, these writers and activists have given voice to the many ways in which today's status quo is unjust.[12] These injustices are often the result of historical and contemporary differentials of power, including those among men, women, and nonbinary people, as well as those among white women and Black women, academic researchers and Indigenous communities, and people in the Global North and the Global South. Feminists analyze these power differentials so that they can change them. Such a broad focus—one that incorporates race, class, ability, and more—would have sounded strange to Friedan or to the white women largely credited for leading the fight for women's suffrage in the nineteenth century.[13] But the reality is that women of color have long insisted that any movement for gender equality must also consider the ways in which privilege and oppression are intersectional.

Because the concept of intersectionality is essential for this whole book, let's get a bit more specific. The term was coined by legal theorist Kimberlé Crenshaw in the late 1980s.[14] In law school, Crenshaw had come across the antidiscrimination case of *DeGraffenreid v. General Motors*. Emma DeGraffenreid was a Black working mother who had sought a job at a General Motors factory in her town. She was not hired and sued GM for discrimination. The factory did have a history of hiring Black people: many Black men worked in industrial and maintenance jobs there. They also had a history of hiring women: many white women worked there as secretaries. These two pieces of evidence provided the rationale for the judge to throw out the case. Because the company did hire Black people and did hire women, it could not be discriminating based on race or gender. But, Crenshaw wanted to know, what about discrimination on the basis of race *and* gender together? This was something different, it was real, and it needed to be named. Crenshaw not only named the concept, but would go on to explain and elaborate the idea of intersectionality in award-winning books, papers, and talks.[15]

Key to the idea of *intersectionality* is that it does not only describe the intersecting aspects of any particular person's identity (or *positionalities*, as they are sometimes

termed).[16] It also describes the intersecting forces of privilege and oppression at work in a given society. *Oppression* involves the systematic mistreatment of certain groups of people by other groups. It happens when power is not distributed equally—when one group controls the institutions of law, education, and culture, and uses its power to systematically exclude other groups while giving its own group unfair advantages (or simply maintaining the status quo).[17] In the case of gender oppression, we can point to the sexism, cissexism, and patriarchy that is evident in everything from political representation to the wage gap to who speaks more often (or more loudly) in a meeting.[18] In the case of racial oppression, this takes the form of racism and white supremacy. Other forms of oppression include ableism, colonialism, and classism. Each has its particular history and manifests differently in different cultures and contexts, but all involve a dominant group that accrues power and privilege at the expense of others. Moreover, these forces of power and privilege on the one hand and oppression on the other mesh together in ways that multiply their effects.

The effects of privilege and oppression are not distributed evenly across all individuals and groups, however. For some, they become an obvious and unavoidable part of daily life, particularly for women and people of color and queer people and immigrants: the list goes on. If you are a member of any or all of these (or other) minoritized groups, you experience their effects everywhere, shaping the choices you make (or don't get to make) each day. These systems of power are as real as rain. But forces of oppression can be difficult to detect when you benefit from them (we call this a *privilege hazard* later in the book). And this is where data come in: it was a set of intersecting systems of power and privilege that Darden was intent on exposing when she posed her initial question to her division chief. And it was that same set of intersecting systems of power and privilege that Darden sought to challenge when she approached Champine. Darden herself didn't need any more evidence of the problem she faced; she was already living it every day.[19] But when her experience was recorded as data and aggregated with others' experiences, it could be used to challenge institutional systems of power and have far broader impact than on her career trajectory alone.

In this way, Darden models what we call *data feminism*: a way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought. The starting point for data feminism is something that goes mostly unacknowledged in data science: power is not distributed equally in the world. Those who wield power are disproportionately elite, straight, white, able-bodied, cisgender men from the Global North.[20] The work of data feminism is first to tune into how standard practices in data science serve to reinforce these existing inequalities and second to use data science to challenge and change the

distribution of power.[21] Underlying data feminism is a belief in and commitment to *co-liberation*: the idea that oppressive systems of power harm all of us, that they undermine the quality and validity of our work, and that they hinder us from creating true and lasting social impact with data science.

We wrote this book because we are data scientists and data feminists. Although we speak as a "we" in this book, and share certain identities, experiences, and skills, we have distinct life trajectories and motivations for our work on this project. If we were sitting with you right now, we would each introduce ourselves by answering the question: What brings you here today? Placing ourselves in that scenario, here is what we would have to say.

Catherine: I am a hacker mama. I spent fifteen years as a freelance software developer and experimental artist, now professor, working on projects ranging from serendipitous news-recommendation systems to countercartography to civic data literacy to making breast pumps not suck. I'm here writing this book because, for one, the hype around big data and AI is deafeningly male and white and technoheroic and the time is now to reframe that world with a feminist lens. The second reason I'm here is that my recent experience running a large, equity-focused hackathon taught me just how much people like me—basically, well-meaning liberal white people—are part of the problem in struggling for social justice. This book is one attempt to expose such workings of power, which are inside us as much as outside in the world.[22]

Lauren: I often describe myself as a professional nerd. I worked in software development before going to grad school to study English, with a particular focus on early American literature and culture. (Early means very early—like, the eighteenth century.) As a professor at an engineering school, I now work on research projects that translate this history into contemporary contexts. For instance, I'm writing a book about the history of data visualization, employing machine-learning techniques to analyze abolitionist newspapers, and designing a haptic recreation of a hundred-year-old visualization scheme that looks like a quilt. Through projects like these, I show how the rise of the concept of "data" (which, as it turns out, really took off in the eighteenth century) is closely connected to the rise of our current concepts of gender and race. So one of my reasons for writing this book is to show how the issues of racism and sexism that we see in data science today are by no means new. The other reason is to help translate humanistic thinking into practice and, in so doing, create more opportunities for humanities scholars to engage with activists, organizers, and communities.[23]

We both strongly believe that data can do good in the world. But for it to do so, we must explicitly acknowledge that a key way that power and privilege operate in the world today has to do with the word *data* itself. The word dates to the mid-seventeenth century, when it was introduced to supplement existing terms such as *evidence* and *fact*. Identifying information as data, rather than as either of those other two terms, served a rhetorical purpose.[24] It converted otherwise debatable information into the solid basis for subsequent claims. But what information needs to become data before it can be trusted? Or, more precisely, whose information needs to become data before it can be considered as fact and acted upon?[25] Data feminism must answer these questions, too.

The story that begins with Christine Darden entering the gates of Langley, passes through her sustained efforts to confront the structural oppression she encountered there, and concludes with her impressive array of life achievements, is a story about the power of data. Throughout her career, in ways large and small, Darden used data to make arguments and transform lives. But that's not all. Darden's feel-good biography is just as much a story about the larger systems of power that required data—rather than the belief in her lived experience—to perform that transformative work. An institutional mistrust of Darden's experiential knowledge was almost certainly a factor in Champine's decision to create her bar chart. Champine likely recognized, as did Darden herself, that she would need the bar chart to be believed.

In this way, the alliance between Darden and Champine, and their work together, underscores the flaws and compromises that are inherent in any data-driven project. The process of converting life experience into data always necessarily entails a reduction of that experience—along with the historical and conceptual burdens of the term. That Darden and Champine were able to view their work as a success despite these inherent constraints underscores even more the importance of listening to and learning from people whose lives and voices are behind the numbers. No dataset or analysis or visualization or model or algorithm is the result of one person working alone. Data feminism can help to remind us that before there are data, there are people—people who offer up their experience to be counted and analyzed, people who perform that counting and analysis, people who visualize the data and promote the findings of any particular project, and people who use the product in the end. There are also, always, people who go uncounted—for better or for worse. And there are problems that cannot be represented—or addressed—by data alone. And so data feminism, like justice, must remain both a goal and a process, one that guides our thoughts and our actions as we move forward toward our goal of remaking the world.

Data and Power

It took five state-of-the-art IBM System/360 Model 75 machines to guide the *Apollo 11* astronauts to the moon. Each was the size of a car and cost $3.5 million dollars. Fast forward to the present. We now have computers in the form of phones that fit in our pockets and—in the case of the 2019 Apple iPhone XR—can perform more than 140 million more instructions per second than a standard IBM System/360.[26] That rate of change is astounding; it represents an exponential growth in computing capacity (figure 0.2a). We've witnessed an equally exponential growth in our ability to collect and record information in digital form—and in the ability to have information collected about us (figure 0.2b).
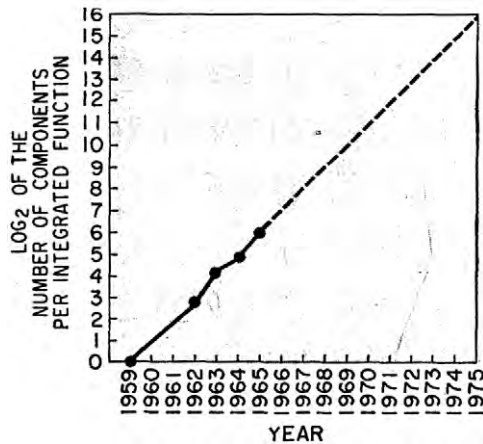
Figure 0.2

(a) The time-series chart included in the original paper on Moore's law, published in 1965, which posited that the number of transistors that could fit on an integrated circuit (and therefore contribute to computing capacity) would double every year. Courtesy of Gordon Moore. (b) Several years ago, researchers concluded that transistors were approaching their smallest size and that Moore's law would not hold. Nevertheless, today's computing power is what enabled Dr. Katie Bouman, a postdoctoral fellow at MIT, to contribute to a project that involved processing and compositing approximately five petabytes of data captured by the Event Horizon Telescope to create the first ever image of a black hole. After the publication of this photo in April 2019 showing her excitement—as one of the scientists on the large team that worked for years to capture the image—Bouman was subsequently trolled and harassed online. Courtesy of Tamy Emma Pepin/ Twitter.

But the act of collecting and recording data about people is not new at all. From the registers of the dead that were published by church officials in the early modern era to the counts of Indigenous populations that appeared in colonial accounts of the Americas, data collection has long been employed as a technique of consolidating knowledge about the people whose data are collected, and therefore consolidating power over their lives.[27] The close relationship between data and power is perhaps most clearly visible in the historical arc that begins with the logs of people captured and placed aboard slave ships, reducing richly lived lives to numbers and names. It passes through the eugenics movement, in the late nineteenth and early twentieth centuries, which sought to employ data to quantify the superiority of white people over all others. It continues today in the proliferation of biometrics technologies that, as sociologist Simone Browne has shown, are disproportionately deployed to surveil Black bodies.[28]

When Edward Snowden, the former US National Security Agency contractor, leaked his cache of classified documents to the press in 2013, he revealed the degree to which the federal government routinely collects data on its citizens—often with minimal regard to legality or ethics.[29] At the municipal level, too, governments are starting to collect data on everything from traffic movement to facial expressions in the interests of making cities "smarter."[30] This often translates to reinscribing traditional urban patterns of power such as segregation, the overpolicing of communities of color, and the rationing of ever-scarcer city services.[31]

But the government is not alone in these data-collection efforts; corporations do it too—with profit as their guide. The words and phrases we search for on Google, the times of day we are most active on Facebook, and the number of items we add to our Amazon carts are all tracked and stored as data—data that are then converted into corporate financial gain. The most trivial of everyday actions—searching for a way around traffic, liking a friend's cat video, or even stepping out of our front doors in the morning—are now hot commodities. This is not because any of these actions are exceptionally interesting (although we do make an exception for Catherine's cats) but because these tiny actions can be combined with other tiny actions to generate targeted advertisements and personalized recommendations—in other words, to give us more things to click on, like, or buy.[32]

This is the data economy, and corporations, often aided by academic researchers, are currently scrambling to see what behaviors—both online and off—remain to be turned into data and then monetized. Nothing is outside of *datafication*, as this process is sometimes termed—not your search history, or Catherine's cats, or the butt that Lauren is currently using to sit in her seat. To wit: Shigeomi Koshimizu, a Tokyo-based professor of engineering, has been designing matrices of sensors that collect data at 360

different positions around a rear end while it is comfortably ensconced in a chair.[33] He proposes that people have unique butt signatures, as unique as their fingerprints. In the future, he suggests, our cars could be outfitted with butt-scanners instead of keys or car alarms to identify the driver.

Although datafication may occasionally verge into the realm of the absurd, it remains a very serious issue. Decisions of civic, economic, and individual importance are already and increasingly being made by automated systems sifting through large amounts of data. For example, PredPol, a so-called predictive policing company founded in 2012 by an anthropology professor at the University of California, Los Angeles, has been employed by the City of Los Angeles for nearly a decade to determine which neighborhoods to patrol more heavily, and which neighborhoods to (mostly) ignore. But because PredPol is based on historical crime data and US policing practices have always disproportionately surveilled and patrolled neighborhoods of color, the predictions of where crime will happen in the future look a lot like the racist practices of the past.[34] These systems create what mathematician and writer Cathy O'Neil, in *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, calls a "pernicious feedback loop," amplifying the effects of racial bias and of the criminalization of poverty that are already endemic to the United States.

O'Neil's solution is to open up the computational systems that produce these racist results. Only by knowing what goes in, she argues, can we understand what comes out. This is a key step in the project of mitigating the effects of biased data. Data feminism additionally requires that we trace those biased data back to their source. PredPol and the "three most objective data points" that it employs certainly amplify existing biases, but they are not the root cause.[35] The cause, rather, is the long history of the criminalization of Blackness in the United States, which produces biased policing practices, which produce biased historical data, which are then used to develop risk models for the future.[36] Tracing these links to historical and ongoing forces of oppression can help us answer the ethical question, Should this system exist?[37] In the case of PredPol, the answer is a resounding no.

Understanding this long and complicated chain reaction is what has motivated Yeshimabeit Milner, along with Boston-based activists, organizers, and mathematicians, to found Data for Black Lives, an organization dedicated to "using data science to create concrete and measurable change in the lives of Black communities."[38] Groups like the Stop LAPD Spying coalition are using explicitly feminist and antiracist methods to quantify and challenge invasive data collection by law enforcement.[39] Data journalists are reverse-engineering algorithms and collecting qualitative data at scale

about maternal harm.[40] Artists are inviting participants to perform ecological maps and using AI for making intergenerational family memoirs (figure 0.3a).[41]

All these projects are *data science*. Many people think of data as numbers alone, but data can also consist of words or stories, colors or sounds, or any type of information that is systematically collected, organized, and analyzed (figures 0.3b, 0.3c).[42] The *science* in data science simply implies a commitment to systematic methods of observation and experiment. Throughout this book, we deliberately place diverse data science examples alongside each other. They come from individuals and small groups, and from across academic, artistic, nonprofit, journalistic, community-based, and for-profit organizations. This is due to our belief in a capacious definition of data science, one that seeks to include rather than exclude and does not erect barriers based on formal credentials, professional affiliation, size of data, complexity of technical methods, or other external markers of expertise. Such markers, after all, have long been used to prevent women from fully engaging in any number of professional fields, even as those fields—which include data science and computer science, among many others—were largely built on the knowledge that women were required to teach themselves.[43] An attempt to push back against this gendered history is foundational to data feminism, too.

Throughout its own history, feminism has consistently had to work to convince the world that it is relevant to people of all genders. We make the same argument: that data feminism is for everybody. (And here we borrow a line from bell hooks.)[44] You will notice that the examples we use are not only about women, nor are they created only by women. That's because *data feminism isn't only about women*. It takes more than one gender to have gender inequality and more than one gender to work toward justice. Likewise, *data feminism isn't only for women*. Men, nonbinary, and genderqueer people are proud to call themselves feminists and use feminist thought in their work. Moreover, *data feminism isn't only about gender*. Intersectional feminists have keyed us into how race, class, sexuality, ability, age, religion, geography, and more are factors that together influence each person's experience and opportunities in the world. Finally, *data feminism is about power—about who has it and who doesn't*. Intersectional feminism examines unequal power. And in our contemporary world, data is power too. Because the power of data is wielded unjustly, it must be challenged and changed.

Data Feminism in Action

Data is a double-edged sword. In a very real sense, data have been used as a weapon by those in power to consolidate their control—over places and things, as well as

Figure 0.3

We define data science expansively in this book—here are three examples. (a) *Not the Only One* by Stephanie Dinkins (2017), is a sculpture that features a Black family through the use of artificial intelligence. The AI is trained and taught by the underrepresented voices of Black and brown individuals in the tech sector. (b) Researcher Margaret Mitchell and colleagues, in "Seeing through the Human Reporting Bias" (2016), have worked on systems to infer what is *not said* in human speech for the purposes of image classification. For example, people say "green bananas" but not "yellow bananas" because yellow is implied as the default color of the banana. Similarly, people say "woman doctor" but do not say "man doctor," so it is the words that are *not spoken* that encode the bias. (c) A gender analysis of Hollywood film dialogue, "Film Dialogue from 2,000 Screenplays Broken Down by Gender and Age," by Hanah Anderson and Matt Daniels, created for The Pudding, a data journalism start-up (2017).

(a) A woman standing next to a **bicycle** with basket.

	Human Label	Visual Label
Bicycle	✓	✓

(b) A city street filled with lots of people walking in the rain.

	Human Label	Visual Label
Bicycle	✗	✓

(c) A **yellow** Vespa parked in a lot with other cars.

	Human Label	Visual Label
Yellow	✓	✓

(d) A store display that has a lot of bananas on sale.

	Human Label	Visual Label
Yellow	✗	✓

Figure 0.3 (continued)
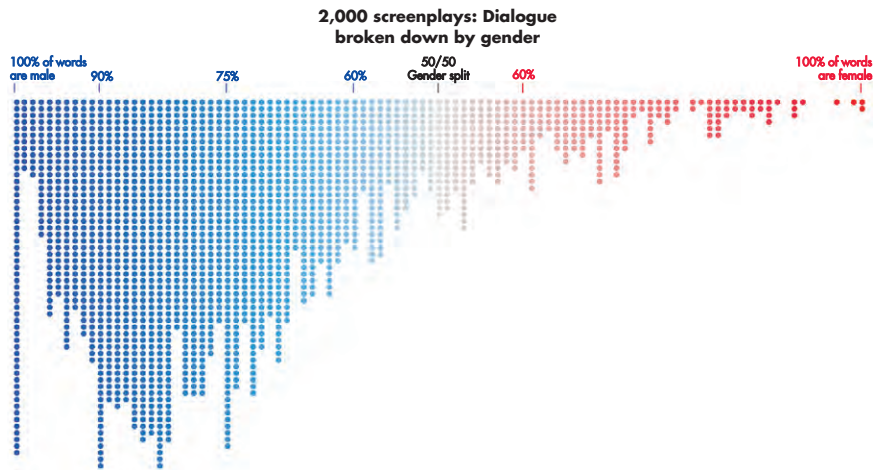
Figure 0.3 (continued)

people. Indeed, a central goal of this book is to show how governments and corporations have long employed data and statistics as management techniques to preserve an unequal status quo. Working with data from a feminist perspective requires knowing and acknowledging this history. To frame the trouble with data in another way: it's not a coincidence that the institution that employed Christine Darden and enabled her professional rise is the same that wielded the results of her data analysis to assert the technological superiority of the United States over its communist adversaries and to plant an American flag on the moon. But this flawed history does not mean ceding control of the future to the powers of the past. Data are part of the problem, to be sure. But they are also part of the solution. Another central goal of this book is to show how the power of data can be wielded back.

To guide us in this work, we have developed seven core principles. Individually and together, these principles emerge from the foundation of intersectional feminist thought. Each of the following chapters is structured around a single principle. The seven principles of data feminism are as follows:

1. **Examine power.** Data feminism begins by analyzing how power operates in the world.
2. **Challenge power.** Data feminism commits to challenging unequal power structures and working toward justice.

3. **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.

4. **Rethink binaries and hierarchies.** Data feminism requires us to challenge the gender binary, along with other systems of counting and classification that perpetuate oppression.

5. **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.

6. **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.

7. **Make labor visible.** The work of data science, like all work in the world, is the work of many hands. Data feminism makes this labor visible so that it can be recognized and valued.

Each of the following chapters takes up one of these principles, drawing upon examples from the field of data science, expansively defined, to show how that principle can be put into action. Along the way, we introduce key feminist concepts like the matrix of domination (Patricia Hill Collins; see chapter 1), situated knowledge (Donna Haraway; see chapter 3), and emotional labor (Arlie Hochschild; see chapter 8), as well as some of our own ideas about what data feminism looks like in theory and practice. To this end, we introduce you to people at the cutting edge of data and justice. These include engineers and software developers, activists and community organizers, data journalists, artists, and scholars. This range of people, and the range of projects they have helped to create, is our way of answering the question: What makes a project feminist? As will become clear, a project may be feminist in *content*, in that it challenges power by choice of subject matter; in *form*, in that it challenges power by shifting the aesthetic and/or sensory registers of data communication; and/or in *process*, in that it challenges power by building participatory, inclusive processes of knowledge production. What unites this broad scope of data-based work is a commitment to action and a desire to remake the world.

Our overarching goal is to take a stand against the status quo—against a world that benefits us, two white college professors, at the expense of others. To work toward this goal, we have chosen to feature the voices of those who speak from the margins, whether because of their gender, sexuality, race, ability, class, geographic location, or any combination of those (and other) subject positions. We have done so, moreover,

because of our belief that those with direct experience of inequality know better than we do about what actions to take next. For this reason, we have attempted to prioritize the work of people in closer proximity to issues of inequality over those who study inequality from a distance. In this book, we pay particular attention to inequalities at the intersection of gender and race. This reflects our location in the United States, where the most entrenched issues of inequality have racism at their source. Our values statement, included as an appendix to this book, discusses the rationale for these authorial choices in more detail.

Any book involves making choices about whose voices and whose work to include and whose voices and work to omit. We ask that those who find their perspectives insufficiently addressed or their work insufficiently acknowledged view these gaps as additional openings for conversation. Our sincere hope is to contribute in a small way to a much larger conversation, one that began long before we embarked upon this writing process and that will continue long after these pages are through.

This book is intended to provide concrete steps to action for data scientists seeking to learn how feminism can help them work toward justice, and for feminists seeking to learn how their own work can carry over to the growing field of data science. It is also addressed to professionals in all fields in which data-driven decisions are being made, as well as to communities that want to resist or mobilize the data that surrounds them. It is written for everyone who seeks to better understand the charts and statistics that they encounter in their day-to-day lives, and for everyone who seeks to communicate the significance of such charts and statistics to others.

Our claim, once again, is that data feminism is for everyone. It's for people of all genders. It's by people of all genders. And most importantly: it's about much more than gender. Data feminism is about power, about who has it and who doesn't, and about how those differentials of power can be challenged and changed using data. We invite you, the readers of this book, to join us on this journey toward justice and toward remaking our data-driven world.

Data Feminism

By: Catherine D'Ignazio, Lauren F. Klein

Citation:

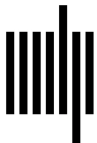*Data Feminism*
By: Catherine D'Ignazio, Lauren F. Klein
DOI: 10.7551/mitpress/11805.001.0001
ISBN (electronic): 9780262358521
Publisher: The MIT Press
Published: 2023

The MIT Press