

Living in Data

A Citizen's Guide to a Better Information Future

Jer Thorp



MCD

Farrar, Straus and Giroux | New York

Begin Reading

Table of Contents

[A Note About the Author](#)

[Copyright Page](#)

**[Thank you for buying this
Farrar, Straus and Giroux ebook.](#)**

[To receive special offers, bonus content,
and info on new releases and other great reads,
sign up for our newsletters.](#)



Sign Up

Or visit us online at
us.macmillan.com/newslettersignup

[For email updates on the author, click here.](#)

The author and publisher have provided this e-book to you for your personal use only. You may not make this e-book publicly available in any way. **Copyright infringement is against the law. If you believe the copy of this e-book you are reading infringes on the author's copyright, please notify the publisher at: us.macmillanusa.com/piracy.**

For Nora.

There are a hundred thousand species of love, separately invented, each more ingenious than the last, and every one of them keeps making things.

—Richard Powers, *The Overstory*

Preface

I am writing this under curfew. The police are, right now, roaming this city in great packs, batons in hand. They are herding protesters, kettling them, beating them, zip-tying their wrists, detaining them for hours without food or water or medical care. It is our fourth night of this. It is our seventy-seventh day of shutdown, of shelter in place; we are at some unclear point in a pandemic that may last many more weeks or months or years.

It has been a spring of data. Numbers of cases, numbers of tests, numbers of deaths. Numbers about social distancing and herd immunity. Numbers of protesters, of Black people killed by police. Numbers, numbers, numbers. On March 27, *The New York Times* published a chart that showed the unprecedented numbers of people seeking unemployment insurance that week in the country; the right-most bar (3.3 million) stretched from the very bottom of the page to the top. The *Times* did it again on May 9, when the monthly figure had passed twenty million. On May 24, the newspaper printed a thousand names of people who had died so far from the coronavirus: a thousand names that spilled off the front page and into the next. Angel Escamilla, sixty-seven, Naperville, Illinois. *Assistant pastor*. April Dunn, thirty-three, Baton Rouge, Louisiana. *Advocate for disability rights*. Joseph Yaggi, sixty-five, Indiana. *Mentor and friend to many*. Alby Kass, eighty-nine, California. *Lead singer of a Yiddish folk group*.

In my neighborhood, on the edge of the East River, someone has been leaving little painted rocks on the trails in the park. “Everything is going to be OK,” one says, painted in tidy black letters on white. “We are all in this together,” reads another.

We are not all in this together. Read those names in *The New York Times* more closely, dig into the numbers, and you’ll find out that the virus has affected Black people and Black communities more severely than white ones. In Minneapolis, where George Floyd was killed on May 25, 2020, the police are seven times more likely to use violence against Black people than they are against white people. It is the “we” in the stone’s platitude that is

problematic.

That same collective “we” has too often been data’s “all lives matter,” a way to soft-pedal concerns about privacy while refusing to speak directly to dangerous inequalities. If I want to talk about the human experience of data, I need to talk about risk, and risk is something that does not affect people equally or at the same time. Living in data may seem like a shared reality, but it is an experience that critically differs from person to person and from group to group. One two-letter word cannot possibly hold all of the varied experiences of data, specifically those of the people who are at the most immediate risk: visible minorities, LGBTQ+ people, indigenous communities, the elderly, the disabled, displaced migrants, the incarcerated.

Much of the book you’re about to read is written in the first-person plural. That is to say that I’m going to use the word “we” a lot, wrapping you and me together. It makes me uncomfortable, and it should make you uncomfortable too. But I believe it’s necessary. To borrow from Julietta Singh, my “we” is “a hopeful summons.” It exists not in the present but in some better future that I have to believe we will arrive at.

This is only the second time in New York’s history that the city has been placed under curfew. The first time was August 1, 1943. It was the night of the Harlem riots, which had erupted after a white police officer, James Collins, shot a Black soldier, Robert Bandy. Mayor LaGuardia sent a force of six thousand police and fifteen hundred civilians into Harlem to quell “outbreaks of hoodlumism.” Six hundred people were arrested, and several were killed. That night was also the funeral for James Baldwin’s stepfather, and he wrote of driving to the graveyard “through a wilderness of smashed plate glass.” The riots and his stepfather’s death would inspire one of his most famous essays, “Notes of a Native Son,” which ends with these words:

It began to seem that one would have to hold in the mind forever two ideas which seemed to be in opposition. The first idea was acceptance, the acceptance, totally without rancor, of life as it is, and men as they are: in the light of this idea, it goes without saying that injustice is a commonplace. But this did not mean that one could be complacent, for the second idea was of equal power: that one must never, in one’s own life, accept these injustices as commonplace but must fight them with all one’s strength. This fight begins, however, in the heart and it now had been laid to my charge to keep my own heart free of hatred and despair. This intimation made my heart heavy and, now that my father was irrecoverable, I wished that he had been beside me so that I could have searched his face for the answers which only the future would give me now.

How does one speak about something that is
both fish and water, means as well as end?

—URSULA FRANKLIN,
THE REAL WORLD OF TECHNOLOGY

Prologue

Hal Fisk stood in the Louisiana sun, wiping the sweat from his brow with a clean white handkerchief. It was one of a few dozen that Emma always packed for him on these trips, for the sweat and for the mud. There was mud still in the deep furrows of his geologist's hands, even after he'd wiped them clean and then wiped them clean again. There was mud, too, caked and dried up both of Hal's arms to where he'd rolled up his shirtsleeves neatly in the morning. To do any sort of work out here was to be very intimate with mud, no way around it.

They'd been trying to get the bore rig started all morning, but the big bit kept on getting caught on something before it could build up enough torque. An old tree root, it turned out, so Bill Hendy was neck-deep in the muck with a hacksaw trying to get the thing out. Hal hoped they could get the machine started soon; around noontime the heat would be downright unbearable. This was the second last drill between Harrisonburg and Natchez, and he was eager to get back to Vicksburg, to Emma and the lab and to the data.

Where Hal was standing was a good ten miles from the Mississippi. There was another river that ran much closer by, the Tensas. It was a stream really, choked with turtle grass, half decent for fishing according to the locals but barely wide enough for a boat even after a rain. What the mud would tell them, Hal knew, was that the Tensas was the meandering ghost of a Mississippi past, a last trickle of a course abandoned some three thousand years before.

It had been clear from the beginning, to the first humans who had navigated the river, that the Mississippi was restless. Over time its curves would widen, the water becoming more shallow. Eventually the elbows of the river would double back on themselves, and the water, taking what it deemed the most rational path, would decide to avoid the old part altogether, would flow across the ground to the next loop of the river, cutting off a piece of itself. And repeat.

Evidence of this bending and re-bending lives in the mud and in the life

that sprang from it. A cutoff elbow of the river might spend a century or two as a stranded lake. During this time it discards its riverine identity and becomes a closed system, something more still. Largemouth bass and quick silvery dace are replaced by languid longnose gar and deep-burrowing catfish. At the water's edges, new plants arrive, more suited to the idle water of the lake. Year to year, decade to decade, they send their roots into the water to drink, and the soil builds around them. It is a drawn-out argument between land and water, won before it began. The water is consumed, and what is left is a lake-shaped mark of greenery, growing on top of a lake-shaped layer of mud.

From the sample they'd draw today, and some sixteen thousand others, Hal and his team would reconstruct the river as it was. Some of the boreholes the team drilled reached down more than thirteen thousand feet into the earth, and the rock there held evidence of the river's beginnings after the last great ice age, some two and a half million years ago. From mud and rock and aerial photographs, they would draft maps of each of the river's abandoned curves, stacked on top of each other, a record of a huge, timeless, writhing thing. The result was a report, delivered in 1944, titled *Geological Investigation of the Alluvial Valley of the Lower Mississippi River*. From all those columns of earth and images from airplanes they built an epic geological story, one more complete than had ever been told. In the report, you can read of roaring glacial runoffs, of boulders carried across continents. Of the Sunflower system, the Thebes Gap, the Walnut Bayou Segment, the Gulf Coast geosyncline. Of bed materials and bank recession and salt domes and uplifts.



Fisk's work was a gargantuan feat of data collection, of synthesis and mapmaking. From mud and patchwork photographs, he gave the river, with its school-rhyme name, a 170-page origin story, of a place, of a landscape, of a way of life.

There were two hundred copies of Fisk's report printed, and for \$2.50 plus postage it'd be delivered to you, a compact book printed on white paper, and twenty-six map plates neatly folded into quarters and tucked into a brown paper envelope. One of the copies ended up at the New York Public Library, and a few months back I went to the main branch in Bryant Park to visit it. Even though I'd seen images of the maps many times, when I unfolded the first plate onto one of the reading room's broad wooden tables, my heart skipped a beat. The maps are gorgeous, visceral things, rendered in a brilliant color palette of green and yellow and orange and red. And they are big: I could fit only four plates at a time on the 20-foot-long table; to see the whole set of maps in one place, you'd need a hallway 108 feet long.

Seeing that I was trying to get a photograph of the maps from above, one of the librarians brought out a stepladder, and I climbed to the top. It might have been vertigo (I'm not good with heights), but a kind of dizziness

overtook me as I looked down onto the four plates, onto a stretch of the river that spanned twenty-four hundred miles and ten thousand years. For a moment I saw a map of this book you have in front of you, and I realized that the story I have to tell is, too, one of meanderings, of swift currents and slow corners, of silt and history. It's a story that has come to be in part by looking down from above, watching and listening and reading, spending hours in archives. Mostly, though, it has come from time spent waist-deep in data: in research labs at *The New York Times* and the Library of Congress (LOC), on expedition with *National Geographic* field teams, in the galleries at the Museum of Modern Art, at the bottom of the ocean, in cities like Calgary and St. Louis and Manchester and in the middle of Times Square. There is mud on my hands.

There are 164 rivers that join the Mississippi along its two-thousand-mile course from Lake Itasca to the Gulf of Mexico. The L'Anguille, the Skokie, the Edwards, the Grant, the Iowa, the Missouri. West Branch Sugar River and Hay Creek and Bayou Pierre. The biggest of these have their own tributaries, so to consider the river is to stretch from Montana to Pennsylvania; from Saskatchewan to New Orleans. Read about data in the newspaper, read about it online, and it might fool you. It might fool you into believing that the path we're on is a predetermined one, that the way in which we are being carried along its course is the only way we can go. That its history could be drawn with one clean line. To tell the real story of data, though, is to speak not only of the big topics, of Facebook and machine learning and targeted advertising and visualization and facial recognition. To get the whole picture, we need to visit the smallest creeks and brooks and streams, the places where the story bubbled up a century ago, and the places where we can find new stories just beginning to flow.

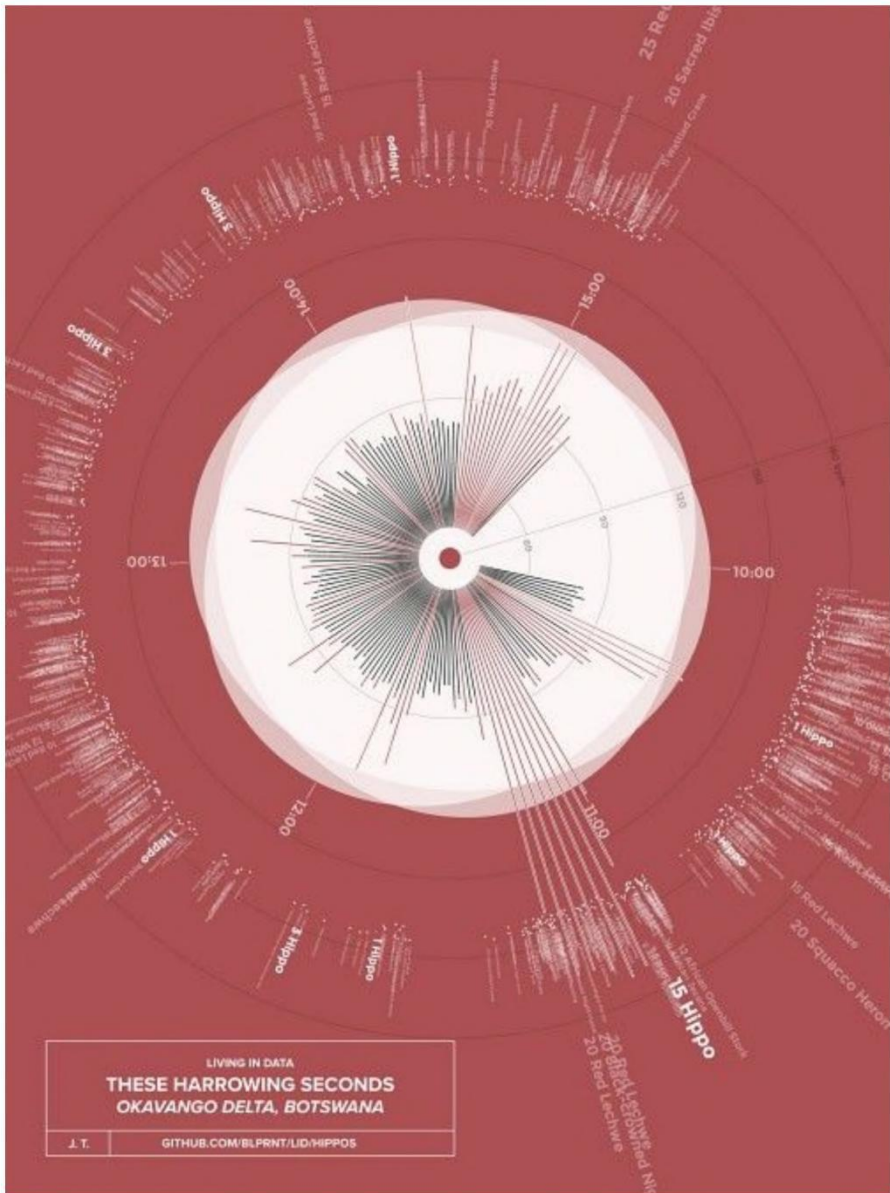
Helen Hall Jennings sits in a classroom in Brooklyn, in 1931, watching the students, writing in her notebook. A seventeen-year-old in Cleveland dials into the nation's largest Free-Net. Outside a school in San Francisco in 1999, children plant two walnut saplings, identical clones. In a theater in Birmingham in 2003, a woman is on her hands and knees, carefully counting out 2,435 grains of rice. In the highlands of Angola in 2015, a convoy of mine-clearing vehicles rumbles through a pristine forest. On the north side of St. Louis, a group of ninth graders project census data onto a hand-drawn map. On a glacier high in the Canadian Rockies, a sensor records a sudden movement of the ice.

There is no straight line to be drawn between these heres and theres, these nows and thens. This book will, by necessity, meander. It will double back and reverse course.

Step back, though, climb up onto a ladder, and I hope you'll see a bigger story.

Now it flows like a broad billow over the whole land, now it divides itself into a gigantic net of thin streams; now it bubbles forth from under the ground ... All run through one another, next to each other, across one another, flow in and over one another; it is an eternal, moving, changing sea of appearances.

—ROSA LUXEMBURG



1. Living in Data

It's 11:01 a.m., and I'm about to be attacked by a hippopotamus.

I've replayed this event many times in my mind's eye: the swell of the wave approaching in the clear water. The rock of the boat as we try to brace ourselves for impact. The shouts from the people around us as they realize what is about to happen. *Kuba! Kuba! Hippo! Hippo!* These harrowing seconds are being recorded as data, as the output of a heart-rate monitor I'm wearing across my sweaty chest. Looking now at those numbers, the actual millisecond-by-millisecond beats of my heart, I can see my distress building. As a graph, it reads like an elevation map of terror, each successive peak taking me closer to the hippo's arrival, or to cardiac arrest.

I recently wrote a piece of software to turn those numbers back into sound, a kind of a thump-by-thump re-creation of the attack, and I've got headphones on right now, listening. As the Jer sitting in that boat gets more and more terrified, so does the Jer sitting here in this chair, in my studio in Brooklyn. It's pretty easy to tell myself that there isn't a hippo here, in this room, but at the same time the data is a convincing record of the most nerve-racking experience of my life.

Despite being the world's largest amphibious animals, hippos aren't great swimmers. The adult males weigh about as much as a minivan, and they don't float. They prefer to stay in the shallows, where their feet can touch the ground. Just deep enough that their eyes and ears and nose—stacked up at the top of their enormous heads—remain out of the water. A scared hippo, though, or a very agitated one, will venture into a lake or a pond or a river channel, moving with great porpoise-like leaps off the bottom. Hydrodynamics be damned.

I wondered, as I watched the hippo-sized bow wave surge toward me, what am I doing here?

I tripped and fell into data, into that boat and this book, one Saturday in the spring of 2009. I was sitting at the little Ikea desk in my East Vancouver flat. The cherry trees that lined my street had just burst into bloom, and the

floor under my chair was sticky with the pink petals I'd tracked in after my morning dog walk. I was just about to give up (again) on a project I'd been working on and reworking for nearly four years. Its central question had come to me one day while I was staring at my screen: What if pixels could do what they want? What if we could unbind them from their tedious life of following instructions: how bright to shine, when to blink on and off, what exact shade of orange they must display.

In my project I'd set the pixels free, letting them trade color with each other in a miniature economy. I coded the pixels to each have a kind of personality: some were conservative; others were happy to take risk. Some of them looked at trends in the color "market" to decide which trades to offer; others listened to a coded oracle, which spit out a series of predictions based on random numbers. Each color block had agency; it was free to make whatever decisions its little programmatic brain might settle on. As a group—a population—these individual foibles would emerge into pattern, and the system would be, in a small sense, alive.

The problem was that it didn't work. No matter how I set the parameters, the economy would collapse within ten thousand or so rounds of trading. I'd be left with two or three extremely wealthy pixels, and the rest would be broke. And dead. I tried changing the starting conditions, setting the color "wealth" of each pixel from different images, photos of sunsets or deserts or wildfires or drawings of national flags or snaps from my webcam. I tried implementing a taxation system, where money was distributed to the poor pixels from the wealthy ones. Some of these solutions worked, for a long minute or two, and then the whole thing collapsed again to the very rich and the very dead.

I decided what the system needed was some chaos, some noise from the real world that might keep the economy on its toes. I looked first at feeds from the stock market, but that seemed far too literal for my pixel population. And then I had an idea: What if the real-world usage of the words "red," "green," and "blue" drove their value in the color economy? If I could get the text from news articles, I could write a program to count these color words and then feed the numbers into my system. I googled. In what I now recognize as a moment that crackled with serendipity, the first result I read was about a new data service that *The New York Times* had released the day before, an interface that allowed anyone to search thirty years of articles and get back lists of results. Headlines, bylines, content summaries, web URLs, and, with a little bit of work, occurrences of specific words and phrases.

I never did finish the color project. I got caught instead in the sweeping currents of data's possibilities. That afternoon I wrote a program to download 972 numbers from the *Times*. The numbers were counts of how many times "red," "green," and "blue" had appeared in the newspaper between the years 1981 and 2008. My computer dutifully packaged up the requests for the numbers, twelve at a time, and after a few minutes of a gray screen a graph appeared. It was my first data visualization.

The graph itself was hardly auspicious. It was rendered in the gaudy primary colors of a day care (or a Google office), the bars sat on top of each other, and there was no way to tell one month from another or one year to the next. Still, looking at this ugly thing, I could see some promise. There was pattern, if you looked closely. While blue and red seemed to oscillate with no regular pattern, the bar graph for green was a line of rounded hummocks, each twelve months long, the color of the seasons reflected in the language of the news. There was a big spike in the red graph in March 2002—the result of Homeland Security Presidential Directive 3 and its rainbow scale showing “the risk of terrorist acts.” I spent hours reading the data returns and matching them to every little peak in the graphs; there was a whole history wrought in color: Deep Blue and Red Square and green energy, Blue Cross, Red Cross, Green Berets.

I tried new combinations of words: first “sex” and “scandal,” then “internet” and “web,” then “Iran” and “Iraq.” “Innovation” and “regulation,” “Christianity” and “Islam.” “Superman,” “Batman,” and “Spider-Man.” “Global warming” and “climate change.” “Hope” and “crisis,” “science” and “religion,” “communism” and “terrorism.” Each of these sets of words told its own visual story; each of them showed some change in how the words were used by the writers and editors at the *Times* and how they were read by millions of readers. How satisfying this simple thing was, this trick of turning numbers into shapes and colors.

I discovered that I could draw connections between people and organizations if they appeared in the same article, and from this realization came dense maps of entire years of news. Ronald Reagan, the Roman Catholic Church, the United Nations, Michael Dukakis, George Bush, Salman Rushdie. The ANC, David Dinkins, General Motors, Bill Clinton, Jim Bakker, the PLO. Reading the maps, year by year, was like a fast-forwarding through history, or at least through the history that had been told by *The New York Times* (presidents, for the most part, occupied the center of the maps, except in the years when the Yankees won the World Series).

I spent months adrift in the possibility space of visualization, where, it seemed, I could conjure pattern from nothing and from everything. When I got tired of the *Times*, I visualized the U.K.’s National DNA Database, the influenza genome, Obama’s foreign policy speeches and State of the Union addresses, and international relief donations to Haiti. I mapped everyone who said “good morning” on Twitter in twenty-four hours, and analyzed language from sixteen hundred issues of *Popular Science*. I plotted vessel traffic in the world’s biggest shipping ports and mapped the narrative structure of Haruki Murakami’s short stories. I created time lines of every character in every issue of the classic *Avengers*. In one of my favorite projects, I reverse engineered a map of global air travel from people tweeting “I just landed” as they touched down in airports all around the world.

I became captivated with what I call “question farming”: using visualization not to simplify something but to unfurl its complexities in

interesting ways, exposing things that weren't before able to be seen. John Tukey, one of the great defining figures of modern statistics, would describe this kind of work as exploratory analysis, rather than the more task-oriented confirmatory analysis. Personally, though, I discovered that this wide-open exploration brought joy, as opposed to data visualization's defining emotion, *satisfaction*.

In the early fall of 2010, I walked into the New York Times Building on Forty-Second Street and took the elevator to the fourteenth floor, where I'd spend two and a half years as the company's first data artist in residence (a title I made up). It was fertile soil. With my colleagues there I built the first large-scale tool for exploring social media data, a kind of interactive forensic instrument for conversations on Twitter. With it you could clearly see both the exhilarating expansiveness and the tangled complications of the then-nascent social network. While I was at the *Times*, I started teaching at New York University's Interactive Telecommunications Program (ITP), a kind of punk rock version of MIT's Media Lab, where I set my students out into the loamy dirt at the edges of data's possibility space, digging, planting, seeing what would grow.

When I left the *Times* in 2013, I started a studio, the Office for Creative Research (OCR), and for almost a decade we tried to break as many of data's rules as we could. We performed data at the Museum of Modern Art and built it into a sculpture in the middle of Times Square. We made tools to give people ways to navigate data's wash: a browser extension that analyzes the web ads that swamp our browsers, a pop-up data community center in North St. Louis, a citizen science platform for chronic pain sufferers. Somewhere in the midst of all this I was (much to my surprise) named a *National Geographic Explorer*, and my work (and the OCR's) seeped out of screens and cities and into wilder places. Keen to put my new credentials to work, I joined an expedition into the heart of Botswana's Okavango Delta, having traded my data skills for three square feet in the front of a boat.

Spoiler: I didn't die. The hippo decided, having run the energetic equations, that we weren't worth the effort of a capsizing. Or all the mess and noise of a thorough goring. He came out of the water a stone's throw away, with an openmouthed roar, showing us four gleaming tusks the size of short swords. We poled quickly away. My pulse didn't settle down for eleven minutes. That afternoon we negotiated past eleven more hippos on the way to camp. Three the next day and twenty the next. Each of them filed into a database with an exact time, a latitude and longitude, and a clear uptick in the speed of my heart.

Back in New York, we kept working. When our little office on the Bowery got too crowded, we moved into a bright space in an old telephone company building in downtown Brooklyn. Before the new OCR opened, I paid a sign maker to hand letter a Pynchon quotation in deep black letters around the whole office:

She looked down a slope, needing to squint for the sunlight, onto a vast sprawl of houses which had grown up all together, like a well-tended crop, from the dull brown earth; and she thought of the time she'd opened a transistor radio to replace a battery and seen her first printed circuit. The ordered swirl of houses and streets, from this high angle, sprang at her now with the same unexpected, astonishing clarity as the circuit card had. Though she knew even less about radios than about Southern Californians, there were to both outward patterns a hieroglyphic sense of concealed meaning, of an intent to communicate. There'd seemed no limit to what the printed circuit could have told her (if she had tried to find out); so in her first minute of San Narciso, a revelation also trembled just past the threshold of her understanding.

As the team grew, so did our commitment to doing work that wasn't tangled up in advertising, in the selling of more stuff to more people. If an email arrived with the word "branding" in it, it'd go straight into the trash. We said no to Google, no to Facebook. Yes to the epidemiologist with no budget. Yes to the community art center in St. Louis. For every project we did that paid us, we'd do two more in the service of our own curiosities and convictions. More farming. People were often confused about whether the OCR was a design studio, or an R&D lab, or a nonprofit. We were somewhere in between. When people asked me what kind of business I ran, I'd say the OCR was a "not-for-enough-profit." Which was a joke, until it wasn't.

After the studio closed in 2017, I spent eighteen months at the Library of Congress, this nation's house of data. I waded, waist-deep, into its millions of books and manuscripts, maps and photographs and recordings. Again I dug, into the library's infrastructures and file formats, its hallways and its card catalogs and its open APIs. I planted ideas into neat rows, watered them, tended to them. Through toolmaking and performance and storytelling, I learned how data at the library works and how it might work differently if we freed it—and ourselves—from technology's incessant expectations and constraints. While I was there, I started writing.

This book is a record. It is data. It's data about these last ten years of my life, a document of the work that I've done to expand my own ideas about what data is and what data can be. It maps, in more detail and with less linearity, my path from that basement studio in Vancouver to the *Times* to the Library of Congress. From a submarine at the bottom of the Gulf of Mexico to the rocking boat in the middle of Africa's Okavango Delta to a windy rock face in the Canadian Rockies. It follows my work as I've explored new methods for showing and exploring data, from visualization to sound and sculpture and performance. Importantly, it also tracks the changes in how I've thought about data, first as a kind of inert fuel for investigation and then as something much more ominous. How we might navigate the risks it

presents to ourselves and others. How we might look directly at its harms without breaking our gaze. How, with some bold revisions, data might offer a rich and fractal medium for personal and community growth.

Mostly, though, this book is a guide. It's a guide for those who have to live in data and for those who want to create data worlds that are more livable. It's a guide for the person who woke up one day, thinking about how they are being tracked moment by moment by their phone and their social media platforms and their cars and their cities, and thought to themselves, how did we get here? It's a guide for the entrepreneur who is setting out with a new company, looking back at the havoc that big data has wrought on our selves and our society, and thinking, we can do better than this. It's a guide for individuals and communities who are looking to speak their own data stories louder than those that are being told about them by others.

Because we can do better than this. We can create new data worlds that put humans first.



On a cloudy September morning in 2013, the students at Hunter College High School filed past security and into the hallways, to be faced with a much different kind of trauma: their school had been named the saddest place in Manhattan.

Five months earlier, researchers in Cambridge, Massachusetts, had pulled more than six hundred thousand tweets from Twitter's public API and fed them through sentiment analysis routines. If a tweet contained words that were deemed sad—maybe “cry” or “frown” or “miserable”—an emotional mark for sadness would be placed on a map of the city. As more tweets classified as sad happened near a specific area, more marks would be put on the map, and more sadness would be ascribed to that particular place. The result was a kind of sedimentary layer of emotion data on top of New York City. If you looked at the sadness map that came out of the study, you'd probably ask about a deep purple spot of lament just to the right of Central Park's reservoir. If you looked that spot up in Google Maps, you'd find that it sat right on top of Hunter College High. If you then thought about when the tweets had been collected (in the spring), you might arrive at a hypothesis similar to that of the president of the New England Complex Systems Institute (NECSI), Yaneer Bar-Yam:

I checked the high school calendar and found that the spring vacation period in 2012 was April 9–13, so that students would be returning to school on the 16th, just during the period of the data collection, April 13–26, 2012. This provided a rationale for the low sentiment there.

Sad students returning from spring break—it seemed to Bar-Yam like an

interesting finding. The press agreed. Stories about the study, and about this sad Upper East Side school, appeared in *Science*, then in *The New York Times*. In advance of its piece, the city's newspaper of record dispatched a reporter to talk to the students at this "saddest spot in Manhattan" to gauge their reaction to the study. "I mean, I can see why it could make sense," one fourteen-year-old student told the reporter. "The school has no windows, so being inside can seem dark and depressing. And some kids do get stressed out from the workload."

Unwittingly, the staff and students at Hunter College High had found themselves inside a microcosm of the data world that by 2012 we were already very much inhabiting. It's a world in which we are all being data-fied from a distance, where our movements and conversations are processed into product recommendations and sociology papers and watch lists, where average citizens don't know the role they are playing, surrounded by machinery built by and for others.

Three decades ago, Ursula Franklin, a Canadian metallurgist, physicist, author, and ardent pacifist, delivered a series of lectures about what she termed the "real world of technology." This place, in which Franklin believed by 1989 we were already firmly entrenched, was one where "most people live and work under conditions that are not structured for their well-being." One defining condition of Franklin's real world of technology is that the authority of data overrides that of lived experience. It's a place where, as she writes, "abstract knowledge is forcing people to perceive their experience as being unreal or wrong." A place where students might accept that they are sad because the data told them so.

It's very easy, as the Cambridge scientists found out, to get carried away with the long-distance magic of APIs and machine learning, to use these technologies to scan from afar. But our scanning isn't harmless; our analyses are not without effect. Those high school students that you've classified as sad feel something when they read the results of your study.



Two years after the *Times*'s "saddest school" article was published, I stood in the lobby of Hunter College High School, waiting to speak with Lisa Siegmann, the school's assistant principal. I couldn't get the story out of my head. I didn't at the time have an inkling that I'd be writing this book; somehow I needed to understand how the story had felt to those who had been so publicly data-fied. I was interested in how the *New York Times* article had been received by students and staff and what the aftermath (if any) was of being so publicly labeled as the saddest spot in the city.

Actually, I wanted their reaction to the school being labeled as sad and then *unlabeled* three weeks later. As it turned out, the researchers from New England had made a big mistake. Their geocoding code, the part that turns a place name or an address into a point on a map, was faulty. Hunter High was

not the saddest place in the data set; it was merely sad-adjacent, coincidentally located near a single Twitter user who had been posting a lot of content that was getting labeled as unhappy. If that weren't bad enough, the scientists had missed a deeper error in their sad-student hypothesis that left the premise completely indefensible.

The assistant principal was running late, so I stood by the security desk with a small group of parents and waited. There was a poster on the wall for an upcoming TEDx event that would feature students and teachers; the theme was "The Care of the Future Is Mine." I couldn't help but look at the students who walked by and try to assess their emotional state. When Siegmann arrived, she led me upstairs, past a small lineup of students waiting to see her, and into her office. Siegmann seemed wary of dredging the article up again, but she was candidly direct about how she and the school had reacted.

"Nobody believed the article," she said. "First, this is not a sad school," she added, taking a minute to explain the various activities that the students participate in and the awards that the school and the students had won.

"Second," she said, "no one in this school uses Twitter." She paused for a minute, then laughed. "Twitter is for old people."

As it turns out, Hunter College High didn't permit students to use social media while they're in the building. The administration knew that students were breaking the rules, and they also knew which social platforms they were surreptitiously using: Instagram and Snapchat, mostly, and sometimes Facebook. But not Twitter.

The Hunter College High fiasco is a perfect example of how data can and does fail end to end: a retracted story about a false statement, spit out by a faulty algorithm, feeding on bad data. All balanced on top of an impossible premise. It's also a clear reflection of the kinds of data stories that we've been so eager to believe: where a large data set combined with novel algorithms shows us some secret that we would not otherwise have seen. "To be human," as Richard Powers wrote, "is to confuse a satisfying story with a meaningful one." Stories told through data can be very satisfying indeed.

What brought me to the assistant principal's office that morning, though, was not to find another way to critique a study or to place blame on the researchers or their broken algorithms. Though I didn't know it at the time, reading that *New York Times* story in 2013 had set a fracture into my thinking. Until then I'd given little thought to the downstream effects of working with data, to the idea that the kind of work I myself was making might have real-world consequences

At first our work at the OCR was centered on building new ways to explore and tell stories with data. We visualized language patterns across 140 years of *Popular Science*, made a touch-screen interface to sift through terabytes of botnet activity, plotted the import and export of the nation's fruits and vegetables. Slowly, though, the Hunter College High fracture found its way into our work, and we began to think more about how data was

tangling itself into the quotidian, how it was affecting the way people worked and lived. Our projects—many of which you’ll read about in this book—made their way out of browser windows and art galleries and into public spaces. They became less about finding answers in data and more about finding agency, less about exploration and more about empowerment.



In writing this, I’ve realized there was another reason I was there in that office. I was there because I can remember so clearly what it was like to be in high school, to be vulnerable and afraid and powerless. How being labeled could feel like being struck. How nothing seemed to be under my control, and no one seemed to hear my voice. I can also remember how I found agency in the face of all that overwhelming possibility through, of all things, computer programming. How could it be that the very same thing that offered me escape three decades ago was now being used to make the lives of high school students worse?



Eighty years before and one borough over, a young woman named Helen Hall Jennings sat in the back of the seventh-grade classroom at P.S. 181. The school was a mile from Prospect Park and was set into the tree-lined streets of East Flatbush, which since World War I had been home to large families of recent immigrants. The 1930 census counts significant numbers of people living near the school who were born in Russia, Syria, Ireland, Scotland, Czechoslovakia, and Austria. This part of Brooklyn was then one of wide boulevards, serviced by busy trolley lines. Maps drawn by fire insurance companies in the late 1920s show narrow streets of single-family houses with modest backyards. P.S. 181 was built just after the turn of the century on New York Avenue between Tilden and Snyder, a sturdy four-story brick building that still stands today.

Jennings observed the classrooms at first and then asked each student a question: Given the choice, who else in the classroom would you sit beside? Back in Manhattan, she worked with her collaborator J. L. Moreno to assemble diagrammatic maps of the students and their chosen relationships. These maps, which Moreno called “sociograms,” show, fifty-three years before Mark Zuckerberg would be born, a language of social networks. Indeed much of the visual language that we find in social network renderings today can be seen in these 1931 drawings: people are shown as nodes (in Jennings’s case, triangles for boys and circles for girls) with lines connecting them into a kind of structured web. If you close your eyes and imagine a network, you might very well be seeing something like these sociograms, nearly a century old.

Looking at all of the P.S. 181 sociograms together (Jennings recorded data

MCD
Farrar, Straus and Giroux
120 Broadway, New York 10271

Copyright © 2021 by Jer Thorp
All rights reserved
First edition, 2021

Grateful acknowledgment is made for permission to reprint the photograph of *Infinite Weft* by Eric Wolfe,
© snapwolf.com.

Ebook ISBN: 978-0-374-72051-3

Our ebooks may be purchased in bulk for promotional, educational, or business use. Please contact the
Macmillan Corporate and Premium Sales Department at 1-800-221-7945, extension 5442, or by email at
MacmillanSpecialMarkets@macmillan.com.

www.mcdbooks.com • www.fsgbooks.com
Follow us on Twitter, Facebook, and Instagram at @mcdbooks