EVERYBODY LIES

BIG DATA, NEW DATA, AND WHAT THE INTERNET CAN TELL US ABOUT WHO WE REALLY ARE

SETH STEPHENS-DAVIDOWITZ

FOREWORD BY STEVEN PINKER

CONTENTS

INTRODUCTION

THE OUTLINES OF A REVOLUTION

Surely he would lose, they said.

In the 2016 Republican primaries, polling experts concluded that Donald Trump didn't stand a chance. After all, Trump had insulted a variety of minority groups. The polls and their interpreters told us few Americans approved of such outrages.

Most polling experts at the time thought that Trump would lose in the general election. Too many likely voters said they were put off by his manner and views.

But there were actually some clues that Trump might actually win both the primaries and the general election—on the internet.

I am an internet data expert. Every day, I track the digital trails that people leave as they make their way across the web. From the buttons or keys we click or tap, I try to understand what we really want, what we will really do, and who we really are. Let me explain how I got started on this unusual path.

The story begins—and this seems like ages ago—with the 2008 presidential election and a long-debated question in social science: How significant is racial prejudice in America?

Barack Obama was running as the first African-American presidential nominee of a major party. He won—rather easily. And the polls suggested that race was not a factor in how Americans voted. Gallup, for example, conducted numerous polls before and after Obama's first election. Their conclusion? American voters largely did not care that Barack Obama was black. Shortly after the election, two well-known professors at the University of California, Berkeley pored through other survey-based data, using more sophisticated data-mining techniques. They reached a similar conclusion.

And so, during Obama's presidency, this became the conventional wisdom in many parts of the media and in large swaths of the academy. The sources that the media and social scientists have used for eighty-plus years to understand the world told us that the overwhelming majority of Americans did not care that Obama was black when judging whether he should be their president.

This country, long soiled by slavery and Jim Crow laws, seemed finally to have stopped judging people by the color of their skin. This seemed to suggest that racism was on its last legs in America. In fact, some pundits even declared that we lived in a post-racial society.

In 2012, I was a graduate student in economics, lost in life, burnt-out in my field, and confident, even cocky, that I had a pretty good understanding of how the world worked, of what people thought and cared about in the twenty-first century. And when it came to this issue of prejudice, I allowed myself to believe, based on everything I had read in psychology and political science, that explicit racism was limited to a small percentage of Americans—the majority of them conservative Republicans, most of them living in the deep South.

Then, I found Google Trends.

Google Trends, a tool that was released with little fanfare in 2009, tells users how frequently any word or phrase has been searched in different locations at different times. It was advertised as a fun tool—perhaps enabling friends to discuss which celebrity was most popular or what fashion was suddenly hot. The earliest versions included a playful admonishment that people "wouldn't want to write your PhD dissertation" with the data, which immediately motivated me to write my dissertation with it.*

At the time, Google search data didn't seem to be a proper source of information for "serious" academic research. Unlike surveys, Google search data wasn't created as a way to help us understand the human psyche. Google was invented so that people could learn about the world, not so researchers could learn about people. But it turns out the trails we leave as we seek knowledge on the internet are tremendously revealing.

In other words, people's search for information is, in itself, information. When and where they search for facts, quotes, jokes, places, persons, things, or help, it turns out, can tell us a lot more about what they really think, really desire, really fear, and really do than anyone might have guessed. This is especially true since people sometimes don't so much query Google as confide in it: "I hate my boss." "I am drunk." "My dad hit me."

The everyday act of typing a word or phrase into a compact, rectangular white box leaves a small trace of truth that, when multiplied by millions, eventually reveals profound realities. The first word I typed in Google Trends was "God." I learned that the states that make the most Google searches mentioning "God" were Alabama, Mississippi, and Arkansas—the Bible Belt. And those searches are most frequently on Sundays. None of which was surprising, but it was intriguing that search data could reveal such a clear pattern. I tried "Knicks," which it turns out is Googled most in New York City. Another no-brainer. Then I typed in my name. "We're sorry," Google Trends informed me. "There is not enough search volume" to show these results. Google Trends, I learned, will provide data only when lots of people make the same search.

But the power of Google searches is not that they can tell us that God is popular down South, the Knicks are popular in New York City, or that I'm not popular anywhere. Any survey could tell you that. The power in Google data is that people tell the giant search engine things they might not tell anyone else.

Take, for example, sex (a subject I will investigate in much greater detail later in this book). Surveys cannot be trusted to tell us the truth about our sex lives. I analyzed data from the General Social Survey, which is considered one of the most influential and authoritative sources for information on Americans' behaviors. According to that survey, when it comes to heterosexual sex, women say they have sex, on average, fifty-five times per year, using a condom 16 percent of the time. This adds up to about 1.1 billion condoms used per year. But heterosexual men say they use 1.6 billion condoms every year. Those numbers, by definition, would have to be the same. So who is telling the truth, men or women?

Neither, it turns out. According to Nielsen, the global information and measurement company that tracks consumer behavior, fewer than 600 million condoms are sold every year. So everyone is lying; the only difference is by how much.

The lying is in fact widespread. Men who have never been married claim to use on average twenty-nine condoms per year. This would add up to more than the total number of condoms sold in the United States to married and single people combined. Married people probably exaggerate how much sex they have, too. On average, married men under sixty-five tell surveys they have sex once a week. Only 1 percent say they have gone the past year without sex. Married women report having a little less sex but not much less.

Google searches give a far less lively—and, I argue, far more accurate—picture of sex during marriage. On Google, the top complaint about a marriage is not having sex. Searches for "sexless marriage" are three and a half times more common than "unhappy marriage" and eight times more common than "loveless marriage." Even unmarried couples complain somewhat frequently about not having sex. Google searches for "sexless relationship" are second only to searches for "abusive relationship." (This data, I should emphasize, is all presented anonymously. Google, of course, does not report data about any particular individual's searches.)

And Google searches presented a picture of America that was strikingly different from that post-racial utopia sketched out by the surveys. I remember when I first typed "nigger" into Google Trends. Call me naïve. But given how toxic the word is, I fully expected this to be a low-volume search. Boy, was I wrong. In the United States, the word "nigger"—or its plural, "niggers"—was included in roughly the same number of searches as the word "migraine(s)," "economist," and "Lakers." I wondered if searches for rap lyrics were skewing the results? Nope. The word used in rap songs is almost always "nigga(s)." So what was the motivation of Americans searching for "nigger"? Frequently, they were looking for jokes mocking African-Americans. In fact, 20 percent of searches with the word "nigger" also included the word "jokes." Other common searches included "stupid niggers" and "I hate niggers."

There were millions of these searches every year. A large number of Americans were, in the privacy of their own homes, making shockingly racist inquiries. The more I researched, the more disturbing the information got.

On Obama's first election night, when most of the commentary focused on praise of Obama and acknowledgment of the historic nature of his election, roughly one in every hundred Google searches that included the word "Obama" also included "kkk" or "nigger(s)." Maybe that doesn't sound so high, but think of the thousands of nonracist reasons to Google this young outsider with a charming family about to take over the world's most powerful job. On election night, searches and sign-ups for Stormfront, a white nationalist site with surprisingly high popularity in the United States, were more than ten times higher than normal. In some states, there were more searches for "nigger president" than "first black president."

There was a darkness and hatred that was hidden from the traditional sources but was quite apparent in the searches that people made.

Those searches are hard to reconcile with a society in which racism is a small factor. In 2012 I knew of Donald J. Trump mostly as a businessman and reality show performer. I had no more idea than anyone else that he would, four years later, be a serious presidential candidate. But those ugly searches are not hard to reconcile with the success of a candidate who—in his attacks on immigrants, in his angers and resentments—often played to people's worst inclinations.

The Google searches also told us that much of what we thought about the location of racism was wrong.

Surveys and conventional wisdom placed modern racism predominantly in the South and mostly among Republicans. But the places with the highest racist search rates included upstate New York, western Pennsylvania, eastern Ohio, industrial Michigan and rural Illinois, along with West Virginia, southern Louisiana, and Mississippi. The true divide, Google search data suggested, was not South versus North; it was East versus West. You don't get this sort of thing much west of the Mississippi. And racism was not limited to Republicans. In fact, racist searches were no higher in places with a high percentage of Republicans than in places with a high percentage of Democrats. Google searches, in other words, helped draw a new map of racism in the United States—and this map looked very different from what you may have guessed. Republicans in the South may be more likely to admit to racism. But plenty of Democrats in the North have similar attitudes.

Four years later, this map would prove quite significant in explaining the political success of Trump.

In 2012, I was using this map of racism I had developed using Google searches to reevaluate exactly the role that Obama's race played. The data was clear. In parts of the country with a high number of racist searches, Obama did substantially worse than John Kerry, the white Democratic presidential candidate, had four years earlier. The relationship was not explained by any other factor about these areas, including education levels, age, church attendance, or gun ownership. Racist searches did not predict poor performance for any other Democratic candidate. Only for Obama.

And the results implied a large effect. Obama lost roughly 4 percentage points nationwide just from explicit racism. This was far higher than might have been expected based on any surveys. Barack Obama, of course, was elected and reelected president, helped by some very favorable conditions for Democrats, but he had to overcome quite a bit more than anyone who was relying on traditional data sources—and that was just about everyone—had realized. There were enough racists to help win a primary or tip a general election in a year not so favorable to Democrats.

My study was initially rejected by five academic journals. Many of the peer reviewers, if you will forgive a little disgruntlement, said that it was impossible to believe that so many Americans harbored such vicious racism. This simply did not fit what people had been saying. Besides, Google searches seemed like such a bizarre dataset.

Now that we have witnessed the inauguration of President Donald J. Trump, my finding seems more plausible.

The more I have studied, the more I have learned that Google has lots of information that is missed by the polls that can be helpful in understanding—among many, many other subjects—an election.

There is information on who will actually turn out to vote. More than half of citizens who don't vote tell surveys immediately before an election that they intend to, skewing our estimation of turnout, whereas Google searches for "how to vote" or "where to vote" weeks before an election can accurately predict which parts of the country are going to have a big showing at the polls.

There might even be information on who they will vote for. Can we really predict which candidate people will vote for just based on what they search? Clearly, we can't just study which candidates are searched for most frequently. Many people search for a candidate because they love him. A similar number of people search for a candidate because they hate him. That said, Stuart Gabriel, a professor of finance at the University of California, Los Angeles, and I have found a surprising clue about which way people are planning to vote. A large percentage of election-related searches contain queries with both candidates' names. During the 2016 election between Trump and Hillary Clinton, some people searched for "Trump Clinton polls." Others looked for highlights from the "Clinton Trump debate." In fact, 12 percent of search queries with "Trump" also included the word "Clinton." More than one-quarter of search queries with "Clinton" also included the word "Trump."

We have found that these seemingly neutral searches may actually give us some clues to which candidate a person supports.

How? The order in which the candidates appear. Our research suggests that a person is significantly more likely to put the candidate they support first in a search that includes both candidates' names.

In the previous three elections, the candidate who appeared first in more searches received the most votes. More interesting, the order the candidates were searched was predictive of which way a particular state would go.

The order in which candidates are searched also seems to contain information that the polls can miss. In the 2012 election between Obama and Republican Mitt Romney, Nate Silver, the virtuoso statistician and journalist, accurately predicted the result in all fifty states. However, we found that in states that listed Romney before Obama in searches most frequently, Romney actually did better than Silver had predicted. In states that most frequently listed Obama before Romney, Obama did better than Silver had predicted.

This indicator could contain information that polls miss because voters are either lying to themselves or uncomfortable revealing their true preferences to pollsters. Perhaps if they claimed that they were undecided in 2012, but were consistently searching for "Romney Obama polls," "Romney Obama debate," and "Romney Obama election," they were planning to vote for Romney all along.

So did Google predict Trump? Well, we still have a lot of work to do—and I'll have to be joined by lots more researchers—before we know how best to use Google data to predict election results. This is a new science, and we only have a few elections for which this data exists. I am certainly not saying we are at the point—or ever will be at the point—where we can throw out public opinion polls completely as a tool for helping us predict elections.

But there were definitely portents, at many points, on the internet that Trump might do better than the polls were predicting.

During the general election, there were clues that the electorate might be a favorable one for Trump. Black Americans told polls they would turn out in large numbers to oppose Trump. But Google searches for information on voting in heavily black areas were way down. On election day, Clinton would be hurt by low black turnout.

There were even signs that supposedly undecided voters were going Trump's way. Gabriel and I found that there were more searches for "Trump Clinton" than "Clinton Trump" in key states in the Midwest that Clinton was expected to win. Indeed, Trump owed his election to the fact that he sharply outperformed his polls there.

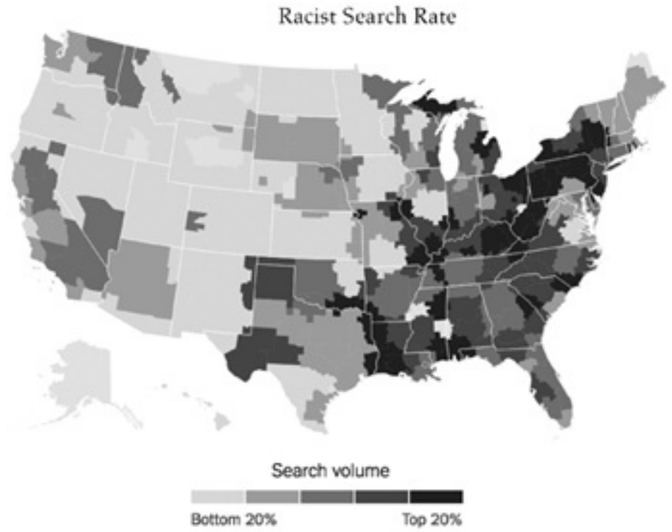
But the major clue, I would argue, that Trump might prove a successful candidate—in the primaries, to begin with—was all that secret racism that my Obama study had uncovered. The Google searches revealed a darkness and hatred among a meaningful number of Americans that pundits, for many years, missed. Search data revealed that we lived in a very different society from the one academics and journalists, relying on polls, thought that we lived in. It revealed a nasty, scary, and widespread rage that was waiting for a candidate to give voice to it.

People frequently lie—to themselves and to others. In 2008, Americans told surveys that they no longer cared about race. Eight years later, they elected as president Donald J. Trump, a man who retweeted a false claim that black people are responsible for the majority of murders of white Americans, defended his supporters for roughing up a Black Lives Matters protester at one of his rallies, and hesitated in repudiating support from a former leader of the Ku Klux Klan. The same hidden racism that hurt Barack Obama helped Donald Trump.

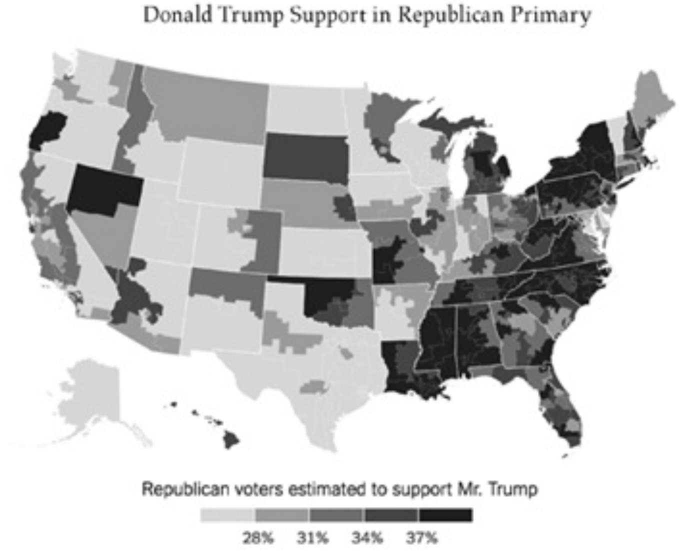
Early in the primaries, Nate Silver famously claimed that there was virtually no chance that Trump would win. As the primaries progressed and it became increasingly clear that Trump had widespread support, Silver decided to look at the data to see if he could understand what was going on. How could Trump possibly be doing so well?

Silver noticed that the areas where Trump performed best made for an odd map. Trump performed well in parts of the Northeast and industrial Midwest, as well as the South. He performed notably worse out West. Silver looked for variables to try to explain this map. Was it unemployment? Was it religion? Was it gun ownership? Was it rates of immigration? Was it opposition to Obama?

Silver found that the single factor that best correlated with Donald Trump’s support in the Republican primaries was that measure I had discovered four years earlier. Areas that supported Trump in the largest numbers were those that made the most Google searches for “nigger.”



Racist Search Rate



Donald Trump Support in Republican Primary

I have spent just about every day of the past four years analyzing Google data. This included a stint as a data scientist at Google, which hired me after learning about my racism research. And I continue to explore this data as an opinion writer and data journalist for the *New York Times*. The revelations have kept coming. Mental illness; human sexuality; child abuse; abortion; advertising; religion; health. Not exactly small topics, and this dataset, which didn't exist a couple of decades ago, offered surprising new perspectives on all of them. Economists and other social scientists are always hunting for new sources of data, so let me be blunt: I am now convinced that Google searches are the most important dataset ever collected on the human psyche.

This dataset, however, is not the only tool the internet has delivered for understanding our world. I soon realized there are other digital gold mines as well. I downloaded all of Wikipedia, pored through Facebook profiles, and scraped Stormfront. In addition, PornHub, one of the largest pornographic sites on the internet, gave me its complete data on the searches and video views of anonymous people around the world. In other words, I have taken a very deep dive into what is now called Big Data. Further, I have interviewed dozens of others—academics, data journalists, and entrepreneurs—who are also exploring these new realms. Many of their studies will be discussed here.

But first, a confession: I am not going to give a precise definition of what Big Data is. Why? Because it's an inherently vague concept. How big is big? Are 18,462 observations Small Data and 18,463 observations Big Data? I prefer to take an inclusive view of what qualifies: while most of the data I fiddle with is from the internet, I will discuss other sources, too. We are living through an explosion in the amount and quality of all kinds of available information. Much of the new information flows from Google and social media. Some of it is a product of digitization of information that was previously hidden away in cabinets and files. Some of it is from increased resources devoted to market research. Some of the studies discussed in this book don't use huge datasets at all but instead just employ a new and creative approach to data—approaches that are crucial in an era overflowing with information.

So why exactly is Big Data so powerful? Think of all the information that is scattered online on a given day—we have a number, in fact, for just how much information there is. On an average day in the early part of the twenty-first century, human beings generate 2.5 million trillion bytes of data.

And these bytes are clues.

A woman is bored on a Thursday afternoon. She Googles for some more "funny clean jokes." She checks her email. She signs on to Twitter. She Googles "nigger jokes."

A man is feeling blue. He Googles for "depression symptoms" and "depression stories." He plays a game of solitaire.

A woman sees the announcement of her friend getting engaged on Facebook. The woman, who is single, blocks the friend.

A man takes a break from Googling about the NFL and rap music to ask the search engine a question: "Is it normal to have dreams about kissing men?"

A woman clicks on a BuzzFeed story showing the "15 cutest cats."

A man sees the same story about cats. But on his screen it is called "15 most adorable cats." He doesn't click.

A woman Googles "Is my son a genius?"

A man Googles "how to get my daughter to lose weight."

A woman is on a vacation with her six best female friends. All her friends keep saying how much fun they're having. She sneaks off to Google "loneliness when away from husband."

A man, the previous woman's husband, is on a vacation with his six best male friends. He sneaks off to Google to type "signs your wife is cheating."

Some of this data will include information that would otherwise never be admitted to anybody. If we aggregate it all, keep it anonymous to make sure we never know about the fears, desires, and behaviors of any specific individuals, and add some data science, we start to get a new look at human beings—their behaviors, their desires, their natures. In fact, at the risk of sounding grandiose, I have come to believe that the new data increasingly available in our digital age will radically expand our understanding of humankind. The microscope showed us there is more to a drop of pond water than we think we see. The telescope showed us there is more to the night sky than we think we see. And new, digital data now shows us there is more to human society than we think we see. It may be our era's microscope or telescope—making possible important, even revolutionary insights.

There is another risk in making such declarations—not just sounding grandiose but also trendy. Many people have been making big claims about the power of Big Data. But they have been short on evidence.

This has inspired Big Data skeptics, of whom there are also many, to dismiss the search for bigger datasets. "I am not saying here that there is no information in Big Data," essayist and statistician Nassim Taleb has written. "There is plenty of information. The problem—the central issue—is that the needle comes in an increasingly larger haystack."

One of the primary goals of this book, then, is to provide the missing evidence of what can be done with Big Data—how we can find the needles, if you will, in those larger and larger haystacks. I hope to provide enough examples of Big Data offering new insights into human psychology and behavior so that you will begin to see the outlines of something truly revolutionary.

"Hold on, Seth," you might be saying right about now. "You're promising a revolution. You're waxing poetic about these big, new datasets. But thus far, you have used all of this amazing, remarkable, breathtaking, groundbreaking data to tell me basically two things: there are plenty of racists in America, and people, particularly men, exaggerate how much sex they have."

I admit sometimes the new data does just confirm the obvious. If you think these findings were obvious, wait until you get to Chapter 4, where I show you clear, unimpeachable evidence from Google searches that men have tremendous concern and insecurity around—wait for it—their penis size.

There is, I would claim, some value in proving things you may have already suspected but had otherwise little evidence for. Suspecting something is one thing. Proving it is another. But if all Big Data could do is confirm your suspicions, it would not be revolutionary. Thankfully, Big Data can do a lot more than that. Time and again, data shows me the world works in precisely the opposite way as I would have guessed. Here are some examples you might find more surprising.

You might think that a major cause of racism is economic insecurity and vulnerability. You might naturally suspect, then, that when people lose their jobs, racism increases. But, actually, neither racist searches nor membership in Stormfront rises when unemployment does.

You might think that anxiety is highest in overeducated big cities. The urban neurotic is a famous stereotype. But Google searches reflecting anxiety—such as "anxiety symptoms" or "anxiety help"—tend to be higher in places with lower levels of education, lower median incomes, and where a larger portion of the population lives in rural areas. There are higher search rates for anxiety in rural, upstate New York than New York City.

You might think that a terrorist attack that kills dozens or hundreds of people would automatically be followed by massive, widespread anxiety. Terrorism, by definition, is supposed to instill a sense of terror. I looked at Google searches reflecting anxiety. I tested how much these searches rose in a country in the days, weeks, and months following every major European or American terrorist attack since 2004. So, on average, how much did anxiety-related searches rise? They didn't. At all.

You might think that people search for jokes more often when they are sad. Many of history's greatest thinkers have claimed that we turn to humor as a release from pain. Humor has long been thought of as a way to cope with the frustrations, the pain, the inevitable disappointments of life. As Charlie Chaplin put it, "Laughter is the tonic, the relief, the surcease from pain."

However, searches for jokes are lowest on Mondays, the day when people report they are most unhappy. They are lowest on cloudy and rainy days. And they plummet after a major tragedy, such as when two bombs killed three and injured hundreds during the 2013 Boston Marathon. People are actually more likely to seek out jokes when things are going well in life than when they aren't.

Sometimes a new dataset reveals a behavior, desire, or concern that I would have never even considered. There are numerous sexual proclivities that fall into this category. For example, did you know that in India the number one search beginning "my husband wants . . ." is "my husband wants me to breastfeed him”? This comment is far more common in India than in other countries. Moreover, porn searches for depictions of women breastfeeding men are four times higher in India and Bangladesh than in any other country in the world. I certainly never would have suspected that before I saw the data.

Further, while the fact that men are obsessed with their penis size may not be too surprising, the biggest bodily insecurity for women, as expressed on Google, is surprising indeed. Based on this new data, the female equivalent of worrying about the size of your penis may be—pausing to build suspense—worrying about whether your vagina smells. Women make nearly as many searches expressing concern about their genitals as men do worrying about theirs. And the top concern women express is its odor—and how they might improve it. I certainly didn’t know that before I saw the data.

Sometimes new data reveals cultural differences I had never even contemplated. One example: the very different ways that men around the world respond to their wives being pregnant. In Mexico, the top searches about “my pregnant wife” include “frases de amor para mi esposa embarazada” (words of love to my pregnant wife) and “poemas para mi esposa embarazada” (poems for my pregnant wife). In the United States, the top searches include “my wife is pregnant now what” and “my wife is pregnant what do I do.”

But this book is more than a collection of odd facts or one-off studies, though there will be plenty of those. Because these methodologies are so new and are only going to get more powerful, I will lay out some ideas on how they work and what makes them groundbreaking. I will also acknowledge Big Data’s limitations.

Some of the enthusiasm for the data revolution’s potential has been misplaced. Most of those enamored with Big Data gush about how immense these datasets can get. This obsession with dataset size is not new. Before Google, Amazon, and Facebook, before the phrase “Big Data” existed, a conference was held in Dallas, Texas, on “Large and Complex Datasets.” Jerry Friedman, a statistics professor at Stanford who was a colleague of mine when I worked at Google, recalls that 1977 conference. One distinguished statistician would get up to talk. He would explain that he had accumulated an amazing, astonishing five gigabytes of data. The next distinguished statistician would get up to talk. He would begin, “The last speaker had gigabytes. That’s nothing. I’ve got terabytes.” The emphasis of the talk, in other words, was on how much information you could accumulate, not what you hoped to do with it, or what questions you planned to answer. “I found it amusing, at the time,” Friedman says, that “the thing that you were supposed to be impressed with was how large their dataset is. It still happens.”

Too many data scientists today are accumulating massive sets of data and telling us very little of importance—e.g., that the Knicks are popular in New York. Too many businesses are drowning in data. They have lots of terabytes but few major insights. The size of a dataset, I believe, is frequently overrated. There is a subtle, but important, explanation for this. The bigger an effect, the fewer the number of observations necessary to see it. You only need to touch a hot stove once to realize that it’s dangerous. You may need to drink coffee thousands of times to determine whether it tends to give you a headache. Which lesson is more important? Clearly, the hot stove, which, because of the intensity of its impact, shows up so quickly, with so little data.

In fact, the smartest Big Data companies are often cutting down their data. At Google, major decisions are based on only a tiny sampling of all their data. You don’t always need a ton of data to find important insights. You need the right data. A major reason that Google searches are so valuable is not that there are so many of them; it is that people are so honest in them. People lie to friends, lovers, doctors, surveys, and themselves. But on Google they might share embarrassing information, about, among other things, their sexless marriages, their mental health issues, their insecurities, and their animosity toward black people.

Most important, to squeeze insights out of Big Data, you have to ask the right questions. Just as you can't point a telescope randomly at the night sky and have it discover Pluto for you, you can't download a whole bunch of data and have it discover the secrets of human nature for you. You must look in promising places—Google searches that begin "my husband wants . . ." in India, for example.

This book is going to show how Big Data is best used and explain in detail why it can be so powerful. And along the way, you'll also learn about what I and others have already discovered with it, including:

- How many men are gay?
- Does advertising work?
- Why was American Pharoah a great racehorse?
- Is the media biased?
- Are Freudian slips real?
- Who cheats on their taxes?
- Does it matter where you go to college?
- Can you beat the stock market?
- What's the best place to raise kids?
- What makes a story go viral?
- What should you talk about on a first date if you want a second?

. . . and much, much more.

But before we get to all that, we need to discuss a more basic question: why do we need data at all? And for that, I am going to introduce my grandmother.