

7 Show Your Work

Principle: 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.

If you work in software development, chances are that you have a GitHub account. As of June 2018, the online code-management platform had over twenty-eight million users worldwide. By allowing users to create web-based repositories of source code (among other forms of content) to which project teams of any size can then contribute, GitHub makes collaborating on a single piece of software or a website or even a book much easier than it has ever been before.

Well, easier if you're a man. A 2016 study found that female GitHub users were less likely to have their contributions accepted if they identified themselves in their user profiles as women. (The study did not consider nonbinary genders.)¹ Critics of GitHub's commitment to inclusivity, or the lack thereof, also point to the company's internal politics. In 2014, GitHub's cofounder was forced to resign after allegations of sexual harassment were brought to light.² More recently, in 2018, Agnes Pak, a former top attorney at GitHub, sued the company for allegedly altering her performance reviews after she complained about her gender and race contributing to a lower compensation package, giving them the grounds to fire her.³ Pak's suit came only shortly after transgender software developer Coraline Ada Ehmke, in 2017, declined a significant severance package so that she could talk publicly about her negative experience of working at GitHub.⁴ Clearly, GitHub has several major issues of corporate culture that it must address.

But a corporate culture that is hostile to women does not necessarily preclude other feminist interventions. And here GitHub makes one important one: its platform helps *show the work* of writing collaborative code. In addition to basic project management tools, like bug tracking and feature requests, the GitHub platform also generates visualizations of each team member's contributions to a project's codebase. Area charts, arranged in small multiples, allow viewers to compare the quantity, frequency, and duration of any particular member's contributions (figure 7.1a). A line graph reveals patterns in the day of the week when those contributions took place (figure 7.1b). And a flowchart-like diagram of the relationships between various branches of the project's

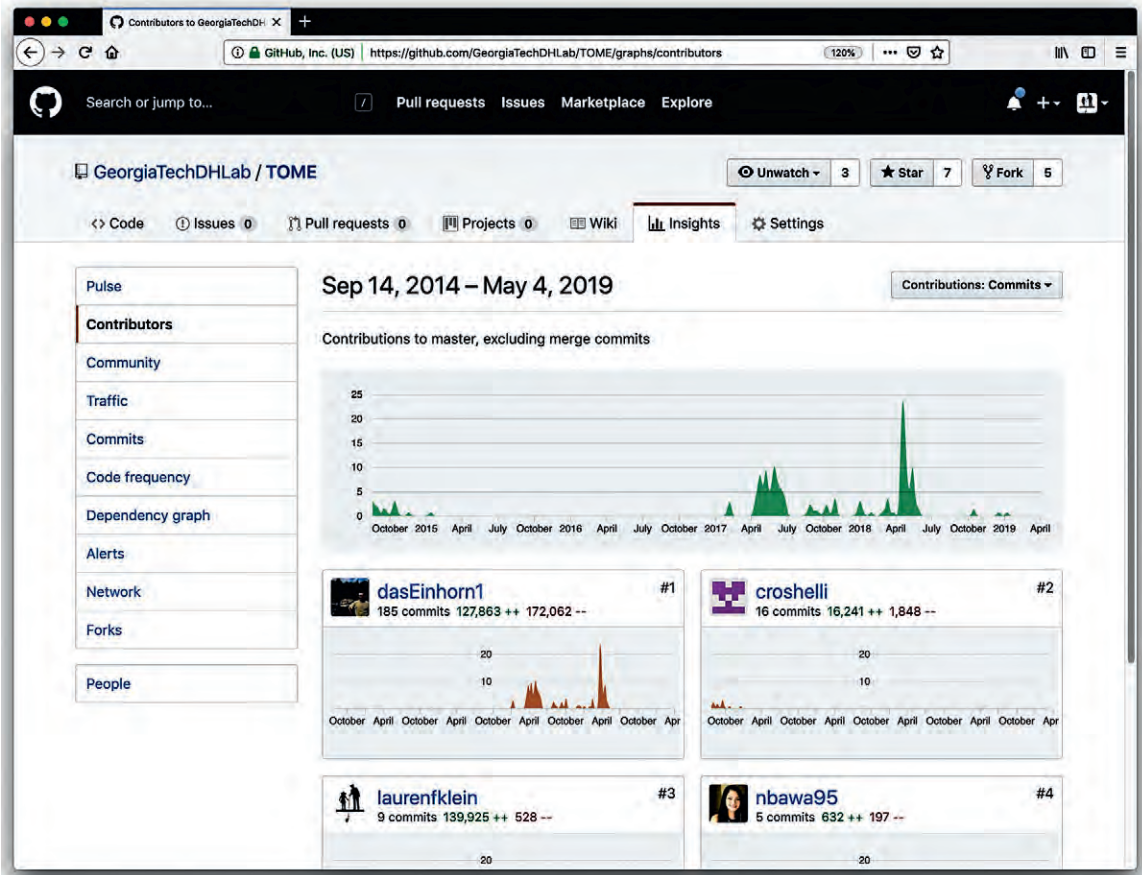


Figure 7.1

(a) The first of three visualizations of the code associated with a project from Lauren's research group, showing the significant contributions of student researchers between the years 2014 and 2019. (b) A bar chart shows the frequency of code commits over time, and a line graph shows any patterns in the day of the week when the commits were made. Screenshot by Lauren F. Klein. (c) A flowchart-like diagram documents the relationships between the various branches of the project's codebase. Screenshots by Lauren F. Klein.

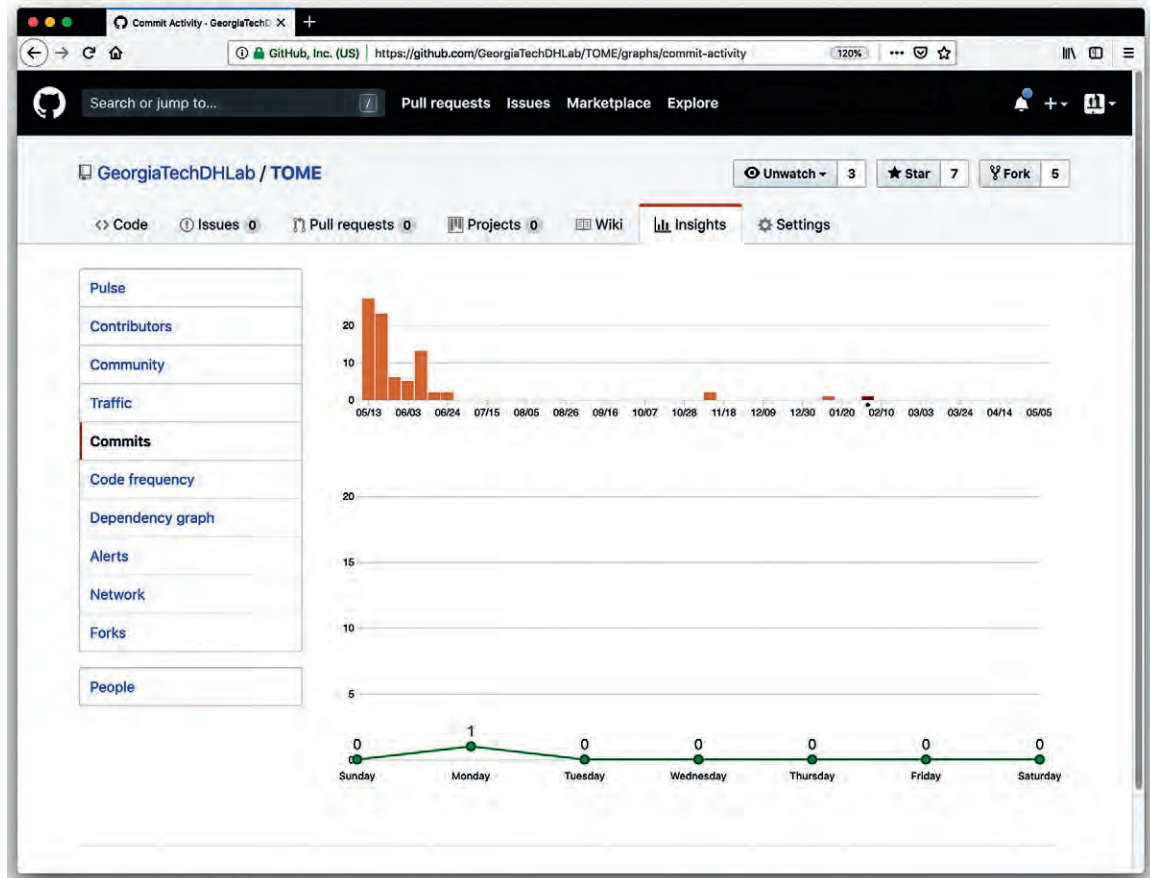


Figure 7.1 (continued)

code helps to acknowledge any sources for the project that might otherwise go uncredited, as well as any additional projects that might build upon the project's initial work (figure 7.1c).

Coding is work, as anyone who's ever programmed anything knows well. But it's not always work that is easy to see. The same is true for collecting, analyzing, and visualizing data. We tend to marvel at the scale and complexity of an interactive visualization like the *Ship Map*, in figure 7.2, which plots the paths of the global merchant fleet over the course of the 2012 calendar year.⁵ By showing every single sea voyage, the *Ship Map* exposes the networks of waterways that constitute our global product supply chain. But we are less often exposed to the networks of processes and people that

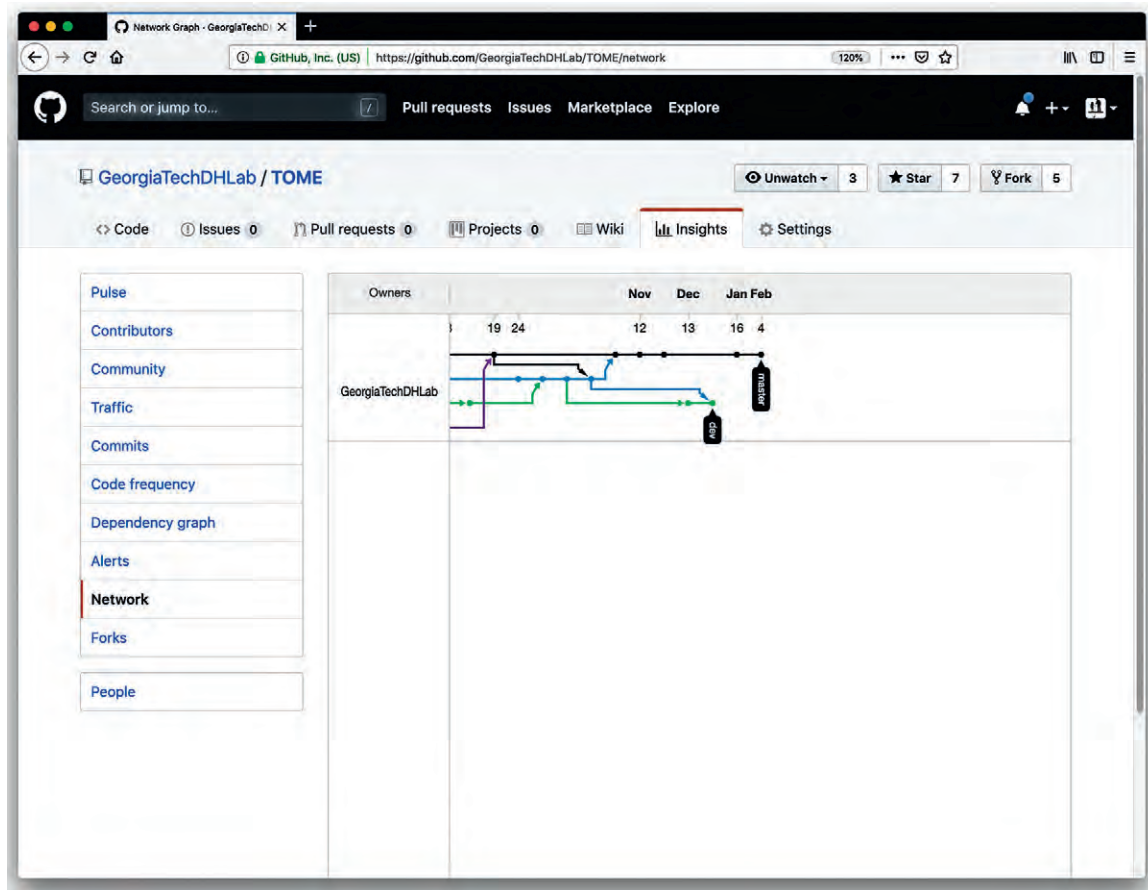


Figure 7.1 (continued)

help constitute the visualization itself—from the seventy-five corporate researchers at Clarksons Research UK who assembled and validated the underlying dataset, to the academic research team at University College London’s Energy Institute that developed the data model, to the design team at Kiln that transformed the data model into the visualization that we see. And that is to say nothing of the tens of thousands of commercial ships that served as the source of data in the first place. Visualizations like the *Ship Map* involve the work of many hands.

Unfortunately, however, when releasing a data product to the public, we tend not to credit the many hands who perform this work. We often cite the source of the dataset, and the names of the people who designed and implemented the code and graphic elements. But we rarely dig deeper to discover who created the data in first

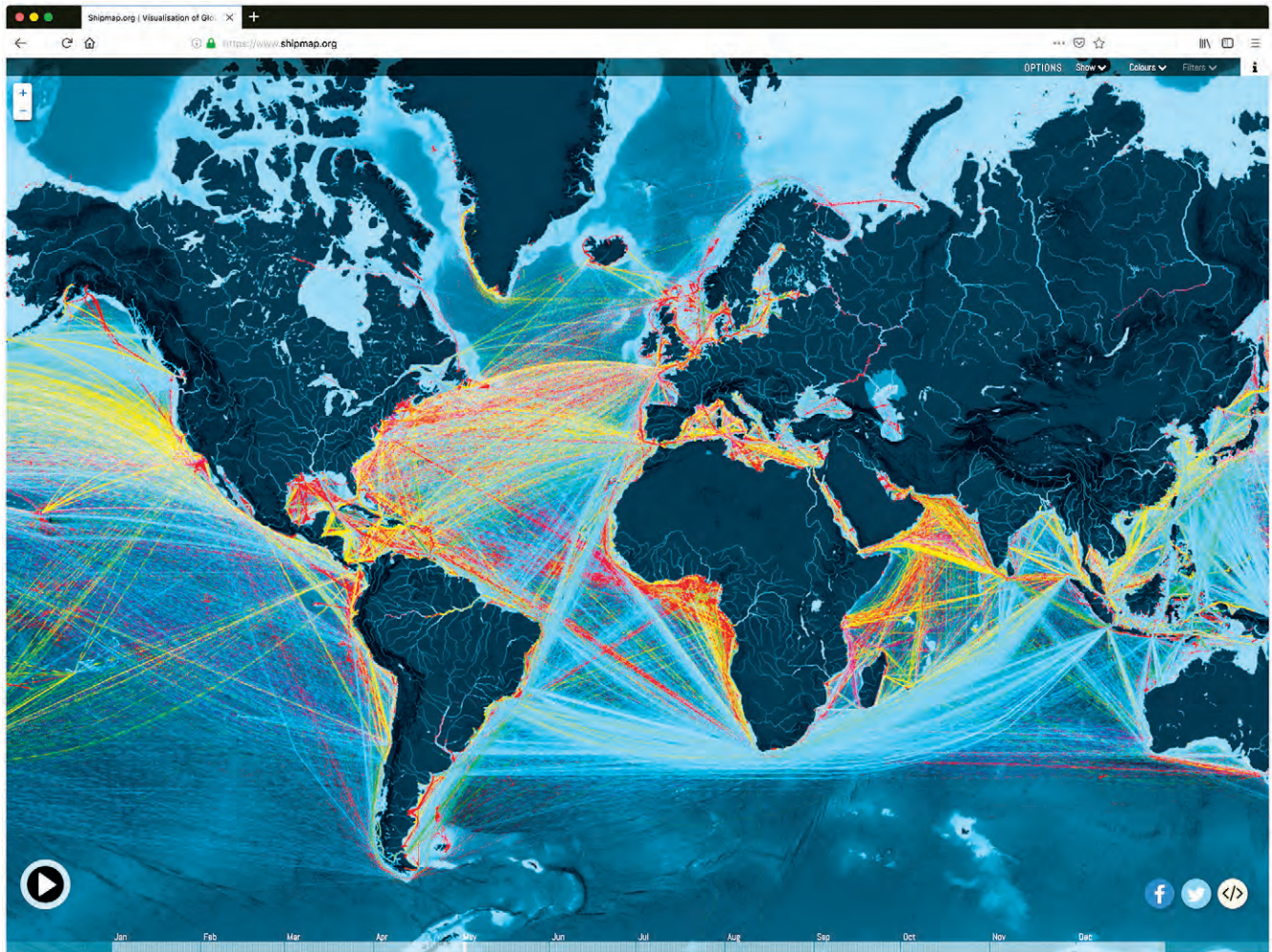


Figure 7.2

A time-based visualization of global shipping routes, designed by Kiln in 2016, based on data from the University College London Energy Institute (UCL EI). The Ship Map website was created by Duncan Clark and Robin Houston from Kiln, and the dataset was compiled by Julia Schaumeier and Tristan Smith from the UCL EI. The website also includes a soundtrack: Bach's Goldberg Variations, played by Kimiko Ishizaka.

place, who collected the data and processed them for use, and who else might have labored to make creations like the *Ship Map* possible. Admittedly, this information is sometimes hard to find. And when project teams (or individuals) are already operating at full capacity, or under budgetary strain, this information can—ironically—simply be too much additional work to pursue.⁶ Even in cases in which there are both resources and desire, information about the range of the contributors to any particular project sometimes can't be found at all. But the various difficulties we encounter when trying to acknowledge this work reflects a larger problem in what information studies scholar Miriam Posner calls our *data supply chain*.⁷ Like the contents of the ships visualized on the *Ship Map*, about which we only know sparse details—the map can tell us *if* a shipping container was loaded onto the boat, but not *what* the shipping container contains—the invisible labor involved in data work, as Posner argues, is something that corporations have an interest in keeping out of public view.

To put it more simply, it's not a coincidence that much of the work that goes into designing a data product—visualization, algorithm, model, app—remains invisible and uncredited. In our capitalist society, we tend to value work that we can see. This is the result of a system in which the cultural worth of any particular form of work is directly connected to the price we pay for it; because a service costs money, we recognize its larger value. But more often than not, the reverse also holds true: we fail to recognize the larger value of the services we get for free. When, in the early 1970s, the International Feminist Collective launched the Wages for Housework campaign, it was this phenomenon of *invisible labor*—labor that was unpaid and therefore unvalued—that the group was trying to expose (figure 7.3).⁸ The precise term they used to describe this work was *reproductive labor*, which comes from the classical economic distinction between the paid and therefore economically *productive* labor of the marketplace, and the unpaid and therefore economically *unproductive* labor of everything else. By reframing this latter category of work as reproductive labor, rather than simply (and inaccurately) unproductive labor, groups like the International Feminist Collective sought to emphasize how the range of tasks that the term encompassed, like cooking and cleaning and child-rearing, were precisely the tasks that enabled those who performed “productive” labor, like office or factory work, to continue to do so.

The Wages for Housework movement began in Italy and migrated to the United States with the help of labor organizer and theorist Silvia Federici. It eventually claimed chapters in several American cities, and did important consciousness-raising work.⁹ Still, as prominent feminists like Angela Davis pointed out, while housework might have been unpaid for white women, women of color—especially Black women in the United States—had long been paid, albeit not well, for their housework in other people's homes: “Because of the added intrusion of racism, vast numbers of Black women



Figure 7.3

A Wages for Housework march, 1977. Photograph by Bettye Lane. Courtesy of the Schlesinger Library, Radcliffe Institute/Bettye Lane.

have had to do their own housekeeping and other women's home chores as well."¹⁰ Here, Davis is making an important point about *racialized labor*: just as housework is structured along the lines of gender, it is also structured along the lines of race and class. The domestic labor of women of color was and remains *underwaged labor*, as feminist labor theorists would call it, and its low cost was what permitted (and continues to permit) many white middle- and upper-class women to participate in the more lucrative *waged labor* market instead.¹¹

Since the 1970s, the term *invisible labor* has come to encompass the various forms of labor, unwaged, underwaged, and even waged, that are rendered invisible because they take place inside of the home, because they take place out of sight, or because they lack physical form altogether.¹² Visit WagesforFacebook.com and you'll find a version of the Wages for Housework argument updated for a new form of invisible work. This invisible labor can be found all over the web, as digital labor theorists such as Tiziana Terranova have helped us to understand.¹³ "They call it sharing. We call it stealing," is one of the statements that scrolls down the screen in large black type. The word *it* refers to work that most of us perform every day, in the form of our Facebook likes, Instagram

posts, and Twitter tweets. The point made by Laurel Ptak, the artist behind Wages for Facebook—a point also made by Terranova—is that the invisible unpaid labor of our likes and tweets is precisely what enables the Facebooks and Twitters of the world to profit and thrive.

The Invisible Labor of Data Science

The world of data science is able to profit and thrive because of unpaid invisible labor as well. How did Netflix improve its movie recommendation algorithm? The company crowdsourced it.¹⁴ How did the *Guardian*, the British newspaper, determine which among two million leaked documents might contain incriminating information about government misspending? The paper crowdsourced it.¹⁵ The optical character recognition (OCR) error correction performed on the dataset of early modern books that you downloaded for your text-analysis project? That was crowdsourced, too.¹⁶

Each of these crowdsourcing projects were framed as acts of benevolence (and, in the case of Netflix, an opportunity to win a million-dollar prize). People should want to contribute to these projects, their proponents claimed, since their labor would further the public good.¹⁷ However, Ashe Dryden, the software developer and diversity consultant, points out that people can only help crowdsource if they have the inclination and the time.¹⁸ Think back to that study of GitHub. If you were a woman and you knew your contributions to a programming project were less likely to be accepted than if you were a man, would that motivate you to contribute the project? Or, for another example, consider Wikipedia. Although the exact gender demographics of Wikipedia contributors are unknown, numerous surveys have indicated that those who contribute content to the crowdsourced encyclopedia are between 84 percent and 91.5 percent men.¹⁹ Why? It could be that there, too, edits are less likely to be accepted if they come from women editors.²⁰ It could also be attributed to Wikipedia's exclusionary editing culture and technological infrastructure, as science and technology studies (STS) scholars Heather Ford and Judy Wajcman have argued.²¹ And there is also reason to go back to the housework argument. Dryden cites a 2011 study showing that women in twenty-nine countries spend more than twice as much time on household tasks than men do, even when controlling for women who hold full-time jobs.²² The study did not consider nonbinary genders or same-sex (or other non-hetero-typical) households. But even as a rough estimate, it seems that women simply don't have as much time.²³

In capitalist societies, it's very often the case that time is money. But it's also important to remember to ask whose time is being spent and whose money is being saved. The premise behind Amazon's Mechanical Turk—or MTurk, as the crowdsourcing

platform is more commonly known—is that data scientists want to save their own time and their own bottom line.²⁴ The MTurk website touts its access to a global marketplace of “on-demand Workers,” who are advertised as being more “scalable and cost-effective” than the “time consuming [and] expensive” process of hiring actual employees.²⁵ But the data-entry and data-processing tasks performed by these workers earn them less than minimum wage, even as a recent study by the Pew Research Center showed that 51 percent of US-based Turkers, as they are known, hold college degrees, and 88 percent are below the age of fifty, among other metrics that would otherwise rank them among the most desired demographic for salaried employees.²⁶ This form of underwaged work is also increasingly outsourced from the United States to countries with fewer (or worse) labor laws and fewer (or worse) opportunities for economic advancement. A 2010 University of California, Irvine study measured a 20 percent drop in the number of US-based Turkers over the eighteen months that it monitored.²⁷ This trend has continued, the real-time MTurk tracker shows. (The gender split, interestingly, has evened out over time.)

Even at resource-rich companies like Amazon and Google, the work of data entry is profoundly undervalued in proportion to the knowledge it helps to create. Andrew Norman Wilson’s 2011 documentary *Workers Leaving the Googleplex* (figure 7.4) exposes how the workers tasked with scanning the books for the Google Books database are hired as a separate but unequal class of employee, with ID cards that restrict their access to most of the Google campus and that prevent them from enjoying the company’s famed employee perks.²⁸ (Evidently, working overtime to preserve the world’s cultural heritage still does not entitle you to a free lunch, let alone a free class on how to cook Pad Kee Mao.)²⁹

Wilson also observes that Google’s book-scanning workers are disproportionately women and people of color—a fact that would not surprise the long line of women of color scholar-activists, including Angela Davis, Patricia Hill Collins, and Evelyn Nakano Glenn, who have insisted that economic oppression be recognized as a vector that cuts across the matrix of domination as a whole. Information studies scholar Lilly Irani confirms that “today’s hierarchy of data labor echoes older gendered, classed, and raced technology hierarchies.”³⁰ Here, Irani compares the hierarchy of data labor to the hierarchy encountered by the first generation of female computers, like Christine Darden, whom we discussed in this book’s introduction.³¹ But Irani’s own research also considers contemporary digital labor practices, and in particular, Amazon’s Mechanical Turk, the people it employs, and the people it exploits. In 2008, Irani and collaborators built a web tool called the Turkopticon, which enabled Turkers to anonymously report unfair labor conditions, as well as any additional information that might help

**Figure 7.4**

Andrew Norman Wilson's *Workers Leaving the Googleplex* (2011) documents the hidden inequities at Google's Mountain View headquarters. Still courtesy of Andrew Norman Wilson.

them decide whether to accept future tasks.³² Irani envisioned the Turkopticon as a worker-led project. However, the same unfair labor conditions that necessitated the tool also ultimately limited its reach. In 2018, after ten years of service, its all-volunteer team of moderators called it quits. “We’re all burned out,” they wrote on Twitter.³³ And amid tagging images and correcting error-laden text, no additional Turkers could find the time.

The people who perform this *cultural data work*, as Irani terms it, are not only found on the MTurk platform, however. They’re also increasingly the people on whom the entire information economy depends. Cultural data workers are responsible for the invisible labor involved in moderating the veritable deluge of content produced online every day, ensuring that your Facebook feed is free of, for example, child pornography and violent propaganda videos. When a 2014 exposé in *Wired* magazine documented

the emotional costs of this labor, performed by some of the least empowered of these workers—women in the Global South—it was met with an outpouring of shock and outrage.³⁴ But subsequent studies like *Ghost Work*, by anthropologist Mary Gray and computer scientist Siddharth Suri, have documented the existence of a large “global underclass” performing this work of content moderation, transcription, and captioning.³⁵ They make the point that the so-called automation of artificial intelligence relies on a vast number of human beings in the loop.³⁶ Moreover, while the demographics of Silicon Valley tech workers remain steadily young, white, and male, these global “ghost workers” are often older women of color, and always required to accept precarious labor conditions.

Those who study the human costs of global capitalism would be quick to point out that this exploitation of precarious, racialized, colonial labor has a long history, one that has its roots in the original form of human exploitation: slavery. Slavery and capitalism are closely connected, after all, and one infamous story is often told to illustrate this point: in 1781, the British slave ship *Zong* made a series of navigational errors while crossing the Atlantic, resulting in a shortage of drinking water for the seventeen crew members and 133 captives on board.³⁷ After performing a cost-benefit analysis, the captain decided to throw the crew’s enslaved human “cargo” overboard so that the crew members could consume all of the remaining water and rations themselves. The decision was made because the captain calculated that he could collect enough insurance money on his captives’ loss of life to come out ahead, even if he couldn’t sell them once they landed ashore. He was thinking about human lives solely in terms of their market value—the notion that capitalism holds in highest regard.

The stark inhumanity of this calculation has prompted numerous scholars and artists to return to the *Zong* as they reckon with what Christina Sharpe calls, in the language of the ship, the “wake” of slavery.³⁸ Poet M. NourbeSe Philip, for example, composed a book-length poem, *Zong!*, using only the words of the legal case that serves as the sole documentation of the original event.³⁹ Written over the course of many years and published in 2011, Philip’s poem plucks words and short phrases out of the language of the court case, arranging them across the printed page. Philip’s poem regularly shifts tenses from past to present, and from present to past, lending an additional voice to Sharpe’s claim that the effects of that originary crime—the exploitation of Black bodies for white financial gain—are far from resolved.⁴⁰

Our present technological infrastructure follows this same pattern of exploitation. In the United States, the scarcely paid or altogether unpaid labor of those who endure a contemporary form of enslavement—incarceration—has been used for everything from packaging Windows software to cleaning up the 2010 BP oil spill.⁴¹ In a global

colonial context, we might consider how the cobalt required to produce the lithium-ion batteries that power our cell phones and laptops is associated with significant human rights violations, including coercing labor from Congolese children as young as seven.⁴² The unregulated disposal of electronics has resulted in roadside salvage and repair shops in places like Agbogbloshie, just outside of Accra, Ghana, which have long served as sites of invention and ingenuity, being transformed into toxic “e-waste” sites, with profound consequences for the health of those who live and work there, as well as for the environment.⁴³ The humanitarian and ecological stakes of our attachments to data and technology cannot be higher, nor can their source be any more clear: the capitalist and colonial forces that encourage the exploitation of Black and brown bodies so that white bodies can thrive.⁴⁴

Examining Data Production

The forces of global capitalism can feel overwhelming. And as people who use data and technology in our everyday work, we are each complicit to varying degrees. But there are certain small things we can do in our work, and in our work with others, to push back against this weight. In prior chapters, we have described some of these possibilities: incorporating an examination of power into a data analysis project (chapters 1 and 2); pushing back against false binaries and hierarchies (chapter 4); including multiple and marginalized voices in the design process (chapter 5); and contextualizing data so that they are not imagined to “speak for themselves” (chapter 6).

Along with these starting points, we can also begin to carve out additional space for the scholars, journalists, and other researchers who are explicitly studying the labor of data science—those who are examining and challenging power by tracing visualizations and algorithms and bots back to their human and material sources. This growing area of research might be called *data production studies*, borrowing a rubric from the field of production studies that currently sits at the intersection of film and media studies and labor studies. The primary focus of production studies, as it relates to film and media, is how media artifacts are produced. Media studies scholar Miranda Banks has asserted that “production studies is a feminist methodology” because it pays particular attention to the power differentials involved in the media production process, as well as the material conditions of media workers.⁴⁵ Work focused on data production is already happening in fields like STS, the digital humanities, library and information science, and archival studies, among others.⁴⁶ It looks at the production process of datasets, algorithms, and models, and traces those products back to the people and conditions that enabled their creation.

As an example of work in this emerging area, we might consider “Anatomy of an AI System,” a project by technology researcher Kate Crawford and design scholar Vladan Joler that seeks to describe and diagram the human labor, data dependencies, and material resources that contribute to a single Amazon Echo. The project was published online as a diagram of Borgesian proportions, too big to view in its entirety on a standard laptop screen (figure 7.5a); it was accompanied by a nine-thousand-word essay. Viewers are first introduced to the mineral extraction required to produce the electronics components for the device and made aware of the hard labor (and sometimes child labor) this task requires. The chart (and narrative) proceeds through processes of refining, assembling, and distributing these components, then transporting them physically, then transporting them virtually—through the infrastructure of the internet. Once within the Amazon corporate boundary, the chart depicts the layers of workers who provide everything from network maintenance to training datasets (figure 7.5b). Crawford and Joler also diagram patterns in the organization of Amazon’s labor force, which they describe in terms of “fractal chains of production and exploitation.” But what is required for this replication is people: “At every level contemporary technology is deeply rooted in and running on the exploitation of human bodies,” the essay concludes.⁴⁷

“Anatomy of an AI System” is an investigation and exposé of the invisible labor involved in making a single product on a global scale. In this way, it is an ambitious example of the seventh principle of data feminism: *show the work*. Behind the magic and marketing of data products, there is always hidden labor—often performed by women and people of color, which is both a cause and effect of the fact that this labor is both underwaged and undervalued. Data feminism seeks to make this labor visible so that it can be acknowledged and appropriately valued, and so that its truer cost—for people and for the planet—can be recognized.

Crediting Data Work

The emphasis on giving formal credit for a broad range of work derives from feminist practices of citation. Feminist theorist Sara Ahmed describes this practice as a way of resisting how certain types of people—usually cis and white and male—“take up spaces by screening out others.”⁴⁸ When those other people are screened out, they become

Figure 7.5 (following three pages)

Overview (a) and detail (b) of “Anatomy of an AI System” (2018)—a diagram and essay by Kate Crawford and Vladan Joler that attempts to chart all of the human labor, data, and planetary resources used to create an Amazon Echo device. Courtesy of Kate Crawford and Vladan Joler.

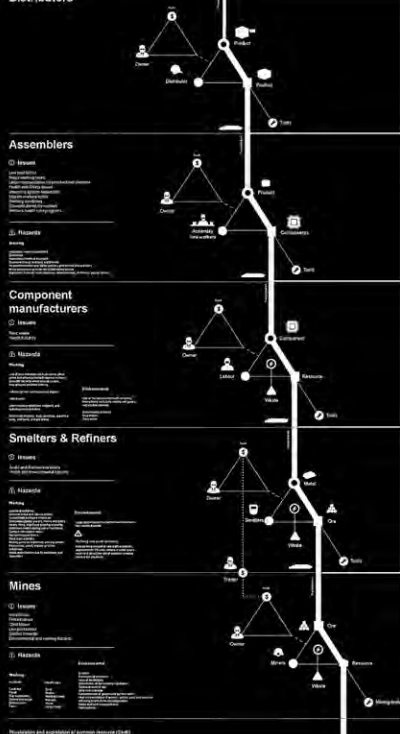
Anatomy of an AI system

An anatomical case study of the Amazon echo as a artificial intelligence system made of human labor

Income distribution



Distributors



Elements



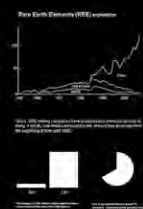
Geological process



Amazon storage systems



Transportation



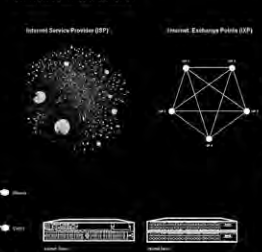
Human operator



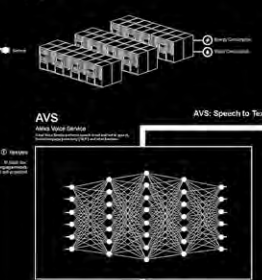
Domestic infrastructure



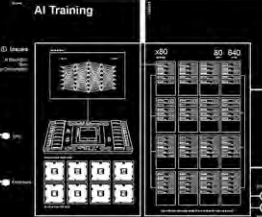
Internet infrastructure



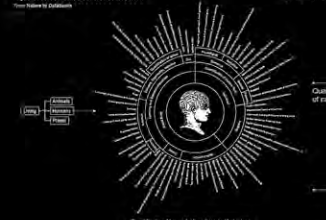
Internet Platforms & Services



AI Training

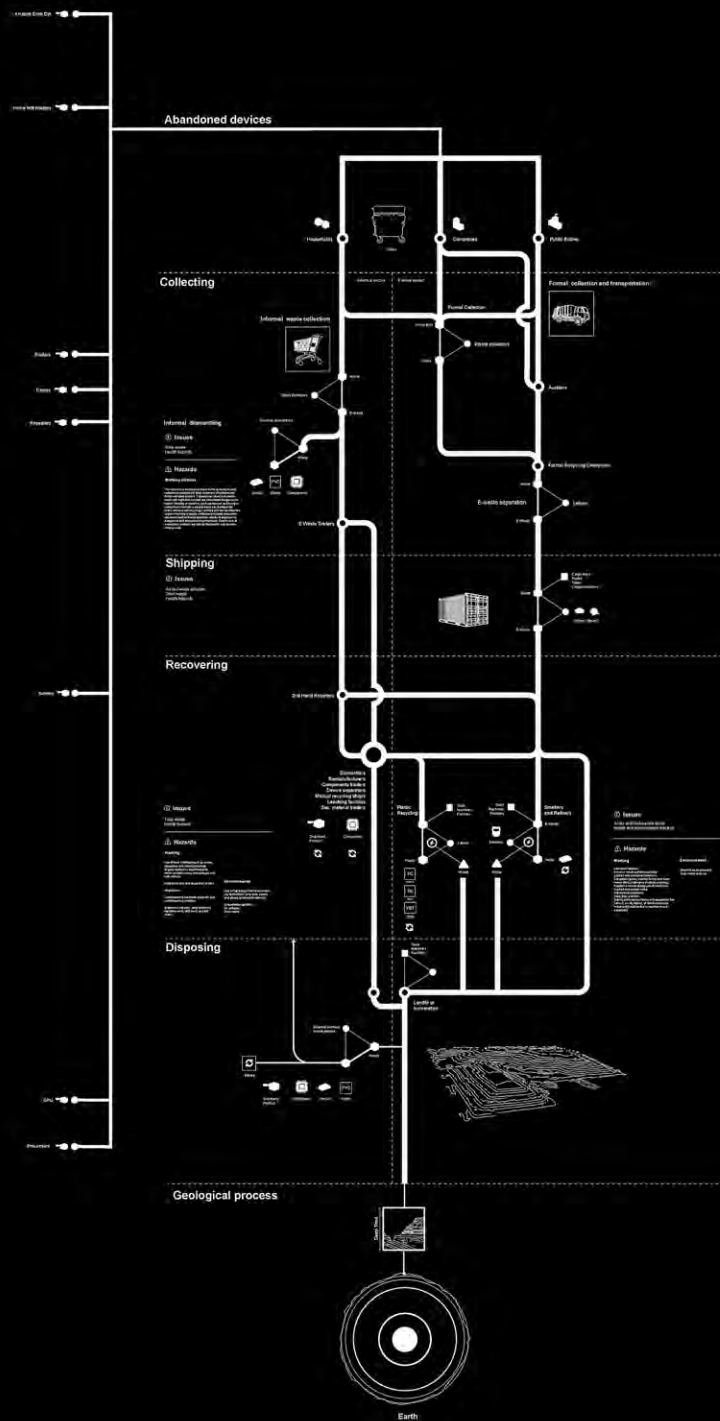
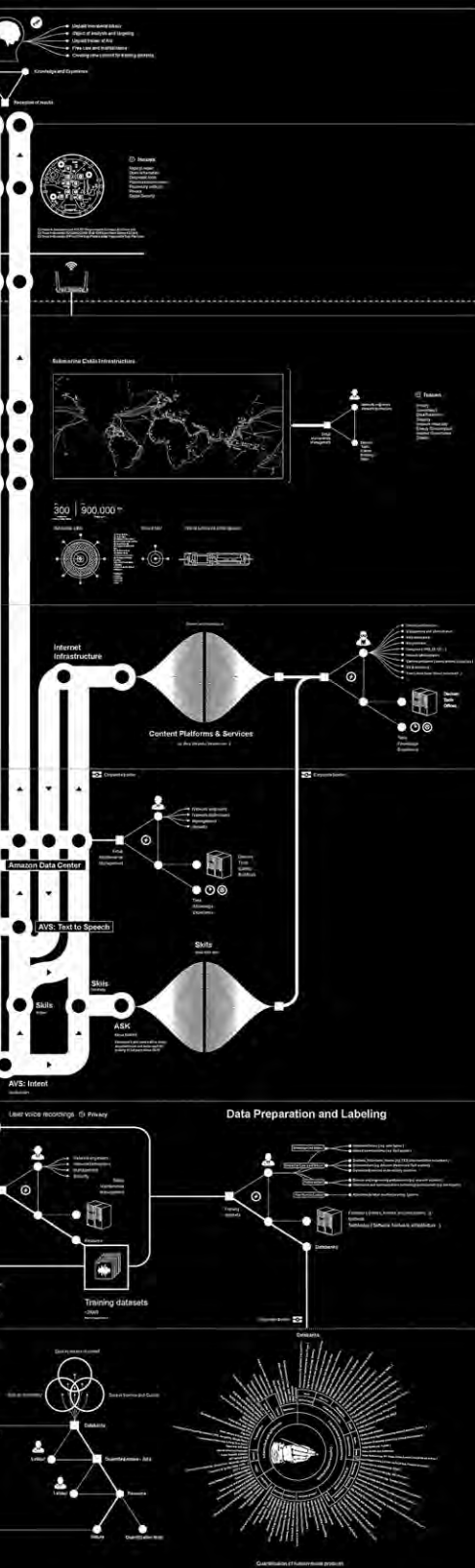


Quantification of nature



Data exploitation





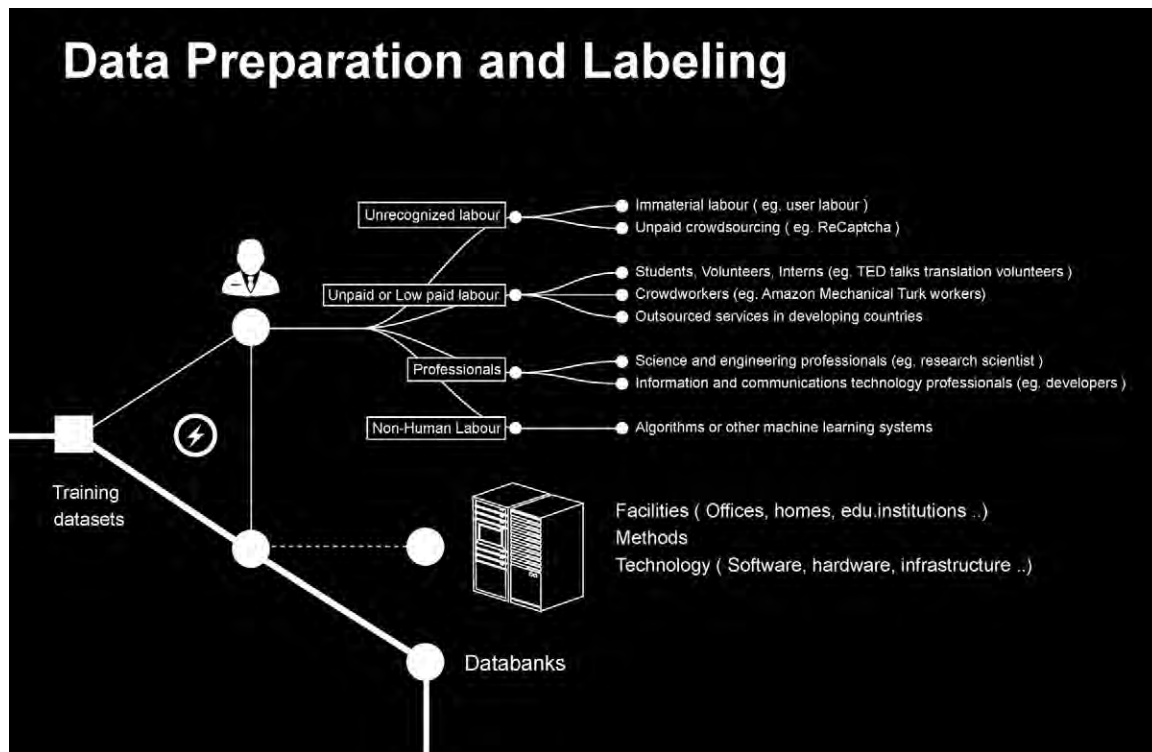


Figure 7.5 (continued)

invisible, and their contributions go unrecognized. The *screening techniques* that lead to their erasure, as Ahmed terms them, are not always intentional, but they are, unfortunately, self-perpetuating. Ahmed gives the example of sinking into a leather armchair that is comfortable because it's molded to the shape of your body over time. You probably wouldn't notice how the chair would be uncomfortable for those who haven't spent time sitting in it—those with different bodies or with different demands on their time. Which is why those of us who occupy those comfortable leather seats—or, more likely in the design world, molded plastic Eames chairs—must remain vigilant in reminding ourselves of the additional forms of labor, and the additional people, that our own data work rests upon.

This gets complicated quickly even on the scale of a single data science project. The names of all the people and the work they perform are not always easy to locate—if they can be located at all. But taking steps to document all the people who work on a particular project at the time that it is taking place can help to ensure that a record of that work remains after the project has been completed. In fact, this is among the four

core principles that comprise the Collaborators' Bill of Rights, a document developed by an interdisciplinary team of librarians, staff technologists, scholars, and postdoctoral fellows in 2011 in response to the proliferation of types of positions, at widely divergent ranks, that were being asked to contribute to data-based (and other digital) projects.⁴⁹

When designing data products from a feminist perspective, we must similarly aspire to show the work involved in the entire lifecycle of the project. This remains true even as it can be difficult to name each individual involved or when the work may be collective in nature and not able to be attributed to a single source. In these cases, we might take inspiration from the Next System Project, a research group aimed at documenting and visualizing alternative economic systems.⁵⁰ In one report, the group compiled information on the diversity of community economies operating in locations as far-ranging as Negros Island, in the Philippines; Quebec province, in Canada; and the state of Kerala, in India. The report employs the visual metaphor of an iceberg (figure 7.6), in which wage labor is positioned at the tip of the iceberg, floating above the water, while dozens of other forms of labor—informal lending, consumer cooperatives, and work within families, among others—are positioned below the water, providing essential economic ballast but remaining out of sight.

With the idea of underwater labor in mind, we might return to the example of GitHub, which began this chapter, to ask what additional forms of labor might contribute to the production of code but cannot be represented by the visualization scheme that GitHub currently employs. We might think of the work of the project manager, which is not directly expressed in a particular number or size or frequency of contributions, but nevertheless ensures the quality and consistency of all project code. We might wonder about the work of the designer on a project or of the technical writer—both of whom might have helped to shape the project in its initial phases, but who have likely moved on to other tasks. In the case of a consumer-facing project, we might also consider the contributions of the customer support teams. Or in a community-oriented project, we might include organizers who have spent years developing strong relationships with community members. These forms of labor, both productive and reproductive, are essential to the success of any project but are not currently rendered visible, nor could they ever be easily visualized, by a scheme that considers project contributions to consist of code alone.⁵¹

But in more instances than you might think, the labor associated with data work can be surfaced through the data themselves. For instance, historian Benjamin Schmidt, whose research centers on the role of government agencies in shaping public knowledge, decided to visualize the metadata associated with the digital catalog of the US Library of Congress, the largest library in the world (figure 7.7).⁵² Schmidt's initial goal was to understand the collection and the classification system that structured the



Figure 7.6

The “Diverse Economies Iceberg” (2017), a diagram of multiple labor practices created by the Next System Project for a report on cultivating community economies. Image courtesy of J. K. Gibson-Graham, Jenny Cameron, Kelly Dombrowski, Stephen Healy, and Ethan Miller for the Next System Project.

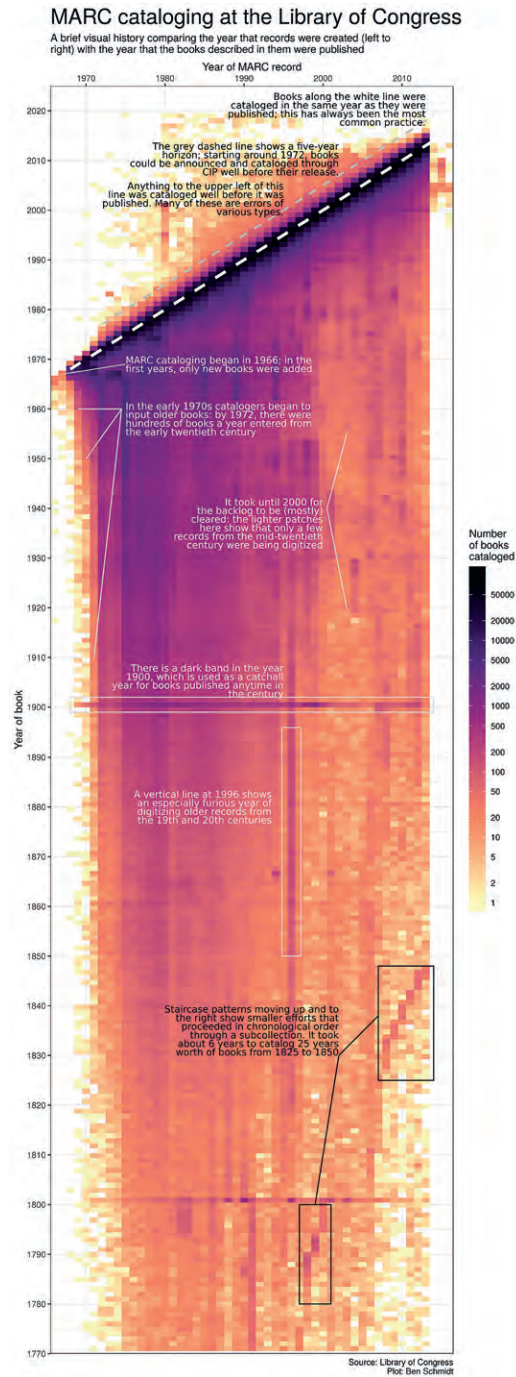


Figure 7.7

“A Brief Visual History of MARC Cataloging at the Library of Congress” (2017) visualizes when books at the Library of Congress entered their digital catalog. Image courtesy of Benjamin M. Schmidt.

catalog. But in the process of visualizing the catalog records, he discovered something else: a record of the labor of the cataloguers themselves. When he plotted the year that each book's record was created against the year that the book was published, he saw some unusual patterns in the image: shaded vertical lines, step-like structures, and dark vertical bands that didn't match up with what one might otherwise assume would be a basic two-step process of (1) acquire a book and (2) enter it in.

The shaded vertical lines, Schmidt soon realized, showed the point at which the cataloguers began to turn back to the books that had been published before the library went digital, filling in the online catalogue with older books. The step-like patterns indicated the periods of time, later in the process, when the cataloguers returned to specific subcollections of the library, entering in the data for the entire set of books in a short period of time. And the horizontal lines? Well, given that they appear only in the years 1800 and 1900, Schmidt inferred that they indicated missing publication information, as best practices for library cataloguing dictate that the first year of the century be entered when the exact publication date is unknown.

With an emphasis on showing the work, these visual artifacts should also prompt us to consider just how much physical work was involved in converting the library's paper records to digital form. The darker areas of the chart don't just indicate a larger number of books entered into the catalog, after all. They also indicate the people who typed them all in. (Schmidt estimates the total number of records at ten million and growing.) Similarly, the step-like formations don't just indicate a higher volume of data entry. They indicate strategic decisions made by library staff to return to specific parts of the collection and reflect those staff members' prior knowledge of the gaps that needed to be filled—in other words, their intellectual labor as well. Schmidt's visualization helps to show how the dataset always points back to the *data setting*—to use Yanni Loukissas's helpful phrase—as well as to the people who labored in that setting to produce the data that we see.⁵³

Crediting Emotional Labor and Care Work

In addition to the invisible labor of data work, there is also labor that remains hidden because we are not trained to think of it as labor at all. This is what is known as *emotional labor*, and it's another form of work that feminist theory has helped to bring to light.⁵⁴ As described by feminist sociologist Arlie Hochschild, emotional labor describes the work involved in managing one's feelings, or someone else's, in response to the demands of society or a particular job.⁵⁵ Hochschild coined the term in the late 1970s to describe the labor required of service industry workers, such as flight

attendants, who are required to manage their own fear while also calming passengers during adverse flight conditions, and generally work to ensure that flight passengers feel cared for and content. In the decades that followed, the notion of emotional labor was supplemented by a related concept, *affective labor*, so that the work of projecting a feeling (the definition of emotion) could be distinguished from the work of experiencing the feeling itself (the definition of affect).⁵⁶

We can see both emotional and affective labor at work all across the technology industry today. Consider, for instance, how call center workers and other technical support specialists must exert a combination of affective and emotional labor, as well as technical expertise, to absorb the rage of irate customers (affective labor), reflect back their sympathy (emotional labor), and then help them with—for instance—the configuration of their wireless router (technical expertise).⁵⁷ In the workplace, we might also consider the affective labor required by women and minoritized groups, in all situations, who must take steps to disprove (or simply ignore) the sexist, racist, or otherist assumptions they face—about their technical ability or about anything else. And they must do so while also performing the emotional labor that ensures that they do not threaten those who hold those assumptions, who often also hold positions of power over them.⁵⁸ Are there ways to visualize these forms of labor, giving visual presence—and therefore acknowledgement and credit—to these outlays of work?

One example that strives to visualize emotional and affective labor is the Atlas of Caregiving (figure 7.8), an ongoing project that aims to document the work involved in caring for a chronically ill family member. The project's name plays on the concept of the anatomy atlas, a compendium of illustrations of the human body that doctors can consult for information and reference. In this case, the goal was to illustrate the sometimes physical and sometimes emotional or affective work of care. The research team outfitted its participants with a variety of biometric sensors, including accelerometers and heart rate monitors, as well as with body cameras programmed to take a picture every fifteen minutes. They then visualized these data alongside excerpts from personal interviews and from the activity logs they asked the caregivers in the study to complete.

The result is a complex picture of caregiving, one that marshals data in the interest of creating a comprehensive view of the range of labor involved in caregiving work.⁵⁹ The stress of serving as a caregiver—a form of affective labor—is broken down into six distinct levels, and then visualized as a gradient (figure 7.8a). The work of caregiving itself is divided into seven subtypes of work, including concrete tasks like healthcare management and household chores, and more abstract forms of labor like being available and social support (figure 7.8b). This, too, helps others recognize the wide range

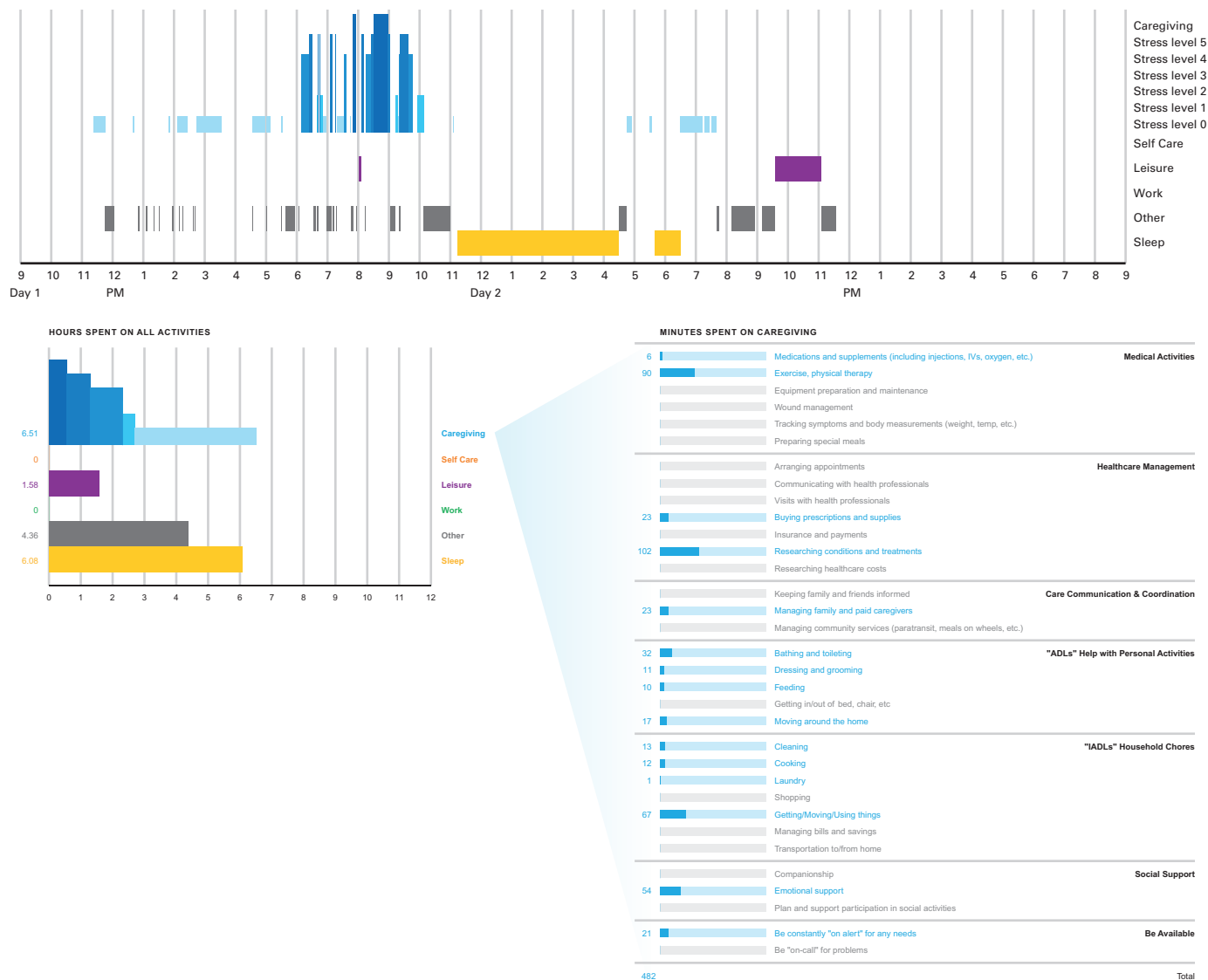


Figure 7.8

The Atlas of Caregiving visualizes the labor of caring for chronically ill family members. (a) A thirty-six-hour log of caregiving activities; (b) caregiving activities separated by type; (c) a photo log created during that same time. Image courtesy of the Atlas of Caregiving, 2016.

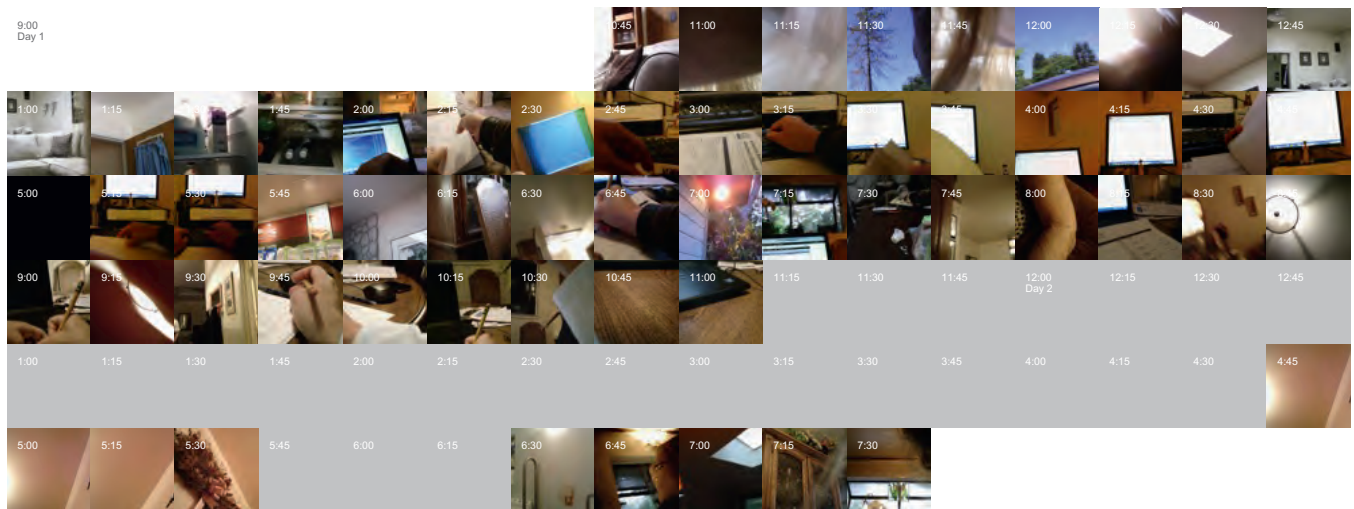


Figure 7.8 (continued)

of work—indeed, expertise—associated with caregiving. And as some of the study’s participants reported, it helped them to recognize that work for themselves.⁶⁰

Of course, a diagram of work is only a proxy for the work itself—and that is to say nothing about the complexity of human feelings. This understanding served as the genesis for “Bruises—the Data we Don’t See.”⁶¹ This artful visualization, created by designer Giorgia Lupi and accompanied by a musical score composed by Kaki King, attempts to get closer to a visual representation of the emotional toll of caregiving (figure 7.9). The project began when King’s daughter was diagnosed with a rare autoimmune disease, idiopathic thrombocytopenic purpura (ITP). ITP is described as a “very visual disease,” and presents as bruises and burst blood vessels all over the body. For this reason, King was instructed to watch her daughter’s skin and record any significant changes. She also recorded her own feelings in terms of hope, stress, and fear, creating subjective data to complement the hard numbers she received from the blood tests her daughter was required to endure.

When Lupi, who knew King from previous collaborations, set out to design her visualization, her goal was to “evoke empathy” and help her audience “feel a part of

Figure 7.9 (following three pages)

Still from *Bruises—the Data We Don’t See* (2018) and the legend that helps decode the data visualization. Image courtesy of Giorgia Lupi and Kaki King.





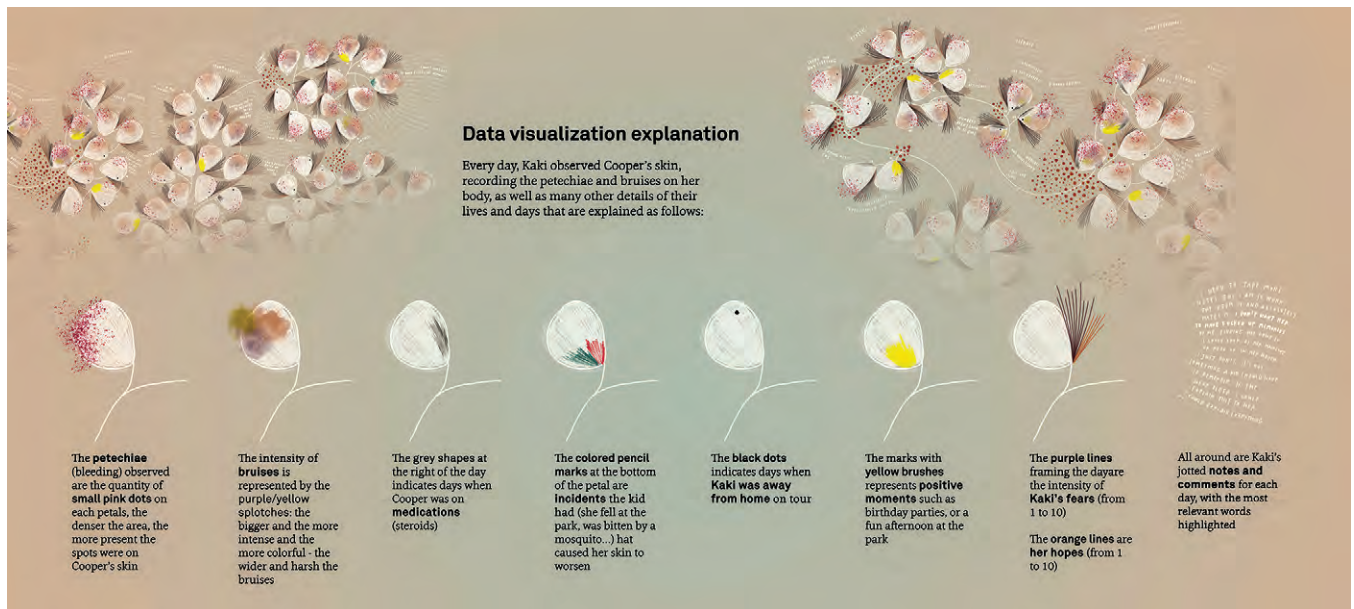


Figure 7.9 (continued)

a story of a human's life."⁶² In contrast to the Atlas of Caregiving, which relies upon standard visualization techniques like radial timelines and Gantt-style charts to legitimate the work of care, Lupi sought alternative visualization strategies to emphasize the particularity and specificity of a single family's situation. She employed a fluid timeline to reflect the subjective nature of what disability studies scholars call *crip time*. With this term, as Ellen Samuels explains it, "Sometimes we just mean that we're late all the time—maybe because we need more sleep than nondisabled people, maybe because the accessible gate in the train station was locked."⁶³ But it can also mean something more profound, as described by Alison Kafer: "Rather than bend disabled bodies and minds to meet the clock, crip time bends the clock to meet disabled bodies and minds."⁶⁴

In Lupi's depiction of how the clock met King's daughter's body and King's own mind, days became white aspen-shaped leaves, segmented not by weeks or years but by hospital visits. Red dots were employed to indicate platelet counts, with color deployed mimetically to convey the intensity of the bruises, as well as the visuality of the data recorded by King. Lupi also employed color to represent King's record of her feelings, with black corresponding to stress and fear and yellow to signify hope. King's fear and hope were also visualized by hand-drawn lines that reflected each on a scale of one to ten. The result is rendered as an animation that unfolds over time and is set to

music, a visually and aurally affecting composition of the affective labor of mothering and care.

Of course, neither Lupi and King nor the Atlas of Caregiving project team are the first to want to identify and make visible the work of care. As early as 1969, shortly after the birth of her own child, artist Mierle Laderman Ukeles penned the Manifesto for Maintenance Art, which called on the art world to elevate the care and maintenance of human life to an art, over and above the solitary creative (male) genius.⁶⁵ In the years that followed, *care work* would become a significant topic of interest for feminist scholars—especially after the mid-1990s, when Nancy Folbre formalized the term. Folbre’s primary model of care work was the everyday work of caring for a child, although care work, like housework, isn’t necessarily performed for free. It can also include the underwaged work performed by daycare workers or home health aids, as well as the waged work of doctors, nurses, physical therapists, mental health professionals, and so on. What binds these forms of work together across economic lines is their motivation. As theorized by Folbre, care work is undertaken out of a sense of compassion or responsibility for others, rather than with a goal of monetary gain. But when it comes to the market, altruism is a double-edged sword. These same professional care workers—who are predominantly women and people of color—are often paid less than they would be in other fields.⁶⁶ Why? Because they care.

So how do we “show the work” of care workers? How do we ensure that this work is sufficiently recognized and valued? And can we do anything more to challenge the root cause of this undervalued work? In the academy, groups like the Maintainers have sought to learn from the theories of care developed by feminist labor studies scholars such as Folbre as they attempt to make visible and value the labor of data work.⁶⁷ Through workshops, conferences, and publications, the Maintainers seek to counter the current tendency in technology fields to celebrate innovation and discovery alone. The work that maintains and sustains the world we live in today should also be celebrated, they insist. Among their current areas of research are the people they call *Info-Maintainers*: the people who work in libraries and archives and in related preservation fields to ensure that the knowledge of the present remains accessible for generations to come. Because the work of librarians and archivists and curators is focused on facilitating access to future knowledge, the Maintainers argue, it can be viewed as a form of care work too.

Across many technical fields, there is an increasing amount of attention paid to care work, and to other forms of invisible labor, now that so much work is virtual rather than physical; as well as to issues of job insecurity now that white-collar jobs have begun to be outsourced to freelancers as well. In this context, it is important to

recall that professional care workers have long dealt with issues of undercompensated and precarious work; and for just as long, they have been involved with efforts to resist and organize against the inequities they have faced. Today, these efforts are being enhanced by data and technology, as unions and other advocacy groups are making use of new platforms and data streams for their work. But they are also being obstructed, as Uber-style apps to connect caregivers and employers increasingly abound. These apps do nothing to solve the systemic problems that caregivers face. A 2016 study of on-demand domestic worker apps by the UK's Overseas Development Institute (ODI) reports that because they displace risk onto workers, these platforms potentially reinforce discrimination and “further entrenchment of unequal power relations within the traditional domestic work sector.”⁶⁸

As a corrective, we might look to emerging prototypes that center the needs of workers, those that are developed by and with workers themselves. In the US, for example, the National Domestic Workers Alliance (NDWA) has developed an app, Alia, in order to serve as a portable benefits platform.⁶⁹ It allows clients to contribute a small amount into the worker's benefits account each time that worker provides them with a service. Workers can then pool contributions from multiple clients to purchase benefits on-demand, such as paid time off and various forms of insurance. Caveats remain, of course: Shouldn't the government require that all workers receive paid time off as a matter of course? Shouldn't we be advocating for a single-payer healthcare system? Yes and yes. But while the NDWA continues to lobby for systemic change, its app offers one way to provide essential benefits to domestic workers right now. It is a harm reduction strategy, one that can be pursued while simultaneously advocating for more transformative change. Thinking back to Kimberly Seals Allers's app, Irth, discussed in chapter 1, we might also begin to imagine how its successful use would contribute to a dataset that could be used to support future advocacy efforts.

Show Your Work

Data work is part of a larger ecology of knowledge, one that must be both sustainable and socially just. Like the ship paths visualized on the *Ship Map* or the source code stored on GitHub or the global assemblage of people and materials that make an Amazon Echo device, the network of people who contribute to data projects is vast and complex. Showing this work is an essential component of data feminism, and it is the reason why “show your work” is the seventh and final principle in this book. An emphasis on labor opens the door to the interdisciplinary area of data production studies: taking a data visualization, model, or product and tracing it back to its

material conditions and contexts, as well as to the quality and character of the work and the people required to make it. This kind of careful excavation can be undertaken in academic, journalistic, or general contexts, in all cases helping to make more clearly visible—and therefore to value—the work that data science rests upon.

We can also look to the data themselves in order to honor the range of forms of invisible labor involved in data science. Who is credited on each project? Whose work has been “screened out”? While one strategy is to show the work behind making data products themselves, another strategy for honoring work of all forms is to use data science to show the work of people (mostly women) who labor in other sectors of the economy, those that involve emotional labor, domestic work, and care work. We see this in action in the Atlas of Caregivers, which focuses on legitimizing care work, and the Alia app, which provides more financial security for domestic workers. Designing in solidarity with domestic workers can begin to challenge the structural inequalities that relegate their work to the margins in the first place.

This point brings us back to the ideas about power that began this book. Power imbalances are everywhere in data science: in our datasets, in our data products, and in the environments that enable our data work. Showing the work is crucial to ensure that undervalued and invisible labor receives the credit it deserves, as well as to understand the true cost and planetary consequences of data work.

This is a section of [doi:10.7551/mitpress/11805.001.0001](https://doi.org/10.7551/mitpress/11805.001.0001)

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 open access edition of this book was made possible by generous funding and support from the MIT Libraries Experimental Collections Fund



The MIT Press

© 2020 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC BY-NC-ND license. Subject to such license, all rights are reserved.



This book was set in ITC Stone Serif Std and ITC Stone Sans Std by Toppan Best-set Premedia Limited.

Library of Congress Cataloging-in-Publication Data

Names: Kanarinka, author. | Klein, Lauren F., author.

Title: Data feminism / Catherine D’Ignazio and Lauren F. Klein.

Description: Cambridge, Massachusetts : The MIT Press, [2020] | Series: ideas series | Includes bibliographical references and index.

Identifiers: LCCN 2019036137 | ISBN 9780262044004 (hardcover)

Subjects: LCSH: Feminism. | Feminism and science. | Big data—Social aspects. | Quantitative research—Methodology—Social aspects. | Power (Social sciences)

Classification: LCC HQ1190 .K375 2020 | DDC 305.42—dc23

LC record available at <https://lcn.loc.gov/2019036137>

10 9 8 7 6 5 4 3 2 1