Short Contents

Contents

CHAPTER 1

Thinking Clearly in a Data-Driven Age

What You'll Learn

- Learning to think clearly and conceptually about quantitative information is important for lots of reasons, even if you have no interest in a career as a data analyst.
- Even well-trained people often make crucial errors with data.
- Thinking and data are complements, not substitutes.
- The skills you learn in this book will help you use evidence to make better decisions in your personal and professional life and be a more thoughtful and well-informed citizen.

Introduction

We live in a data-driven age. According to former Google CEO Eric Schmidt, the contemporary world creates as much new data every two days as had been created from the beginning of time through the year 2003. All this information is supposed to have the power to improve our lives, but to harness this power we must learn to think clearly about our data-driven world. Clear thinking is hard—especially when mixed up with all the technical details that typically surround data and data analysis.

Thinking clearly in a data-driven age is, first and foremost, about staying focused on ideas and questions. Technicality, though important, should serve those ideas and questions. Unfortunately, the statistics and quantitative reasoning classes in which most people learn about data do exactly the opposite—that is, they focus on technical details. Students learn mathematical formulas, memorize the names of statistical procedures, and start crunching numbers without ever having been asked to think clearly and conceptually about what they are doing or why they are doing it. Such an approach can work for people to whom thinking mathematically comes naturally. But we believe it is counterproductive for the vast majority of us. When technicality pushes students to stop thinking and start memorizing, they miss the forest for the trees. And it's also no fun.

Our focus, by contrast, is on conceptual understanding. What features of the world are you comparing when you analyze data? What questions can different kinds of comparisons answer? Do you have the right question and comparison for the problem you are trying to solve? Why might an answer that sounds convincing actually be misleading? How might you use creative approaches to provide a more informative answer?

It isn't that we don't think the technical details are important. Rather, we believe that technique without conceptual understanding or clear thinking is a recipe for disaster. In our view, once you can think clearly about quantitative analysis, and once you understand why asking careful and precise questions is so important, technique will follow naturally. Moreover, this way is more fun.

In this spirit, we've written this book to require no prior exposure to data analysis, statistics, or quantitative methods. Because we believe conceptual thinking is more important, we've minimized (though certainly not eliminated) technical material in favor of plain-English explanations wherever possible. Our hope is that this book will be used as an introduction and a guide to how to think about and do quantitative analysis. We believe anyone can become a sophisticated consumer (and even producer) of quantitative information. It just takes some patience, perseverance, hard work, and a firm resolve to never allow technicality to be a substitute for clear thinking.

Most people don't become professional quantitative analysts. But whether you do or do not, we are confident you will use the skills you learn in this book in a variety of ways. Many of you will have quantitative analysts working for or with you. And all of you will read studies, news reports, and briefings in which someone tries to convince you of a conclusion using quantitative analyses. This book will equip you with the clear thinking skills necessary to ask the right questions, be skeptical when appropriate, and distinguish between useful and misleading evidence.

Cautionary Tales

To whet your appetite for the hard work ahead, let's start with a few cautionary tales that highlight the importance of thinking clearly in a data-driven age.

Abe's Hasty Diagnosis

Ethan's first child, Abe, was born in July 2006. As a baby, he screamed and cried almost non-stop at night for five months. Abe was otherwise happy and healthy, though a bit on the small side. When he was one year old the family moved to Chicago, without which move, you'd not be reading this book. (That last sentence contains a special kind of claim called a *counterfactual.* Counterfactuals are really important, and you are going to learn all about them in chapter 3.) After noticing that Abe was small for his age and growing more slowly than expected, his pediatrician decided to run some tests.

After some lab work, the doctors were pretty sure Abe had celiac disease—a digestive disease characterized by gluten intolerance. The good news: celiac disease is not life threatening or even terribly serious if properly managed through diet. The bad news: in 2007, the gluten-free dietary options for kids were pretty miserable.

It turns out that Abe actually had two celiac-related blood tests. One came back positive (indicating that he had the disease), the other negative (indicating that he did not have the disease). According to the doctors, the positive test was over 80 percent accurate. "This is a strong diagnosis," they said. The suggested course of action was to put Abe on a gluten-free diet for a couple of months to see if his weight increased. If it did, they could either do a more definitive biopsy or simply keep Abe gluten-free for the rest of his life.

Ethan asked for a look at the report on Abe's bloodwork. The doctors indicated they didn't think that would be useful since Ethan isn't a doctor. This response was neither surprising nor hard to understand. People, especially experts and authority figures, often don't like acknowledging the limits of their knowledge. But Ethan wanted to make the right decision for his son, so he pushed hard for the information. One of the goals of this book is to give you some of the skills and confidence to be your own advocate in this way when using information to make decisions in your life.

Two numbers characterize the effectiveness of any diagnostic test. The first is its false negative rate, which is how frequently the test says a sick person is healthy. The second is its false positive rate, which is how frequently the test says a healthy person is sick. You need to know *both* the false positive rate and the false negative rate to interpret a diagnostic test's results. So Abe's doctors' statement that the positive blood test was 80 percent accurate wasn't very informative. Did that mean it had a 20 percent false negative rate? A 20 percent false positive rate? Do 80 percent of people who test positive have celiac disease?

Fortunately, a quick Google search turned up both the false positive and false negative rates for both of Abe's tests. Here's what Ethan learned. The test on which Abe came up positive for celiac disease has a false negative rate of about 20 percent. That is, if 100 people with celiac disease took the test, about 80 of them would correctly test positive and the other 20 would incorrectly test negative. This fact, we assume, is where the claim of 80 percent accuracy came from. The test, however, has a false positive rate of 50 percent! People who don't have celiac disease are just as likely to test positive as they are to test negative. (This test, it is worth noting, is no longer recommended for diagnosing celiac disease.) In contrast, the test on which Abe came up negative for celiac disease had much lower false negative and false positive rates.

Before getting the test results, a reasonable estimate of the probability of Abe having celiac disease, given his small size, was around 1 in 100. That is, about 1 out of every 100 small kids has celiac disease. Armed with the lab reports and the false positive and false negative rates, Ethan was able to calculate how likely Abe was to have celiac disease given his small size and the test results. Amazingly, the combination of testing positive on an inaccurate test and testing negative on an accurate test actually meant that the evidence suggested that Abe was much *less* likely than 1 in 100 to have celiac disease. In fact, as we will show you in chapter 15, the best estimate of the likelihood of Abe having celiac, given the test results, was about 1 in 1,000. The blood tests that Abe's doctors were sure supported the celiac diagnosis actually strongly supported the opposite conclusion. Abe was almost certain not to have celiac disease.

Ethan called the doctors to explain what he'd learned and to suggest that moving his pasta-obsessed son to a gluten-free diet, perhaps for life, was not the prudent next step. Their response: "A diagnosis like this can be hard to hear." Ethan found a new pediatrician.

Here's the upshot. Abe did not have celiac disease. The kid was just a bit small. Today he is a normal-sized kid with a ravenous appetite. But if his father didn't know how to think about quantitative evidence or lacked the confidence to challenge a mistaken expert, he'd have spent his childhood eating rice cakes. Rice cakes are gross, so he might still be small.

Civil Resistance

As many around the world have experienced, citizens often find themselves in deep disagreement with their government. When things get bad enough, they sometimes decide to organize protests. If you ever find yourself doing such organizing, you will face many important decisions. Perhaps none is more important than whether to build a movement with a non-violent strategy or one open to a strategy involving more violent forms of confrontation. In thinking through this quandry, you will surely want to consult your personal ethics. But you might also want to know what the evidence says about the costs and benefits of each approach. Which kind of organization is most likely to succeed in changing government behavior? Is one or the other approach more likely to land you in prison, the hospital, or the morgue?

There is some quantitative evidence that you might use to inform your decisions. First, comparing anti-government movements across the globe and over time, governments more often make concessions to fully non-violent groups than to groups that use violence. And even comparing across groups that do use violence, governments more frequently make concessions to those groups that engage in violence against military and government targets rather than against civilians. Second, the personal risks associated with violent protest are greater than those associated with non-violent protest. Governments repress violent uprisings more often than they do non-violent protests, making concerns about prison, the hospital, and the morgue more acute.

This evidence sounds quite convincing. A non-violent strategy seems the obvious choice. It is apparently both more effective and less risky. And, indeed, on the basis of this kind of data, political scientists Erica Chenoweth and Evan Perkoski conclude that "planning, training, and preparation to maintain nonviolent discipline is key—especially (and paradoxically) when confronting brutal regimes."

But let's reconsider the evidence. Start by asking yourself, In what kind of a setting is a group likely to engage in non-violent rather than violent protest? A few thoughts occur to us. Perhaps people are more likely to engage in non-violent protest when they face a government that they think is particularly likely to heed the demands of its citizens. Or perhaps people are more likely to engage in non-violent protest when they have broad-based support among their fellow citizens, represent a group in society that can attract media attention, or face a less brutal government.

If any of these things are true, we should worry about the claim that maintaining non-violent discipline is key to building a successful anti-government movement. (Which isn't to say that we are advocating violence.) Let's see why.

Empirical studies find that, on average, governments more frequently make concessions in places that had non-violent, rather than violent, protests. The claimed implication rests on a particular interpretation of that difference—namely, that the higher frequency of government concessions in non-violent places is *caused* by the use of non-violent tactics. Put differently, all else held equal, if a given movement using violent methods had switched to using non-violent methods, the government would have been more likely to grant concessions. But is this causal interpretation really justified by the evidence?

Suppose it's the case that protest movements are more likely to turn to violence when they do not have broad-based support among their fellow citizens. Then, when we compare places that had violent protests to places that had non-violent protests, all else (other than protest tactics) is not held equal. Those places differ in at least two ways. First, they differ in terms of whether they had violent or non-violent protests. Second, they differ in terms of how supportive the public was of the protest movement.

This second difference is a problem for the causal interpretation. You might imagine that public opinion has an independent effect on the government's willingness to grant concessions. That is, all else held equal (including protest tactics), governments might be more willing to grant concessions to protest movements with broad-based public support. If this is the case, then we can't really know whether the fact that governments grant concessions more often to non-violent protest movements than to violent protest movements is because of the difference in protest tactics or because the non-violent movements also happen to be the movements with broad-based public support. This is the classic problem of mistaking correlation for causation.

It is worth noting a few things. First, if government concessions are in fact due to public opinion, then it could be the case that, were we actually able to hold all else equal in our comparison of violent and non-violent protests, we would find the opposite relationship—that is, that non-violence is not more effective than violence (it could even be less effective). Given this kind of evidence, we just can't know.

Second, in this example, the conclusion that appears to follow if you don't force yourself to think clearly is one we would all like to be true. Who among us would not like to live in a world where non-violence is always preferred to violence? But the whole point of using evidence to help us make decisions is to force us to confront the possibility that the world may not be as we believe or hope it is. Indeed, it is in precisely those situations where the evidence seems to say exactly what you would like it to say that it is particularly important to force yourself to think clearly.

Third, we've pointed to one challenge in assessing the effects of peaceful versus violent protest, but there are others. For instance, think about the other empirical claim we discussed: that violent protests are more likely to provoke the government into repressive crack-downs than are non-violent protests. Recall, we suggested that people might be more likely to engage in non-violent protest when they are less angry at their government, perhaps because the government is less brutal. Ask yourself why, if this is true, we have a similar problem of interpretation. Why might the fact that there are more government crack-downs following violent protests than non-violent protests *not* mean that switching from violence to non-violence will reduce the risk of crack-downs? The argument follows a similar logic to the one we just made regarding concessions. If you don't see how the argument works yet, that's okay. You will by the end of chapter 9.

Broken-Windows Policing

In 1982, the criminologist George L. Kelling and the sociologist James Q. Wilson published an article in *The Atlantic* proposing a new theory of crime and policing that had enormous and long-lasting effects on crime policy in the United States and beyond.

Kelling and Wilson's theory is called *broken windows*. It was inspired by a program in Newark, New Jersey, that got police out of their cars and walking a beat. According to Kelling and Wilson, the program reduced crime by elevating "the level of public order." Public order is important, they argue, because its absence sets in motion a vicious cycle:

> A piece of property is abandoned, weeds grow up, a window is smashed. Adults stop scolding rowdy children…Families move out, unattached adults move in. Teenagers gather in front of the corner store. The merchant asks them to move; they refuse. Fights occur. Litter accumulates. People start drinking in front of the grocery…
>
> Residents will think that crime, especially violent crime, is on the rise…They will use the streets less often…Such an area is vulnerable to criminal invasion.

This idea that policing focused on minimizing disorder can reduce violent crime had a big impact on police tactics. Most prominently, the broken-windows theory was the guiding philosophy in New York City in the 1990s. In a 1998 speech, then New York mayor Rudy Giuliani said,

> We have made the "Broken Windows" theory an integral part of our law enforcement strategy...
>
> You concentrate on the little things, and send the clear message that this City cares about maintaining a sense of law and order...then the City as a whole will begin to become safer.

And, indeed, crime in New York city did decline when the police started focusing "on the little things." According to a study by Hope Corman and H. Naci Mocan, misdemeanor arrests increased 70 percent during the 1990s and violent crime decreased by more than 56 percent, double the national average.

To assess the extent to which broken-windows policing was responsible for this fall in crime, Kelling and William Sousa studied the relationship between violent crime and broken-windows approaches across New York City's precincts. If minimizing disorder causes a reduction in violent crime, they argued, then we should expect the largest reductions in crime to have occurred in neighborhoods where the police were most focused on the broken-windows approach. And this is just what they found. In precincts where misdemeanor arrests (the "little things") were higher, violent crime decreased more. They calculated that "the average NYPD precinct...could expect to suffer one less violent crime for approximately every 28 additional misdemeanor arrests."

This sounds pretty convincing. But let's not be too quick to conclude that arresting people for misdemeanors is the answer to ending violent crime. Two other scholars, Bernard Harcourt and Jens Ludwig, encourage us to think a little more clearly about what might be going on in the data.

The issue that Harcourt and Ludwig point out is something called *reversion to the mean* (which we'll talk about a lot more in chapter 8). Here's the basic concern. In any given year, the amount of crime in a precinct is determined by lots of factors, including policing, drugs, the economy, the weather, and so on. Many of those factors are unknown to us. Some of them are fleeting; they come and go across precincts from year to year. As such, in any given precinct, we can think of there being some "baseline" level of crime, with some years randomly having more crime and some years randomly having less (relative to that precinct-specific baseline).

In any given year, if a precinct had a high level of crime (relative to its baseline), then it had bad luck on the unknown and fleeting factors that help cause crime. Probably next year its luck won't be as bad (that's what *fleeting* means), so that precinct will likely have less crime. And if a precinct had a low level of crime (relative to its baseline) this year, then it had good luck on the unknown and fleeting factors, and it will probably have worse luck next year (crime will go back up). Thus, year to year, the crime level in a precinct tends to revert toward the *mean* (i.e., the precinct's baseline level of crime).

Now, imagine a precinct that had a really high level of violent crime in the late 1980s. Two things are likely to be true of that precinct. First, it is probably a precinct with a high baseline of violent crime. Second, it is also probably a precinct that had a bad year or two—that is, for idiosyncratic and fleeting reasons, the level of crime in the late 1980s was high relative to that precinct's baseline. The same, of course, is true in reverse for precincts that had a low level of crime in the late 1980s. They probably have a low baseline of crime, and they also probably had a particularly good couple of years.

Why is this a problem for Kelling and Sousa's conclusions? Because of reversion to the mean, we would expect the most violent precincts in the late 1980s to show a reduction in violent crime on average, even with no change in policing. And unsurprisingly, given the police's objectives, but unfortunately for the study, it was precisely those high-crime precincts in the 1980s that were most likely to get broken-windows policing in the early 1990s. So, when we see a reduction in violent crime in the precincts that had the most broken-windows policing, we don't know if it's the policing strategy or reversion to the mean that's at work.

Harcourt and Ludwig go a step further to try to find more compelling evidence. Roughly speaking, they look at how changes in misdemeanor arrests relate to changes in violent crime in precincts that had similar levels of violent crime in the late 1980s. By comparing precincts with similar starting levels of violent crime, they go some way toward eliminating the problem of reversion to the mean. Surprisingly, this simple change actually flips the relationship! Rather than confirming Kelling and Sousa's finding that misdemeanor arrests are associated with a reduction in violent crime, Harcourt and Ludwig find that precincts that focused more on misdemeanor arrests actually appear to have experienced an *increase* in violent crime. Exactly the opposite of what we would expect if the broken-windows theory is correct.

Now, this reversal doesn't settle the matter on the efficacy of broken-windows policing. The relationship between misdemeanor arrests and violent crime that Harcourt and Ludwig find could be there for lots of reasons other than misdemeanor arrests causing an increase in violent crime. For instance, perhaps the neighborhoods with increasing misdemeanors are becoming less safe in general and would have experienced more violent crime regardless of policing strategies. What these results do show is that the data, properly considered, certainly don't offer the kind of unequivocal confirmation of the broken-windows ideas that you might have thought from Kelling and Sousa's finding. And you can only see this if you have the ability to think clearly about some subtle issues.

This flawed thinking was important. Evidence-based arguments like Kelling and Sousa's played a role in convincing politicians and policy makers that broken-windows policing was the right path forward when, in fact, it might have diverted resources away from preventing and investigating violent crime and may have created a more adversarial and unjust relationship between the police and the disproportionately poor and minority populations who were frequently cited for "the small stuff."

Thinking and Data Are Complements, Not Substitutes

Our quantitative world is full of lots of exciting new data and analytic tools to analyze that data with fancy names like machine learning algorithms, artificial intelligence, random forests, and neural networks. Increasingly, we are even told that this new technology will make it possible for the machines to do the thinking for us. But that isn't right. As our cautionary tales highlight, no data analysis, no matter how futuristic its name, will work if we aren't asking the right questions, if we aren't making the right comparisons, if the underlying assumptions aren't sound, or if the data used aren't appropriate. Just because an argument contains seemingly sophisticated quantitative data analysis, that doesn't mean the argument is rigorous or right. To harness the power of data to make better decisions, we must combine quantitative analysis with clear thinking.

Our stories also illustrate how our intuitions can lead us astray. It takes lots of care and practice to train ourselves to think clearly about evidence. The doctors' intuition that Abe had celiac disease because of a test with 80 percent accuracy and the researchers' intuition that broken-windows policing works because crime decreased in places where it was deployed seem sensible. But both intuitions were wrong, suggesting that we should be skeptical of our initial hunches. The good news is that clear thinking can become intuitive if you work at it.

Data and quantitative tools are not a substitute for clear thinking. In fact, quantitative skills without clear thinking are quite dangerous. We suspect, as you read the coming chapters, you will be jarred by the extent to which unclear thinking affects even the most important decisions people make. Through the course of this book, we will see how misinterpreted information distorts life-and-death medical choices, national and international counterterrorism policies, business and philanthropic decisions made by some of the world's wealthiest people, how we set priorities for our children's education, and a host of other issues from the banal to the profound. Essentially, no aspect of life is immune from critical mistakes in understanding and interpreting quantitative information.

In our experience, this is because unclear thinking about evidence is deeply ingrained in human psychology. Certainly our own intuitions, left unchecked, are frequently subject to basic errors. Our guess is that yours are too. Most disturbingly, the experts on whose advice you depend—be they doctors, business consultants, journalists, teachers, financial advisors, scientists, or what have you—are often just as prone to making such errors as the rest of us. All too often, because they are experts, we trust their judgment without question, and so do they. That is why it is so important to learn to think clearly about quantitative evidence for yourself. That is the only way to know how to ask the right questions that lead you, and those on whose advice you depend, to the most reliable and productive conclusions possible.

How could experts in so many fields make important errors so often? Expertise, in any area, comes from training, practice, and experience. No one expects to become an expert in engineering, finance, plumbing, or medicine without instruction and years of work. But, despite its fundamental and increasing importance for so much of life in our quantitative age, almost no one invests this kind of effort into learning to think clearly with data. And, as we've said, even when they do, they tend to be taught in a way that over-emphasizes the technical and under-emphasizes the conceptual, even though the fundamental problems are almost always about conceptual mistakes in thinking rather than technical mistakes in calculation.

The lack of expertise in thinking presents us with two challenges. First, if so much expert advice and analysis is unreliable, how do you know what to believe? Second, how can you identify those expert opinions that do in fact reflect clear thinking?

This book provides a framework for addressing these challenges. Each of the coming chapters explains and illustrates, through a variety of examples, fundamental principles of clear thinking in a data-driven world. Part 1 establishes some shared language—clarifying what we mean by correlation and causation and what each is useful for. Part 2 discusses how we can tell whether a statistical relationship is genuine. Part 3 discusses how we can tell if that relationship reflects a causal phenomenon or not. And part 4 discusses how we should and shouldn't incorporate quantitative information into our decision making.

Our hope is that reading this book will help you internalize the principles of clear thinking in a deep enough way that they start to become second nature. You will know you are on the right path when you find yourself noticing basic mistakes in how people think and talk about the meaning of evidence everywhere you turn—as you watch the news, peruse magazines, talk to business associates, visit the doctor, listen to the color commentary during athletic competitions, read scientific studies, or participate in school, church, or other communal activities. You will, we suspect, find it difficult to believe how much nonsense you're regularly told by all kinds of experts. When this starts to happen, try to remain humble and constructive in your criticisms. But do feel free to share your copy of this book with those whose arguments you find are in particular need of it. Or better yet, encourage them to buy their own copy!

Readings and References

The essay on non-violent protest by Erica Chenoweth and Evan Perkoski that we quote can be found at https://politicalviolenceataglance.org/2018/05/08/states-are-far-less-likely-to-engage-in-mass-violence-against-nonviolent-uprisings-than-violent-uprisings/.

The following book contains more research on the relationship between non-violence and efficacy:

Erica Chenoweth and Maria J. Stephan. 2011. *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict.* Columbia University Press.

The following articles were discussed in this order on the topic of broken windows policing:

George L. Kelling and James Q. Wilson. 1982. "Broken Windows: The Police and Neighborhood Safety." *The Atlantic*. March https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/.

Archives of Rudolph W. Giuliani. 1998. "The Next Phase of Quality of Life: Creating a More Civil City." February 24. http://www.nyc.gov/html/rwg/html/98a/quality.html.

Hope Corman and H. Naci Mocan. 2005. "Carrots, Sticks, and Broken Windows." *Journal of Law and Economics* 48(1):235–66.

George L. Kelling and William H. Sousa, Jr. 2001. Do Police Matter? An Analysis of the Impact of New York City's Police Reforms. Civic Report for the Center for Civic Innovation at the Manhattan Institute.

Bernard E. Harcourt and Jens Ludwig. 2006. "Broken Windows: New Evidence from New York City and a Five-City Social Experiment." *University of Chicago Law Review* 73:271–320. *The published version has a misprinted sign in the key table. For the correction, see Errata, 74* U. Chi. L. Rev. 407 (2007).

Index

Note: Page numbers in italic type indicate figures or tables.