CHAPTER 4

Correlation Requires Variation

What You'll Learn

- You can't learn about a correlation without variation in both variables of interest.
- In many realms of life—from education to medicine to rocket science—people fall into the trap of trying to make claims about correlations without such variation.
- A particularly common way people fall into this mistake is by *selecting on the dependent variable*, examining only instances when some phenomenon occurred rather than comparing cases where it occurred to cases where it did not.
- Many institutional procedures push us to select on the dependent variable without noticing it.

Introduction

In chapter 2 we discussed the idea that the correlation between two features of the world is the extent to which they tend to occur together. We opened our discussion of correlation by thinking about whether oil production and autocracy are correlated. To figure this out we looked at the country-level data represented in table 4.1.

To determine whether there is a correlation between oil production and autocracy we compared the percentage of major oil producers that are autocracies to the percentage of countries that aren't major oil producers that are autocracies. To make this comparison, we needed four pieces of information: the number of autocracies that are major oil producers, the number of democracies that are major oil producers, the number of autocracies that are not major oil producers, and the number of democracies that are not major oil producers. Had we been lacking any of these pieces of information, we would not have been able to figure out whether oil production and autocracy are correlated.

To see why, suppose we didn't know the number of democracies that are major oil producers. (Of course, we'd also have to not know the total number of countries, so we couldn't just back out the 9 by subtracting the number of countries in the other three categories from the total number of countries.) We still know that about 20 percent ($\frac{29}{147}$) of countries that aren't major oil producers are autocracies. But now we can't figure out

Table 4.1. Oil production and type of government.

	Not Major Oil Producer	Major Oil Producer	Total
Democracy	118	9	127
Autocracy	29	11	40
Total	147	20	167

what proportion of the major oil producers are autocracies. It could be anything. If the number of democracies that are major oil producers turned out to be (say) 11, then 50 percent ($\frac{11}{22}$) of major oil producers would be autocracies and there would be a positive correlation. If the number of democracies that are major oil producers turned out to be (say) 99, then only 10 percent ($\frac{11}{110}$) of major oil producers would be autocracies, so there would be a negative correlation. If the number of democracies that are major oil producers turned out to be 44, then 20 percent ($\frac{11}{55}$) of major oil producers would be autocracies—the same as for countries that are not major oil producers—and there would be no correlation at all. So, just as we saw in our discussion of scandals and congressional representatives in chapter 2, we need to observe all four pieces of information to figure out the correlation.

This is what we mean when we say that correlation requires variation: If you want to figure out whether two variables are correlated, you have to observe variation in both of them. You must observe the number of countries that are and are not major oil producers. And you must observe the number of autocracies and democracies in each group. Just observing variation in one or the other variable is not enough. In chapter 2, when we asked which of five factual statements described a correlation, the problem with the three statements that did not was a lack of variation in one of the variables.

While it may seem obvious, on the basis of our simple binary example, that correlation requires variation, in our experience, it is anything but. Indeed, failing to look for variation in one or another variable while trying to establish a correlation is an exceptionally common mistake.

In this chapter, we explore this mistake and try to unpack why it is so common. Broadly, we think there are two closely related reasons that people so frequently try to establish a correlation without variation. The first reason is called *selecting on the dependent variable.* The second reason is that the world is often organized in ways that push us to make this mistake.

This chapter, more than most in the book, is built around examples. We do this for a reason. We've found that, once we explain that correlation requires variation, people tend to nod their head in agreement, appearing to understand. Indeed, because the point seems obvious when put in plain English, many people are skeptical that this could be such a big problem. And yet, they themselves go right back to making the same mistake. We hope that by showing you lots of examples of very smart people making this mistake in high-stakes environments, we will convince you that this is a real problem and that avoiding this error requires clear thinking, genuine effort, and concentration.

Selecting on the Dependent Variable

If you want to forecast or explain some phenomenon, it is a natural impulse to start by examining previous instances of that phenomenon occurring. This is called *selecting on the dependent variable.* But if you look only at instances when the phenomenon

occurred, you are trying to assess a correlation without variation, since you have no variation in whether or not the phenomenon occurred. This is like looking for correlates of autocracy without examining any democracies. It won't work.

The phrase *dependent variable* refers to the variable representing the phenomenon you are trying to forecast or explain. This mistake is referred to as *selecting on the dependent variable* because you are selecting which cases to look at based on the value of the dependent variable (e.g., only looking at autocracies) rather than looking at variation in the dependent variable (e.g., comparing autocracies and democracies).

Consider a few examples. Following the financial crisis of 2008, both scholars and journalists who wanted to understand how to predict future financial crises invested enormous time and energy examining the historic record to look for patterns in previous crises. Malcolm Gladwell, in his book *Outliers*, tries to understand the correlates of personal success by recounting the lives of highly accomplished people, looking for similarities. Congress, considering a change to American counterinsurgency strategy in Afghanistan, heard testimony on the correlates of suicide terrorism from an academic expert who had done an exhaustive study of all suicide terrorist campaigns since 1980, looking for shared characteristics.

As natural as it seems to look for commonalities in past instances of events you want to forecast, it really is a mistake. Correlation requires variation. Each of the studies just described would have been far more informative if they'd had variation in the dependent variable.

The claim that we can't learn about the correlates of financial crises or suicide terrorism by looking for commonalities among historic cases of similar events may seem counterintuitive. But, since we know that correlation requires variation, the mistake is actually quite simple to grasp. Put in the terms of our earlier example, each of these examples is analogous to looking for correlates of oil production without any data on non-oil-producing countries!

To see the key conceptual flaw in all of these arguments in another way, let's start by considering the central claim in Gladwell's *Outliers*, the so-called *10,000-hour rule*.

The 10,000-Hour Rule

Gladwell's idea is that it takes about 10,000 hours of serious practice to master any difficult skill. Talent might matter too, but first and foremost, if you are looking for a great achiever, look for someone who put in that 10,000 hours of practice.

Now, of course, Gladwell isn't just interested in forecasting great success. He thinks the 10,000-hour rule might be causal. If true, this would have far-reaching consequences. Given enough practice, perhaps any of us could achieve almost anything.

But talk of causality is premature. Before we can think about causality, we need to figure out whether Gladwell's evidence is even compelling for the claim of a correlation between 10,000 hours of practice and great success. So let's start there.

Gladwell asks, "Is the ten-thousand-hour rule a general rule of success?" The answer, he concludes, is yes. The evidence? "If we scratch below the surface of every great achiever" we see the same pattern (p. 47). "Virtually every success story... involves someone or some group working harder than their peers" (p. 239). In case after case, from Bill Gates to the Beatles, Gladwell shows that great achievers put in their 10,000 hours—overwhelming evidence, he concludes, that practice predicts success.

Let's try to think a little more clearly about Gladwell's evidence. What has Gladwell shown us? Of course, he hasn't actually looked at every great achiever. But he's shown us evidence that lots of great achievers practice at least 10,000 hours. The big problem

Table 4.2. Great achievers practice more than 10,000 hours.

	Great Achiever	Not Great Achiever	Total
10,000 Hours of Practice	Many	?	
Not 10,000 Hours of Practice	Very few	?	
Total			

is that he's told us nothing about all the people who aren't great achievers. A table of evidence for *Outliers* would look something like table 4.2.

Even granting that Gladwell is correct that most great achievers put in 10,000 hours of practice, this doesn't tell us whether 10,000 hours of practice is correlated with great success. Correlation requires variation. Because he has selected on the dependent variable, Gladwell's data lack variation in achievement. If you want to know whether putting in 10,000 hours of practice correlates with success, it is not enough to observe that most great achievers put in 10,000 hours of practice. We need to know about the non-achievers' practice habits as well.

Of course, Gladwell's analysis does provide some information that we didn't previously have. Momentarily, let's suppose that Gladwell didn't cherry pick his stories in order to fit his narrative (although, of course he did: he's a storyteller, not a scientist). In this case, we've learned that most highly successful people put in 10,000 hours of practice before achieving great success.

Although this is not enough information to measure a correlation, Gladwell and his defenders might argue that we already have a rough sense that most members of the general public who are not great achievers have not put in 10,000 hours of practice. In that case, maybe Gladwell's analysis significantly shifts our beliefs about the correlation between practice and great success, even if he didn't explicitly measure the correlation. In these cases where we already have a good sense of the prevalence of something in the general population, perhaps it's useful to show that the prevalence is different for a certain group of interest.

Maybe. But we're still skeptical that Gladwell's analysis teaches us much. That's because most people probably *have* devoted at least 10,000 hours of practice to *something*. Anthony has spent 10,000 hours on the golf course, and he's no Tiger Woods. Ethan has spent 10,000 hours playing guitar, and he's no Jimi Hendrix. If you've worked at something full time for five years but you're not the most successful person in your field, then you're one of the many, many people in the top-right cell of table 4.2 that Gladwell never considered.

We should also remember that Gladwell is a gifted storyteller. In the extremely unlikely scenario in which Anthony wins the Masters, Gladwell might write an inspiring and convincing story about how, despite being a full-time college professor, Anthony's many years of practice, failure, and more practice allowed him to pull off the greatest Cinderella story in sports history (just let us dream for a moment). But far more likely, Anthony will happily continue to be one of millions, if not billions, of people who love something, work hard at it, but never achieve immense success and who are never considered in Gladwell's analysis.

To test your understanding, let's see the problem with claims like Gladwell's in another setting. We are going to repeat his exact argument, but in a fictional example that we hope makes the problem even clearer.

Table 4.3. What sick people drank (made-up data).

	Sick	Not Sick	Total
Drank Beverage	500		
Didn't Drink Beverage	0		
Total	500		

Table 4.4. What sick and healthy people drank (made-up data).

	Sick	Not Sick	Total
Drank Beverage	500	9,500	10,000
Didn't Drink Beverage	0	0	0
Total	500	9,500	10,000

Suppose a town of 10,000 people experiences a surprising spate of illness. In the course of a month, 500 people are taken ill with the same symptoms. Local health officials want to determine the cause of the illness. They take case histories of the 500 sick people, looking for commonalities. In the course of this investigation, they find that all 500 people consumed the same beverage, from the same source, the day before they were hospitalized.

Table 4.3 shows data corresponding to our fictionalized story.

The facts about the beverage and the illness correspond exactly to the facts about practice and success from *Outliers*. Everyone who gets sick (succeeds) drank the same beverage (put in 10,000 hours). Surely, then, drinking that beverage (practicing 10,000 hours) is an important predictor of illness (great success). If we want to know who else is likely to get sick, we should survey the town and find out who else drank the same beverage. Right?

Suppose we tell you that the beverage in question is tap water. The claim that the "pattern" of illness suggests a correlation between the beverage and the disease now seems questionable. Why? Because many people consume tap water every day. Indeed, in our fictional town, all 500 people who got sick consumed tap water, but so too did the 9,500 who didn't get sick. As table 4.4 makes clear, there is in fact no correlation between the beverage and getting sick: 100 percent of sick people and 100 percent of healthy people drank the beverage.

The 10,000-hour rule is similarly unsubstantiated by data of the sort presented by Gladwell. Yes, lots of successful people practice very hard. So too do lots of less successful people. Think of all the bands that practiced countless hours, played countless gigs, and did not become the Beatles.

Corrupting the Youth

American kids who liked rock music in the 1980s (ask your parents) may remember the Parents Music Resource Center (PMRC). The PMRC was a lobbying group whose members opposed what they perceived to be the increasingly inappropriate content of rock music. Most famous among the founders of the PMRC was Tipper Gore, wife

of then Senator and later Vice President Al Gore, who started the group after being shocked by the lyrics of a Prince song.

The PMRC claimed that explicit lyrics were corrupting the youth, causing suicide, sexual violence, and even murder. They denounced "porn rock"—a category that included Bruce Springsteen because the song "I'm on Fire" contained a sexual innuendo—and demanded warning labels be placed on albums. In 1985, the Senate Commerce, Science, and Transportation Committee held hearings. Musicians from across the musical spectrum, from the country singer John Denver to Twisted Sister's Dee Snider testified against the PMRC's position. But the PMRC prevailed.

Let's consider a bit of the argument. Here is the testimony of Jeff Ling, a PMRC consultant:

> Many albums today include songs that encourage suicide, violent revenge, sexual violence, and violence just for violence's sake. . . This is Steve Boucher. Steve died while listening to AC/DC's "Shoot to Thrill." Steve fired his father's gun into his mouth. . . A few days ago I was speaking in San Antonio. The day before I arrived, they buried a young high school student. This young man had taken his tape deck to the football field. He hung himself while listening to AC/DC's "Shoot to Thrill." Suicide has become epidemic in our country among teenagers. Some 6,000 will take their lives this year. Many of these young people find encouragement from some rock stars who present death as a positive, almost attractive alternative. . .Of course, AC/DC is no stranger to violent material. . .One of their fans I know you are aware of is the accused Night Stalker.

Ling's argument, which is typical of crusaders against corruption of the youth, amounts to this:

1. Some young people behave regrettably.
2. The youth who behave regrettably all listen to this terrible rock music.
3. The music must be the cause of the regrettable behavior

Of course, talk of causality is again premature. We'll focus on whether such evidence even suggests a correlation.

Thirty years earlier, in 1954, the Senate heard astoundingly similar testimony about that generation's scourge of the youth, comic books. Here is the neurologist and psychiatrist Fredric Wertham testifying before a Senate subcommittee:

> There is a school in a town in New York State where there has been a great deal of stealing. Some time ago some boys attacked another boy and they twisted his arm so viciously that it broke in two places, and, just like in a comic book, the bone came through the skin.
>
> In the same school about 10 days later 7 boys pounced on another boy and pushed his head against the concrete so that the boy was unconscious and had to be taken to the hospital. He had a concussion of the brain.
>
> In this same high school in 1 year 26 girls became pregnant. The score this year, I think, is eight. Maybe it is nine by now.
>
> Now, Mr. Chairman, this is what I call ethical and moral confusion. I don't think that any of these boys or girls individually vary very much. It cannot be explained individually, alone.
>
> Here is a general moral confusion and I think that these girls were seduced mentally long before they were seduced physically, and, of course, all those people

there are very, very great—not all of them, but most of them, are very great comic book readers, have been and are.

This kind of argument persists in the contemporary environment. We have all heard, and perhaps even made, similar claims about the insidious effects of television or video games or social media. For instance, following the horrific shootings at Columbine High School, the U.S. Department of Education and the Secret Service set up a joint task force to determine what factors would allow school officials to anticipate and prevent school violence. The task force studied all thirty-seven incidents of school violence from 1974 through 2000. While concluding that there is no single profile of a school shooter, they also reported the following (among many other things):

1. "Many attackers felt bullied, persecuted, or injured by others prior to the attack."
2. "Most attackers were known to have had difficulty coping with significant losses or personal failures."
3. "Most attackers engaged in some behavior, prior to the incident, that caused others concern or indicated a need for help."
4. "Over half of the attackers demonstrated some interest in violence, through movies, video games, books, and other media."

A similar commission was convened in 2018. While less focused on specific corrupters of the youth, this commission too at times fell into selecting on the dependent variable. For instance, in a chapter recommending increased focus on character education, the commission notes that many school shooters experienced social isolation, without comparing this to levels of social isolation among those who do not engage in violence:

In the aftermath of the Parkland shooting, multiple reports indicated the alleged shooter experienced feelings of isolation and depression in the years leading up to the shooting. . . . Perpetrators of previous school shootings shared that sense of detachment. For example, one Columbine shooter was characterized as depressed and reclusive. . . . Family members and acquaintances of the Virginia Tech shooter said that, as his isolation grew during his senior year, his "attention to schoolwork and class time dropped." . . . The same was true at Sandy Hook.

At times the commission does avoid selecting on the dependent variable. In a chapter on mental health, they write,

Individuals who commit mass shootings may or may not have a serious mental illness (SMI). There is little population-level evidence to support the notion that those diagnosed with mental illness are more likely than anyone else to commit gun crimes.

But not long after, they return to arguments that suggest they are looking for correlation without variation:

A U.S. Department of Education and U.S. Secret Service analysis found that as many as a quarter of individuals who committed mass shootings had been in treatment for mental illnesses. . . Such individuals often feel aggrieved and extremely angry, and nurture fantasies of violent revenge.

These are not the only such government reports; such analyses are seemingly inevitable after acts of youth violence. But, for reasons we've already seen, these findings, like the Senate testimonies above, are misleading. Even if it were true that virtually every young person who behaves in a troubling manner also listens to rock, reads comic books, or plays video games, this would not establish a correlation between such behavior and these supposed corrupters of the youth. Correlation requires variation. Evidence for the proposition that kids who engage in those activities are *more* likely to be violent than kids who do not engage in those activities must involve a comparison of these two types of kids.

If we want to know if there is a relationship between some putative scourge of the youth and violence, we must not select on the dependent variable—that is, we must compare violent kids to non-violent kids and see whether violent kids are more likely to engage in that scourge than non-violent kids. (Again, even then, we can't say the relationship is causal.) The fact that even experts can fail to think clearly about this means that, for all the expert opinion offered on the topic, we know far less than we could about the correlates of youth violence.

High School Dropouts

Let's stick, for the moment, with troubled youth. Early twenty-first-century America has a high school graduation problem. At a time when the economic returns to education are at an all-time high, almost a third of students in the public schools fail to complete high school on time. Over 10 percent never graduate.

In 2006, the Bill and Melinda Gates Foundation decided to put some resources into addressing this issue. As one step in trying to find a solution, they commissioned a study on the correlates of dropping out of high school. The report's main thrust is that high school dropout is not primarily associated with the things you might have guessed—problems at home, lack of academic preparation, or listening to rock music. Rather, the big problem seems to be that kids aren't engaged by the educational environment and find school boring.

As the report states, "nearly half (47 percent) [of dropouts] said a major reason for dropping out was that classes were not interesting." And "nearly 7 in 10 respondents (69 percent) said they were not motivated or inspired to work hard."

Unfortunately, because correlation requires variation, the evidence in this Gates Foundation study, just like the evidence presented by the PMRC and the anti–comic book lobby before it, is pretty uninformative.

The fact that half of high school dropouts report finding school uninteresting does not mean that finding school uninteresting correlates with dropout. Because correlation requires variation, measuring the correlation has to involve comparing dropouts to non-dropouts to see whether dropouts are more likely to find school uninteresting. The Gates Foundation study, because it looks only at high school dropouts, can't make this comparison.

This point isn't just pedantic. Think about it for a second. Both authors of this book went to high school. Neither dropped out. However, both authors recall finding some classes uninteresting. Didn't you?

Now, our personal experiences also don't constitute compelling evidence. So let's see if we can do a little better in figuring out whether finding classes boring is really a key predictor of dropout. Researchers at Indiana University did a nationally representative survey of high school students in 2009. Most of these students are not going to drop out, yet the researchers report that "two out of three respondents (66%) in 2009 are bored

at least every day in class." That's even more than the 50 percent of dropouts who find school boring in the Gates Foundation study.

But let's be careful. There are many reasons the Gates Foundation survey and the Indiana University survey can't be compared. They sample different groups of students, ask different questions, and are from different years. So we don't want to leap to conclusions. But at the very least, the Indiana University survey should make you worry that finding school boring is in fact a very common experience for high school students, not just those who drop out.

The future of American education is serious stuff. It is admirable that the Gates Foundation is trying to improve education. But their research ignores a key principle of thinking clearly with data; they are trying to learn about the correlates of educational failure without any variation in failure versus success. This approach cannot work.

Suicide Attacks

In 2009, University of Chicago professor and noted terrorism expert Robert Pape testified to the House of Representatives Armed Services Subcommittee on Terrorism. The topic was General Stanley McChrystal's proposal for a forty-thousand-troop surge to fight the Taliban insurgency in Afghanistan. Here is what Pape had to say:

> The picture is clear, the more Western troops have gone to Afghanistan, the more local residents have viewed themselves as under foreign occupation—and are using suicide and other terrorism to resist it. . .As my study of suicide terrorism around the world since 1980 shows, what motivates suicide terrorists is not the existence of a terrorist sanctuary, but the presence of foreign forces on land they prize. So, it is little surprise that US troops are producing anti-American suicide attackers.

Pape goes on to recommend a major rethinking of American military strategy in Afghanistan. His argument is based on the claim that suicide attacks are primarily motivated by foreign occupation. His evidence is the data he collected and analyzed in articles and two books on every suicide terrorist campaign in the world since 1980.

The argument sounds plausible. In Afghanistan, U.S. forces were being attacked by suicide bombers who wanted the United States to leave the country. Tamil Tiger suicide bombers attacked a government in Sri Lanka they believed was occupying their homeland. Palestinian suicide bombers attack Israelis, arguing that they are foreign occupiers. It sure seems like occupation is a major correlate of suicide attacks.

Now, the claim that virtually every suicide attack is targeted against a foreign occupier is, we think, debatable. (For instance, while Osama bin Laden claimed the American troops stationed in Saudi Arabia at the invitation of the Saudi government were an occupying force, are we sure we agree with him?) But, for the sake of argument, let's assume that the basic factual claim is correct. Does this mean that there is a correlation between foreign occupation and suicide attacks?

The answer is, of course, no. Correlation requires variation. To understand the correlates of suicide attacks, you can't just study every single instance of a suicide attack and look for commonalities. That is selecting on the dependent variable. You must compare conflicts with suicide attacks to those without.

An easy thing to do in this case is to simply look at every single country and ask: Are foreign-occupied countries more likely to experience suicide attacks than countries

that are not foreign occupied? It turns out that a recent study did precisely that comparison and found that the answer was no. In particular, if we compare occupied to non-occupied countries, the difference in likelihood of experiencing suicide violence is less than 1 percentage point!

What is going on? All those examples of suicide bombers that we listed involved attacking foreign occupiers. How could it be that there is almost no correlation between foreign occupation and suicide attacks?

The way to get some intuition is to think about how many foreign occupations there have been that didn't lead to suicide terrorism. The British occupation of Ireland, despite sparking a decades-long campaign of violent resistance, never gave rise to suicide terrorism. Basque separatists in Spain fought a decades-long campaign and never resorted to suicide attacks. At various points during the Cold War (and beyond), the United States stationed troops in Germany, Japan, South Korea, Grenada, Panama, and Haiti (arguably, all as much occupations as the putative occupation of Saudi Arabia) but suffered not even one suicide attack in any of these locations. If occupation predicts suicide violence, what was going on in all these places?

This example has another nice feature. It not only illustrates the mistake of looking for correlation without variation. It shows you how misled you can be by trying to reach conclusions by only looking at cases where the phenomenon of interest (here, suicide attacks) occurs—that is, by selecting on the dependent variable. To see this, it helps to go back in history a little.

Suppose you'd started collecting data on suicide violence in the early 1980s. By 1986 you'd have recorded thirty-three attacks and over one thousand deaths. Essentially every single one of those attacks was carried out by the armed Shi'a militia Hezbollah against American, Israeli, and French targets in Lebanon, including the attack on the U.S. Marines Barracks in Beirut, which killed 320 people.

If you'd looked for commonalities amongst every suicide attack ever committed in 1986, you might have noticed that they were all carried out by Muslims in the Middle East. Using the same logic that led to the conclusion that occupation is a major predictor of suicide attacks, you might have concluded that Islam was the key correlate.

Of course, if you had done a proper comparison, you wouldn't have reached this conclusion. There are a whole lot of Muslim-majority countries in the world. In 1986, almost none of them had experienced suicide violence.

Moreover, if you were trying to forecast where the next suicide attack might occur, this conclusion in 1986 would have led you terribly astray. In 1987, the world saw the first suicide attack by the Liberation Tigers of Tamil Eelam (Tamil Tigers), a group of secular separatists in Sri Lanka with no ties to Islam. The attack marked the beginning of what would become the largest campaign of suicide violence the world had ever seen. When you try to establish correlation without variation, you can get things colossally wrong.

The World Is Organized to Make Us Select on the Dependent Variable

As we've seen, it is incredibly easy to fall into the trap of selecting on the dependent variable simply by failing to think clearly. But matters are even worse than that. The world sometimes seems to be organized in a way that almost forces us to look for correlation without variation. In this section we look at three ways in which that is true:

the organization of certain professions, the practice of post-mortem analyses following disasters, and the way we seek life advice.

Doctors Mostly See Sick People

Anyone who has suffered from significant back pain knows that it is rough. When, inevitably, many of you develop back pain, you will likely go to a doctor, who will send you to get an MRI. Usually, the MRI shows some bulging or herniated discs in the afflicted back. These bulging discs are taken to be the cause, in some not fully understood way, of the back pain (maybe by impinging a nerve).

The recommendations following this diagnosis can vary greatly. Some doctors want to operate. Others will refer you to a pain clinic where yet other doctors might stick you with giant needles with medication that dulls pain and reduces inflammation. Still others will suggest you try physical therapy and take lots of painkillers.

Here's the kicker. As best we can tell, there is precious little evidence that having a bulging disc is correlated with back pain. Here are the facts. People with back pain are quite likely to exhibit disc herniation. Indeed, in a 2011 British study published in the journal *Pain*, about two-thirds of back pain sufferers who were referred for an MRI had nerve compression as a result of a disc bulge or herniation. This seems like evidence that those bulging discs really are a problem.

But remember, correlation requires variation. You should be asking yourself: What about people without back pain? How do their discs look? Good question. The answer is, they look exactly the same as the people's discs who do have back pain! A 1994 study published in the *New England Journal of Medicine* found that about two-thirds of people who do not suffer from back pain also have a disc bulge or herniation. Once you compare both variables of interest, the apparent association between bulging discs and back pain disappears.

It is easy to see how doctors could end up associating bulging discs with back pain. Even if they are thinking clearly, by dint of profession, a doctor is almost doomed not to look at variation. Sick people go to the doctor. Healthy people tend not to. Your typical back doctor just doesn't get much of an opportunity to look at the MRIs of people with well-functioning backs.

Post-Mortems

Another way the world is organized to make us look for correlation without variation is through institutional rules or procedures. A particularly common example is the way organizations respond to both great failures and great successes.

Following a crisis or disaster, organizations want to know what went wrong so they can avoid making similar mistakes in the future. Likewise, following great successes they want to know what went right to establish best practices. Achieving these goals is the role of a post-mortem analysis. Looking closely at an instance of great failure or great success is not, in and of itself, a mistake. Indeed, it is a very sensible starting point. But, if you think clearly, you should already be able to see that, on their own, such post-mortem procedures are not sufficient to establish correlations between what went wrong (or right) and existing practices.

The question you should be trying to answer when assessing lessons learned from a crisis is, Which decisions should have been made differently to avoid the crisis, given what we knew at the time? However, when assessing lessons learned, we often slip

Table 4.5. Rehearsal strategies in the week before competitions where your band performed poorly (made-up data).

	Do Well	Do Poorly	Total
Extra Rehearsals	?	80	?
Take It Easy	?	8	?
Total	?	88	?

into answering a slightly different question: Which decisions should have been made differently to avoid the crisis, given what we know now?

The latter isn't a terribly useful question to answer, for the reasons we've already talked about in this chapter. Suppose you find some decision that, it turns out, seems to have led directly to the disaster. After the fact, it is easy to say, "Had we not taken that action, the disaster wouldn't have happened." But does that mean that you shouldn't take similar such actions in the future? To know the answer to that, you'd want to know whether disasters are more likely to occur in the presence of such actions than in their absence. That is, you want to know whether there is a correlation between taking such actions and disasters occurring. To establish a correlation, you need variation. But a post-mortem, almost by definition, has no variation. You are only looking at an instance of the disaster occurring.

To see what we mean a little more intuitively, let's start with a fictional example. Then we'll turn to some real cases.

Imagine you are a high school band director preparing for a regional competition in a week. You have to decide whether to push the kids hard with a grueling schedule of rehearsals or give them time off so they go into the competition relaxed. You weigh the pros and cons, deciding preparation is more important than mental state. So you schedule a week of extra rehearsals. Unfortunately, the band doesn't play terribly well on the day of the competition, and you are eliminated in the first round.

In your post-mortem analysis you ask the question, What should I have done to avoid the loss? It occurs to you that you've seen a lot of bands lose competitions in this same way (i.e., having rehearsed themselves to death the week before), so you decide to collect some data. You look at the history of all the competitions in which your band was eliminated in the early rounds. Just like in this year's competition, you find that in almost every one of these competitions, you scheduled a heavy rehearsal schedule in the week leading up to the competition.

Let's say you did a week of intensive rehearsing prior to 80 out of 88 losses. The post-mortem conclusion seems clear. In over 90 percent of the cases where your band was eliminated early, it was after a week of exhausting rehearsal. Now you feel even more sure: intensive rehearsal is the wrong strategy. Table 4.5 summarizes what you know so far from your post-mortem analysis.

But this conclusion doesn't necessarily follow from the data you've collected. In fact, from this data alone, there's no way to know whether those rehearsals are associated with performing well or poorly, because you have answered the wrong question.

You don't want to know if bands did extra rehearsals prior to most of the competitions where they performed poorly. You want to know if extra rehearsals are positively or negatively correlated with performing well. The answer to that question will help you know whether those extra rehearsals are a good idea for the next competition.

Table 4.6. Rehearsal strategy in the week before competitions where your band performed well or poorly (made-up data).

	Do Well	Do Poorly	Total
Extra Rehearsals	300	80	380
Take It Easy	12	8	20
Total	312	88	400

To answer that question, you have to look at the correlation between extra rehearsals and performing well in competition. But you can't know the correlation from your post-mortem analysis. Correlation requires variation. Your post-mortem, by focusing only on poor performances, guarantees that you lack the variation needed to establish a correlation.

To do a better job, you could look at the history of all the band competitions you've participated in to see whether you performed well or poorly. Now you have variation in both variables and can fill in all the data, as shown in table 4.6.

From this table it is clear that there is in fact a strong positive correlation between scheduling extra rehearsals and performing well. The probability of your band performing well when you rehearsed hard is about 79 percent ($\frac{300}{380} \approx .79$). By contrast, the probability of your band performing well when you took it easy the week prior to a competition is only 60 percent ($\frac{12}{20} = .60$). The only reason that the post-mortem turned up the finding that almost every poor performance involved intensive rehearsals is that those extra rehearsals are so effective that sensible band directors almost always schedule them.

By finding the variation needed to establish the correlation that is actually relevant to the question at hand, you reach a very different conclusion than you did in your original post-mortem. Following the loss, it seemed like intensive rehearsals were a bad idea. But before the fact, given the information available, rehearsing hard was exactly the right call. Faced with the same situation again, you should probably make the same decision.

This problem is endemic to the process of post-mortems following disasters. We tend to look at the factors that seem like they contributed to the disaster, ask if they were also present in past disasters, and, when they were, conclude that we should eliminate those factors in the future. But, in so doing, we are making the same mistake as the band director. Without variation in whether or not a disaster occurred, we can't actually learn whether the presence of those factors is correlated with the occurrence of a disaster. So we don't know if there are lessons to be learned.

We are going to show you what we mean with two examples of post-mortems that followed major disasters—the *Challenger* space shuttle explosion in 1986 and the financial crisis of 2008. In each case, we will see that, while after the fact it sure looks like some serious and obvious mistakes were made, it is less clear that the decision makers could have known that they were making mistakes before the fact. Moreover, once we've grasped this, we will be able to think more clearly about how to design post-mortems that might be more informative about lessons learned.

The Challenger disaster

On January 28, 1986, the space shuttle *Challenger* disintegrated off the coast of Cape Canaveral less than two minutes after launch. Seven crew members were killed. The

night before the *Challenger* exploded, a small group of engineers from the NASA contractor responsible for the shuttle's solid rocket boosters predicted that the cold weather would lead to a catastrophic failure that might well compromise the shuttle. The concern was that the critical O-ring seals responsible for containing gases produced by burning rocket fuel were not certified to operate at the low temperatures that preceded this particular launch. If the O-ring seals failed, the engineers argued, hot pressurized gas could burn through the rocket's casing, causing disaster.

These predictions, shunted aside by managers at NASA and the engineers' own firm, proved tragically correct. Many post-mortem analyses focused on NASA's failure to take these concerns seriously. The conclusion most observers reached was that the disaster was caused by organizational and cultural failures at NASA that facilitated group-think and led managers to systematically ignore important objections from experts. For instance, the *Report of the Presidential Commission on the Space Shuttle Challenger Accident* (the Rogers Commission) concluded, "Failures in communication . . . resulted in a decision to launch 51-L based on incomplete and sometimes misleading information, a conflict between engineering data and management judgements, and a NASA management structure that permitted internal flight safety problems to bypass key Shuttle managers."

The *Challenger* case is interesting. No one questions the physics behind the conclusion that the O-rings failed because of cold temperatures. Indeed, the Rogers Commission included the Nobel Prize–winning physicist Richard Feynman precisely so they could say with authority whether the engineers were right on the science. They were. And so, in this sense, launching the shuttle was clearly a mistake.

Because the science is so clear, it seems natural for a post-mortem to ask what it was about the process that led decision makers to ignore engineers making good scientific arguments. Here is where our knowledge of the pitfalls of post-mortems should make us stop and think. We know that, after the fact, the decision to launch was tragically flawed. But we want to evaluate whether it was a bad decision at the time it was made. To do so, we need to know about the correlation between the presence of scientifically valid engineering concerns and the success of shuttle launches. And to know about that correlation, we need variation; we must compare disastrous launches to successful launches.

We aren't engineers, so we aren't going to try to weigh in on whether or not the decision to launch *Challenger* was reasonable at the time it was made. But we can see how, to analyze this, a post-mortem commission would need to ask questions they aren't accustomed to asking. Post-mortem commissions ask what led to the disaster, whether people had raised the relevant objections, and, if so, why those objections weren't listened to. In addition, such commissions need to ask whether engineers also raised scientifically valid concerns prior to lots of successful launches. This doesn't seem implausible. Space shuttle launches are incredibly complex and dangerous undertakings. Perhaps there is almost always a scientifically valid reason for serious concern. If so, then there actually wouldn't be much (if any) correlation between the presence of such concerns and launch success. If this is the case, unless you are prepared to simply shut down the space program, it isn't fair to say that launching following a scientifically plausible objection by an engineer is always a mistake. This is the sort of thing one would want to know from a post-mortem commission before reaching conclusions about changing NASA's organizational culture or management practices.

The financial crisis of 2008

The financial crisis that shook the world economy in 2007 and 2008 began with a crash in the U.S. subprime housing market. This crash had ripple effects across the banking sector that eventually spread throughout the world. Understandably, in the wake of this crisis—at the time, the worst since the Great Depression—policy makers and the public alike were interested in identifying early warning indicators that might help them forecast and forestall future crises.

Perhaps the most important post-mortem analysis attempting to provide such early warning indicators was the book *This Time Is Different* by the economists Carmen M. Reinhart and Kenneth S. Rogoff. Reinhart and Rogoff collected and analyzed data on every major financial crisis of the last eight hundred years. By doing so, they argued, they could identify a few key indicators that almost always precede such a crisis. These include uncommonly large current account deficits (that is, goods and services exported minus imported net of income from abroad), asset price bubbles, and excessive borrowing. For instance, in 2006 the United States had a current account deficit close to 7 percent of GDP, a bubble in the housing market, and ballooning federal debt. Thus, Reinhart and Rogoff conclude, "we've been here before." The implication is that the 2008 U.S. financial crisis could have been predicted by the presence of those same factors that seem to characterize financial crises across time and around the globe. Similar patterns were true before the financial crises in Latin America in the early 2000s, East Asia in the 1990s, Nordic countries in the 1980s, and so on into history.

The problem with this argument is the same as in our earlier examples. Early warning indicators should be correlates of financial crises. Because correlation requires variation, to know if current account deficits, soaring asset prices, and heavy borrowing correlate with financial crises, we need variation in crises. That is, we need to know not only that these factors tend to be present when crises occur but also how frequently they are present when crises do not occur. Without such variation, we cannot establish a correlation.

Reinhart and Rogoff's plan of studying every major financial crisis for eight hundred years cannot answer the question. And there are reasons to be worried about their conclusions. As the MIT political scientist David Andrew Singer points out, one need only look at recent history to cast some doubt on the story. For instance, in the late 1990s the United States had all the early warning signs for a financial crisis. There were large current account deficits as a result of massive foreign investment in dot-coms. Moreover, when the dot-com bubble burst, "it wiped out approximately $5 trillion in market capitalization." Yet no financial crisis occurred. This, of course, is just one anecdote. But it should make you wonder whether the factors Reinhart and Rogoff point to are really good predictors of financial crises or just common features of the world that happen to exist both when financial crises occur and when they don't occur.

Life Advice

We've been arguing that our world is organized in ways that lead us to try to figure out the correlates of success or failure without looking at variation, even though it won't work. It is important to see that this problem isn't confined to big institutional settings. We are all victims of it every day in many small ways.

One simple example is the ways in which we seek life advice, which almost always involves asking successful people how it is that they succeeded. In our business, for instance, graduate students are encouraged to ask senior professors what they did to succeed on the job market. We imagine something similar is true in other professions. There is certainly no shortage of self-help books describing the habits of successful people.

But such wisdom suffers from exactly the problems we've been pointing to. Successful people, reflecting on their lives, are inclined to identify a few decisions they made or a few personal characteristics that seem important and offer them as advice to the next generation. But those successful people typically have no idea whether many other, less successful people made similar decisions or had similar characteristics. That is, their introspection about the correlates of success lacks variation. As such, successful people don't really know whether the lessons they point to in telling their personal stories are correlates of success or not. And so, we leave you with this happy bit of wisdom of our own: Beware life advice. Most of it is probably nonsense.

Wrapping Up

Correlation requires variation. But unclear thinking and organizational mandates often lead us to select on the dependent variable—trying to establish the correlates of some phenomenon by only looking at instances when it occurred. It requires careful attention to make sure you aren't falling into this trap, whether you are doing quantitative analysis or just trying to think informally about evidence. Even just forcing yourself to think about whether you could fill in all four cells of one of our two-by-two tables is a good starting point for avoiding looking for correlation without variation.

You can be even more rigorous by using quantitative techniques to measure correlations. The most important such technique is called regression, the topic of chapter 5.

Key Term

- **Selecting on the dependent variable:** Examining only instances when the phenomenon of interest occurred, rather than comparing cases where it occurred to cases where it did not occur.

Exercises

4.1   In chapter 2 we discussed the differences between statements about correlations and other factual statements that do not convey information about a correlation. Now that you have a deeper understanding that correlation requires variation, consider the following statements. Which ones describe a correlation, and which ones do not?

(a) Most top-performing schools have small student bodies.
(b) Married people are typically happier than unmarried people.
(c) Among professionals, taller basketball players tend to have lower free-throw percentages than shorter players.
(d) The locations in the United States with the highest cancer rates are typically small towns.
(e) Older houses are more likely to have lead paint than newer ones.

(f) Most colds caught in Cook County are caught on cold days. (This one also doubles as a tongue twister.)

4.2   At least twenty billionaires dropped out of college before earning their fortunes, including Bill Gates and Mark Zuckerberg.

(a) Does this mean that dropping out of college is correlated with becoming a billionaire? Why or why not?

(b) Draw the two-by-two table that would allow you to assess whether dropping out of college is correlated with becoming a billionaire. Let's assume that exactly twenty people have dropped out of college and become billionaires, so you know what to put in one of the four cells. Make your best guess for the other cells. At the time of this writing, there are about 7.8 billion people in the world, and about two thousand billionaires. Do you think there is a positive or negative correlation between dropping out of college and becoming a billionaire?

(c) Given your guesses from part (b), what proportion of the non-billionaires would need to be college dropouts in order for the correlation to be negative? What proportion of the non-billionaires would need to be college dropouts in order for the correlation to be positive?

(d) If you're currently a college student deciding whether you want to drop out in the hopes of becoming a billionaire, you may want to restrict attention to people who actually started college. Do you think the correlation between dropping out of college and becoming a billionaire is more or less likely to be positive if we restrict attention to just people who start college?

(e) About 7 percent of the world's population has a college degree. And about a third of people who start college complete it. If we assume that everyone who becomes a billionaire started college, you should now have all the information you need to assess the correlation between becoming a billionaire and dropping out of college among those who start college. Is it positive, negative, or zero?

4.3   Identify one recent case where an analyst made the mistake discussed in this chapter. That is, find a case where someone (at least implicitly) makes a claim about a correlation but they don't have variation in one of their variables. Your example might come from a newspaper article, an academic study, a policy memo, or a statement from a politician or business leader.

(a) Summarize the claim being made (perhaps implicitly) and explain why the evidence does not necessarily support the claim.

(b) Explain what additional data collection and analysis *would* allow the analyst to assess the correlation of interest.

(c) Draw a two-by-two table that illustrates your argument, and discuss what the unknown numbers in the table would have to be in order for the correlation of interest to be positive, negative, or zero.

Readings and References

We extensively discussed

Malcolm Gladwell. 2008. *Outliers: The Story of Success*. Little, Brown.

For more information on Fredric Wertham and his flawed argument about comic books, we recommend

David Hajdu. 2009. *The Ten-Cent Plague: The Great Comic-Book Scare and How It Changed America*. Picador.

And although Wertham's conclusions don't follow from his empirical findings because correlation requires variation, it also turns out that Wertham may have manipulated and fabricated his data. See

Carol L. Tilley. 2012. "Seducing the Innocent: Fredric Wertham and the Falsifications That Helped Condemn Comics." *Information & Culture: A Journal of History* 47(4):383–413.

We discussed two reports by the U.S. Department of Education and the Secret Service on school safety.

- The 2002 report is here: https://www.govinfo.gov/content/pkg/ERIC-ED466024/pdf/ERIC-ED466024.pdf.
- The 2018 report is here: https://www2.ed.gov/documents/school-safety/school-safety-report.pdf.

The report on the high school dropout problem prepared for the Gates Foundation is

John M. Bridgeland, John J. Dilulio, Jr., and Karen Burke Morison. *The Silent Epidemic: Perspectives of High School Dropouts*. https://docs.gatesfoundation.org/documents/thesilentepidemic3-06final.pdf.

The survey on boredom in school is called the High School Survey of Student Engagement, and it's administered by the Center for Evaluation and Education Policy at Indiana University. The excerpt is from their 2010 study. For more information, see http://newsinfo.iu.edu/news-archive/14593.html.

For more on suicide attacks, including why we can't learn about the causes or correlates of suicide violence from studying only cases where it occurs, see

Robert A. Pape. 2003. "The Strategic Logic of Suicide Terrorism." *American Political Science Review* 97(3):343–61.

Scott Ashworth, Joshua D. Clinton, Adam Meirowitz, and Kristopher W. Ramsay. 2008. "Design, Inference, and the Strategic Logic of Suicide Terrorism." *American Political Science Review* 102(2):269–73.

Robert A. Pape. 2008. "Methods and Findings in the Study of Suicide Terrorism." *American Political Science Review* 102(2):275–77.

Scott Ashworth, Joshua D. Clinton, Adam Meirowitz, and Kristopher W. Ramsay. 2008. "Design, Inference, and the Strategic Logic of Suicide Terrorism: A Rejoinder." Unpublished note: http://home.uchicago.edu/~sashwort/rejoinder3.pdf.

To read about the high rates of disc bulges and herniation among people with and without back pain (and therefore the lack of correlation between these characteristics and back pain), see

Michael J. DePalma, Jessica M. Ketchum, and Thomas Saullo. 2011. "What Is the Source of Chronic Low Back Pain and Does Age Play a Role?" *Pain Medicine* 12(2):224–33.

Maureen C. Jensen, Michael N. Brant-Zawadzki, Nancy Obuchowski, Michael T. Modic, Dennis Malkasian, and Jeffrey S. Ross. 1994. "Magnetic Resonance Imaging of the Lumbar Spine in People without Back Pain." *New England Journal of Medicine* 331:69–73.

These are the two studies we referenced on the financial crisis of 2008:

Carmen M. Reinhart and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly.* Princeton University Press.

David A. Singer. 2010. "Is This Time Different?" *The Political Economist.* Fall, pp. 4–5.