17

Even comparing apples and apples leaves plenty of room for shenanigans. As discussed in the last chapter, one important role of statistics is to describe changes in quantities over time. Are taxes going up? How many cheeseburgers are we selling compared with last year? By how much have we reduced the arsenic in our drinking water? We often use percentages to express these changes because they give us a sense of scale and context. We understand what it means to reduce the amount of arsenic in the drinking water by 22 percent, whereas few of us would know whether reducing arsenic by one microgram (the absolute reduction) would be a significant change or not. Percentages don't lie—but they can exaggerate. One way to make growth look explosive is to use percentage change to describe some change relative to a very low starting point. I live in Cook County, Illinois. I was shocked one day to learn that the portion of my taxes supporting the Suburban Cook County Tuberculosis Sanitarium District was slated to rise by 527 percent! However, I called off my massive antitax rally (which was really still in the planning phase) when I learned that this change would cost me less than a good turkey sandwich. The Tuberculosis Sanitarium District deals with roughly a hundred cases a year; it is not a large or expensive organization. The *Chicago Sun-Times* pointed out that for the typical homeowner, the tax bill would go from $1.15 to $6.[7] Researchers will sometimes qualify a growth figure by pointing out that it is "from a low base," meaning that any increase is going to look large by comparison.

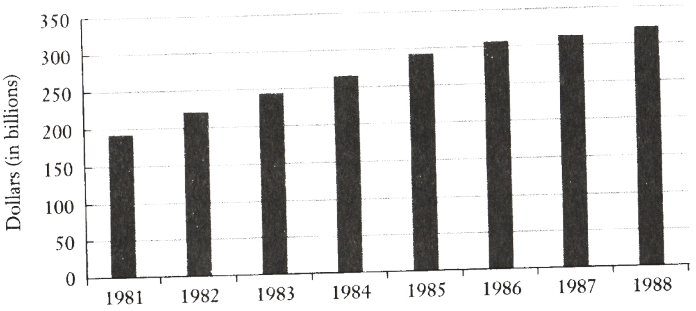
Obviously the flip side is true. A small percentage of an enormous sum can be a big number. Suppose the secretary of defense reports that defense spending will grow only 4 percent this year. Great news! Not really, given that the Defense Department budget is nearly $700 billion. Four percent of $700 billion is $28 billion, which can buy a lot of turkey sandwiches. In fact, that seemingly paltry 4 percent increase in the defense budget is *more than the entire NASA budget and about the same as the budgets of the Labor and Treasury Departments combined.*

In a similar vein, your kindhearted boss might point out that as a matter of fairness, every employee will be getting the same raise this year, 10 percent. What a magnanimous gesture—except that if your boss makes $1 million and you make $50,000, his raise will be $100,000 and yours will be $5,000. The statement "everyone will get the same 10 percent raise this year" just sounds so much better than "my raise will be twenty times bigger than yours." Both are true in this case.

18

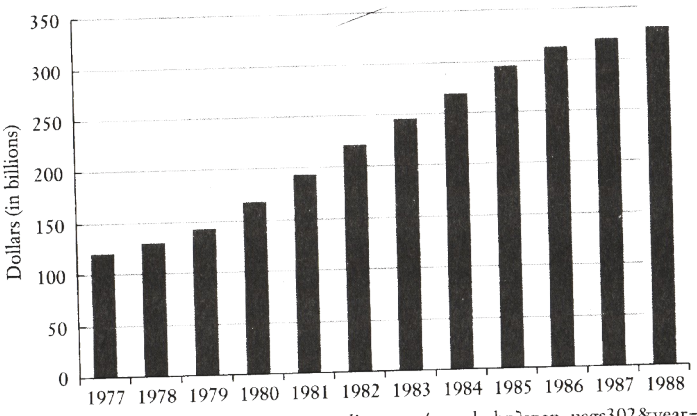
Any comparison of a quantity changing over time must have a start point and an end point. One can sometimes manipulate those points in ways that affect the message. I once had a professor who liked to speak about his "Republican slides" and his "Democratic slides." He was referring to data on defense spending, and what he meant was that he could organize the same data in different ways in order to please either Democratic or Republican audiences. For his Republican audiences, he would offer the following slide with data on increases in defense spending under Ronald Reagan. Clearly Reagan helped restore our commitment to defense and security, which in turn helped to win the Cold War. No one can look at these numbers and not appreciate the steely determination of Ronald Reagan to face down the Soviets.

Defense Spending in Billions, 1981–1988



For the Democrats, my former professor merely used the same (nominal) data, but a longer time frame. For this group, he pointed out that Jimmy Carter deserves credit for beginning the defense buildup. As the following "Democratic" slide shows, the defense spending increases from 1977 to 1980 show the same basic trend as the increases during the Reagan presidency. Thank goodness that Jimmy Carter—a graduate of Annapolis and a former naval officer—began the process of making America strong again!

Defense Spending in Billions, 1977–1988



Source: http://www.usgovernmentspending.com/spend.php?span=usgs302&year=1988&view=1&expand=30&expandC=&units=b&fy=fy12&local=s&state=US&pie=#usgs302.

Excerpts from Wheelan, C. (2013). *Naked Statistics*. New York, NY. W. W. Norton and Company.

19

Consider school quality. This is a crucial thing to measure, since we would like to reward and emulate "good" schools while sanctioning or fixing "bad" schools. (And within each school, we have the similar challenge of measuring teacher quality, for the same basic reason.) The most common measure of quality for both schools and teachers is test scores. If students are achieving impressive scores on a well-conceived standardized test, then presumably the teacher and school are doing a fine job. Conversely, bad test scores are a clear signal that lots of people should be fired, sooner rather than later. These statistics can take us a long way toward fixing our public education system, right?

Wrong. Any evaluation of teachers or schools that is based solely on test scores will present a dangerously inaccurate picture. Students who walk through the front door of different schools have vastly different backgrounds and abilities. We know, for example, that the education and income of a student's parents have a significant impact on achievement, regardless of what school he or she attends. The statistic that we're missing in this case happens to be the only one that matters for our purposes: How much of a student's performance, good or bad, can be attributed to what happens inside the school (or inside a particular classroom)?

Students who live in affluent, highly educated communities are going to test well from the moment their parents drop them off at school on the first day of kindergarten. The flip side is also true. There are schools with extremely disadvantaged populations in which teachers may be doing a remarkable job but the student test scores will still be low—albeit not nearly as low as they would have been if the teachers had not been doing a good job. What we need is some measure of "value-added" at the school level, or even at the classroom level. We don't want to know the absolute level of student achievement; we want to know how much that student achievement has been affected by the educational factors we are trying to evaluate.

At first glance, this seems an easy task, as we can simply give students a pretest and a posttest. If we know student test scores when they enter a particular school or classroom, then we can measure their performance at the end and attribute the difference to whatever happened in that school or classroom.

Alas, wrong again. Students with different abilities or backgrounds may also learn *at different rates*. Some students will grasp the material faster than others for reasons that have nothing to do with the quality of the teaching. So if students in Affluent School A and Poor School B both start algebra at the same time and level, the explanation for the fact that students at Affluent School A test better in algebra a year later may be that the teachers are better, or it may be that the students were capable of learning faster—or both. Researchers are working to develop statistical techniques that measure instructional quality in ways that account appropriately for different student backgrounds and abilities. In the meantime, our attempts to identify the "best" schools can be ridiculously misleading.

Every fall, several Chicago newspapers and magazines publish a ranking of the "best" high schools in the region, usually on the basis of state test score data. Here is the part that is laugh-out-loud funny from a statistical standpoint: Several of the high schools consistently at the top of the rankings are selective enrollment schools, meaning that students must apply to get in, and only a small proportion of those students are accepted. One of the most important admissions criteria is standardized test scores. So let's summarize: (1) these schools are being recognized as "excellent" for having students with high test scores; (2) to get into such a school, one must have high test scores. This is the logical equivalent of giving an award to the basketball team for doing such an excellent job of producing tall students.

20

Even if you have a solid indicator of what you are trying to measure and manage, the challenges are not over. The good news is that "managing by statistics" can change the underlying behavior of the person or institution being managed for the better. If you can measure the proportion of defective products coming off an assembly line, and if those defects are a function of things happening at the plant, then some kind of bonus for workers that is tied to a reduction in defective products would presumably change behavior in the right kinds of ways. Each of us responds to incentives (even if it is just praise or a better parking spot). Statistics measure the outcomes that matter; incentives give us a reason to improve those outcomes.

Or, in some cases, just to make the statistics look better. That's the bad news.

If school administrators are evaluated—and perhaps even compensated—on the basis of the high school graduation rate for students in a particular school district, they will focus their efforts on boosting the number of students who graduate. Of course, they may also devote some effort to improving the graduation rate, which is not necessarily the same thing. For example, students who leave school before graduation can be classified as "moving away" rather than dropping out. This is not merely a hypothetical example; it is a charge that was leveled against former secretary of education Rod Paige during his tenure as the Houston school superintendent. Paige was hired by President George W. Bush to be U.S. secretary of education because of his remarkable success in Houston in reducing the dropout rate and boosting test scores.

If you're keeping track of the little business aphorisms I keep tossing your way, here is another one: "It's never a good day when *60 Minutes* shows up at your door." Dan Rather and the *60 Minutes II* crew made a trip to Houston and found that the manipulation of statistics was far more impressive than the educational improvement.[8] High schools routinely classified students who quit high school as transferring to another school, returning to their native country, or leaving to pursue a General Equivalency Diploma (GED)—none of which count as dropping out in the official statistics. Houston reported a citywide dropout rate of 1.5 percent in the year that was examined; *60 Minutes* calculated that the true dropout rate was between 25 and 50 percent.

The statistical chicanery with test scores was every bit as impressive. One way to improve test scores (in Houston or anywhere else) is to improve the quality of education so that students learn more and test better. This is a good thing. Another (less virtuous) way to improve test scores is to prevent the worst students from taking the test. If the scores of the lowest-performing students are eliminated, the average test score for the school or district will go up, even if all the rest of the students show no improvement at all. In Texas, the statewide achievement test is given in tenth grade. There was evidence that Houston schools were trying to keep the weakest students from reaching tenth grade. In one particularly egregious example, a student spent three years in ninth grade and then was promoted straight to eleventh grade—a deviously clever way of keeping a weak student from taking a tenth-grade benchmark exam without forcing him to drop out (which would have showed up on a different statistic).

It's not clear that Rod Paige was complicit in this statistical trickery during his tenure as Houston superintendent; however, he did implement a rigorous accountability program that gave cash bonuses to principals who met their dropout and test score goals and that fired or demoted principals who failed to meet their targets. Principals definitely responded to the incentives; that's the larger lesson. But you had better be darn certain that the folks being evaluated can't make themselves look better (statistically) in ways that are not consistent with the goal at hand.