1

THE IMPORTANCE OF SOCIAL STATISTICS

Nineteenth-century Americans worried about prostitution; reformers called it "*the* social evil" and warned that many women prostituted themselves. How many? For New York City alone, there were dozens of estimates: in 1833, for instance, reformers published a report declaring that there were "not less than 10,000" prostitutes in New York (equivalent to about 10 percent of the city's female population); in 1866, New York's Methodist bishop claimed there were more prostitutes (11,000 to 12,000) than Methodists in the city; other estimates for the period ranged as high as 50,000. These reformers hoped that their reports of widespread prostitution would prod the authorities to act, but city officials' most common response was to challenge the reformers' numbers. Various investigations by the police and grand juries produced their own, much lower estimates; for instance, one 1872 police report counted only 1,223 prostitutes (by that time, New York's population included nearly half a mil-

lion females). Historians see a clear pattern in these cycles of competing statistics: ministers and reformers "tended to inflate statistics";[1] while "police officials tended to underestimate prostitution."[2]

Antiprostitution reformers tried to use big numbers to arouse public outrage. Big numbers meant there was a big problem: if New York had tens of thousands of prostitutes, something ought to be done. In response, the police countered that there were relatively few prostitutes—an indication that they were doing a good job. These dueling statistics resemble other, more recent debates. During Ronald Reagan's presidency, for example, activists claimed that three million Americans were homeless, while the Reagan administration insisted that the actual number of homeless people was closer to 300,000, one-tenth what the activists claimed. In other words, homeless activists argued that homelessness was a big problem that demanded additional government social programs, while the administration argued new programs were not needed to deal with what was actually a much smaller, more manageable problem. Each side presented statistics that justified its policy recommendations, and each criticized the other's numbers. The activists ridiculed the administration's figures as an attempt to cover up a large, visible problem, while the adminstration insisted that the activists' numbers were unrealistic exaggerations.[3]

Statistics, then, can become weapons in political struggles over social problems and social policy. Advocates of different positions use numbers to make their points ("It's a big problem!"

"No, it's not!"). And, as the example of nineteenth-century estimates of prostitution reminds us, statistics have been used as weapons for some time.

THE RISE OF SOCIAL STATISTICS

In fact, the first "statistics" were meant to influence debates over social issues. The term acquired its modern meaning—numeric evidence—in the 1830s, around the time that New York reformers estimated that the city had 10,000 prostitutes. The forerunner of statistics was called "political arithmetic"; these studies—mostly attempts to calculate population size and life expectancy—emerged in seventeenth-century Europe, particularly in England and France. Analysts tried to count births, deaths, and marriages because they believed that a growing population was evidence of a healthy *state;* those who conducted such numeric studies—as well as other, nonquantitative analyses of social and political prosperity—came to be called *statists*. Over time, the statists' social research led to the new term for quantitative evidence: *statistics*.[4]

Early social researchers believed that information about society could help governments devise wise policies. They were well aware of the scientific developments of their day and, like other scientists, they came to value accuracy and objectivity. Counting—quantifying—offered a way of making their studies more precise, and let them concisely summarize lots of information.

Over time, social research became less theoretical and more quantitative. As the researchers collected and analyzed their data, they began to see patterns. From year to year, they discovered, the numbers of births, deaths, and even marriages remained relatively stable; this stability suggested that social arrangements had an underlying order, that what happened in a society depended on more than simply its government's recent actions, and analysts began paying more attention to underlying social conditions.

By the beginning of the nineteenth century, the social order seemed especially threatened: cities were larger than ever before; economies were beginning to industrialize; and revolutions in America and France had made it clear that political stability could not be taken for granted. The need for information, for facts that could guide social policy, was greater than ever before. A variety of government agencies began collecting and publishing statistics: the United States and several European countries began conducting regular censuses to collect population statistics; courts, prisons, and police began keeping track of the numbers of crimes and criminals; physicians kept records of patients; educators counted students; and so on. Scholars organized statistical societies to share the results of their studies and to discuss the best methods for gathering and interpreting statistics. And reformers who sought to confront the nineteenth-century's many social problems—the impoverished and the diseased, the fallen woman and the child laborer, the factory workforce and dispossessed agricultural labor—found statistics useful in

demonstrating the extent and severity of suffering. Statistics gave both government officials and reformers hard evidence—proof that what they said was true. Numbers offered a kind of precision: instead of talking about prostitution as a vaguely defined problem, reformers began to make specific, numeric claims (for example, that New York had 10,000 prostitutes).

During the nineteenth century, then, statistics—numeric statements about social life—became an authoritative way to describe social problems. There was growing respect for science, and statistics offered a way to bring the authority of science to debates about social policy. In fact, this had been the main goal of the first statisticians—they wanted to study society through counting and use the resulting numbers to influence social policy. They succeeded; statistics gained widespread acceptance as the best way to measure social problems. Today, statistics continue to play a central role in our efforts to understand these problems. But, beginning in the nineteenth century and continuing through today, social statistics have had two purposes, one public, the other often hidden. Their public purpose is to give an accurate, true description of society. But people also use statistics to support particular views about social problems. Numbers are created and repeated because they supply ammunition for political struggles, and this political purpose is often hidden behind assertions that numbers, simply because they are numbers, must be correct. People use statistics to support particular points of view, and it is naive simply to accept numbers as accurate, without examining who is using them and why.

CREATING SOCIAL PROBLEMS

We tend to think of social problems as harsh realities, like gravity or earthquakes, that exist completely independent of human action. But the very term reveals that this is incorrect: *social* problems are products of what people do.

This is true in two senses. First, we picture social problems as snarls or flaws in the social fabric. Social problems have their causes in society's arrangements; when some women turn to prostitution or some individuals have no homes, we assume that society has failed (although we may disagree over whether that failure involves not providing enough jobs, or not giving children proper moral instruction, or something else). Most people understand that social problems are social in this sense.

But there is a second reason social problems are social. Someone has to bring these problems to our attention, to give them names, describe their causes and characteristics, and so on. Sociologists speak of social problems being "constructed"— that is, created or assembled through the actions of activists, officials, the news media, and other people who draw attention to particular problems.[5] "Social problem" is a label we give to some social conditions, and it is that label that turns a condition we take for granted into something we consider troubling. This means that the processes of identifying and publicizing social problems are important. When we start thinking of prostitution or homelessness as a social problem, we are responding to campaigns by reformers who seek to arouse our concern about the issue.

The creation of a new social problem can be seen as a sort of public drama, a play featuring a fairly standard cast of characters. Often, the leading roles are played by *social activists*—individuals dedicated to promoting a cause, to making others aware of the problem. Activists draw attention to new social problems by holding protest demonstrations, attracting media coverage, recruiting new members to their cause, lobbying officials to do something about the situation, and so on. They are the most obvious, the most visible participants in creating awareness of social problems.

Successful activists attract support from others. The *mass media*—including both the press (reporters for newspapers or television news programs) and entertainment media (such as television talk shows)—relay activists' claims to the general public. Reporters often find it easy to turn those claims into interesting news stories; after all, a new social problem is a fresh topic, and it may affect lots of people, pose dramatic threats, and lead to proposals to change the lives of those involved. Media coverage, especially sympathetic coverage, can make millions of people aware of and concerned about a social problem. Activists need the media to provide that coverage, just as the media depend on activists and other sources for news to report.

Often activists also enlist the support of *experts*—doctors, scientists, economists, and so on—who presumably have special qualifications to talk about the causes and consequences of some social problem. Experts may have done research on the problem and can report their findings. Activists use experts to make claims about social problems seem authoritative, and the

mass media often rely on experts' testimonies to make news stories about a new problem seem more convincing. In turn, experts enjoy the respectful attention they receive from activists and the media.[6]

Not all social problems are promoted by struggling, independent activists; creating new social problems is sometimes the work of powerful organizations and institutions. *Government officials* who promote problems range from prominent politicians trying to arouse concern in order to create election campaign issues, to anonymous bureaucrats proposing that their agencies' programs be expanded to solve some social problem. And *businesses, foundations, and other private organizations* sometimes have their own reasons to promote particular social issues. Public and private organizations usually command the resources needed to organize effective campaigns to create social problems. They can afford to hire experts to conduct research, to sponsor and encourage activists, and to publicize their causes in ways that attract media attention.[7]

In other words, when we become aware of—and start to worry about—some new social problem, our concern is usually the result of efforts by some combination of *problem promoters*—activists, reporters, experts, officials, or private organizations—who have worked to create the sense that this is an important problem, one that deserves our attention. In this sense, people deliberately construct social problems.*

*I am not implying that there is anything wrong with calling attention to social problems. In fact, this book can be seen as my

Efforts to create or promote social problems, particularly when they begin to attract attention, may inspire opposition. Sometimes this involves officials responding to critics by defending existing policies as adequate. Recall that New York police minimized the number of prostitutes in the city, just as the Reagan administration argued that activists exaggerated the number of homeless persons. In other cases, opposition comes from private interests; for example, the Tobacco Institute (funded by the tobacco industry) became notorious for, over decades, challenging every research finding that smoking was harmful.

Statistics play an important role in campaigns to create—or defuse claims about—new social problems. Most often, such statistics describe the problem's size: there are 10,000 prostitutes in New York City, or three million homeless people. When social problems first come to our attention, perhaps in a televised news report, we're usually given an example or two (perhaps video footage of homeless individuals living on city streets) and then a statistical estimate (of the number of homeless people). Typically this is a big number. Big numbers warn us that the problem is a common one, compelling our attention, concern, and action. The media like to report statistics because numbers seem to be "hard facts"—little nuggets of indisputable truth. Activists trying to draw media attention to a new social problem often find that the press demands statistics: reporters insist on getting estimates of

---

effort to construct "bad statistics" as a problem that ought to concern people.

the problem's size—how many people are affected, how much it costs, and so on. Experts, officials, and private organizations commonly report having studied the problem, and they present statistics based on their research. Thus, the key players in creating new social problems all have reason to present statistics.

In virtually every case, promoters use statistics as ammunition; they choose numbers that will draw attention to or away from a problem, arouse or defuse public concern. People use statistics to support their point of view, to bring others around to their way of thinking. Activists trying to gain recognition for what they believe is a big problem will offer statistics that seem to prove that the problem is indeed a big one (and they may choose to downplay, ignore, or dispute any statistics that might make it seem smaller). The media favor disturbing statistics about big problems because big problems make more interesting, more compelling news, just as experts' research (and the experts themselves) seem more important if their subject is a big, important problem. These concerns lead people to present statistics that support their position, their cause, their interests. There is an old expression that captures this tendency: "Figures may not lie, but liars figure." Certainly we need to understand that people debating social problems choose statistics selectively and present them to support their points of view. Gun-control advocates will be more likely to report the number of children killed by guns, while opponents of gun control will prefer to count citizens who use guns to defend themselves from attack. Both numbers may be correct, but most people debating gun control present only the statistic that bolsters their position.[8]

THE PUBLIC AS AN INNUMERATE AUDIENCE

Most claims drawing attention to new social problems aim to persuade all of us—that is, the members of the general public. We are the audience, or at least one important audience, for statistics and other claims about social problems. If the public becomes convinced that prostitution or homelessness is a serious problem, then something is more likely to be done: officials will take action, new policies will begin, and so on. Therefore, campaigns to create social problems use statistics to help arouse the public's concern.

This is not difficult. The general public tends to be receptive to claims about new social problems, and we rarely think critically about social problems statistics. Recall that the media like to report statistics because numbers seem to be factual, little nuggets of truth. The public tends to agree; we usually treat statistics as facts.

In part, this is because we are innumerate. Innumeracy is the mathematical equivalent of illiteracy; it is "an inability to deal comfortably with the fundamental notions of number and chance."[9] Just as some people cannot read or read poorly, many people have trouble thinking clearly about numbers.

One common innumerate error involves not distinguishing among large numbers. A very small child may be pleased by the gift of a penny; a slightly older child understands that a penny or even a dime can't buy much, but a dollar can buy some things, ten dollars considerably more, and a hundred dollars a great deal (at least from a child's point of view). Most adults clearly grasp

what one can do with a hundred, a thousand, ten thousand, even one hundred thousand dollars, but then our imaginations begin to fail us. Big numbers blend together: a million, a billion, a trillion—what's the difference? They're all big numbers. (Actually, of course, there are tremendous differences. The difference between a million and a billion is the difference between one dollar and one thousand dollars; the difference between a million and a trillion is the difference between one dollar and a million dollars.)

Because many people have trouble appreciating the differences among big numbers, they tend to uncritically accept social statistics (which often, of course, feature big numbers). What does it matter, they may say, whether there are 300,000 homeless or 3,000,000?—either way, it's a big number. They'd never make this mistake dealing with smaller numbers; everyone understands that it makes a real difference whether there'll be three people or thirty coming by tomorrow night for dinner. A difference (thirty is ten times greater than three) that seems obvious with smaller, more familiar numbers gets blurred when we deal with bigger numbers (3,000,000 is ten times greater than 300,000). If society is going to feed the homeless, having an accurate count is just as important as it is for an individual planning to host three—or thirty—dinner guests.

Innumeracy—widespread confusion about basic mathematical ideas—means that many statistical claims about social problems don't get the critical attention they deserve. This is not simply because an innumerate public is being manipulated by advocates who cynically promote inaccurate statistics. Often,

statistics about social problems originate with sincere, well-meaning people who are themselves innumerate; they may not grasp the full implications of what they are saying. Similarly, the media are not immune to innumeracy; reporters commonly repeat the figures their sources give them without bothering to think critically about them.

The result can be a social comedy. Activists want to draw attention to a problem—prostitution, homelessness, or whatever. The press asks the activists for statistics—How many prostitutes? How many homeless? Knowing that big numbers indicate big problems and knowing that it will be hard to get action unless people can be convinced a big problem exists (and sincerely believing that there is a big problem), the activists produce a big estimate, and the press, having no good way to check the number, simply publicizes it. The general public—most of us suffering from at least a mild case of innumeracy—tends to accept the figure without question. After all, it's a big number, and there's no real difference among big numbers.

ORGANIZATIONAL PRACTICES
AND OFFICIAL STATISTICS

One reason we tend to accept statistics uncritically is that we assume that numbers come from experts who know what they're doing. Often these experts work for government agencies, such as the U.S. Bureau of the Census, and producing statistics is part of their job. Data that come from the government—crime rates,

unemployment rates, poverty rates—are *official statistics.*[10] There is a natural tendency to treat these figures as straightforward facts that cannot be questioned.

This ignores the way statistics are produced. All statistics, even the most authoritative, are created by people. This does not mean that they are inevitably flawed or wrong, but it does mean that we ought to ask ourselves just how the statistics we encounter were created.

Let's say a couple decides to get married. This requires going to a government office, taking out a marriage license, and having whoever conducts the marriage ceremony sign and file the license. Periodically, officials add up the number of marriage licenses filed and issue a report on the number of marriages. This is a relatively straightforward bit of recordkeeping, but notice that the accuracy of marriage statistics depends on couples' willingness to cooperate with the procedures. For example, imagine a couple who decide to "get married" without taking out a license; they might even have a wedding ceremony, yet their marriage will not be counted in the official record. Or consider couples that cohabit—live together—without getting married; there is no official record of their living arrangement. And there is the added problem of recordkeeping: is the system for filing, recording, and generally keeping track of marriages accurate, or do mistakes occur? These examples remind us that the official number of marriages reflects certain bureaucratic decisions about what will be counted and how to do the counting.

Now consider a more complicated example: statistics on suicide. Typically, a coroner decides which deaths are suicides. This

can be relatively straightforward: perhaps the dead individual left behind a note clearly stating an intent to commit suicide. But often there is no note, and the coroner must gather evidence that points to suicide—perhaps the deceased is known to have been depressed, the death occurred in a locked house, the cause of death was an apparently self-inflicted gunshot to the head, and so on. There are two potential mistakes here. The first is that the coroner may label a death a "suicide" when, in fact, there was another cause (in mystery novels, at least, murder often is disguised as suicide). The second possibility for error is that the coroner may assign another cause of death to what was, in fact, a suicide. This is probably a greater risk, because some people who kill themselves want to conceal that fact (for example, some single-car automobile fatalities are suicides designed to look like accidents so that the individual's family can avoid embarrassment or collect life insurance benefits). In addition, surviving family members may be ashamed by a relative's suicide, and they may press the coroner to assign another cause of death, such as accident.

In other words, official records of suicide reflect coroners' judgments about the causes of death in what can be ambiguous circumstances. The act of suicide tends to be secretive—it usually occurs in private—and the motives of the dead cannot always be known. Labeling some deaths as "suicides" and others as "homicides," "accidents," or whatever will sometimes be wrong, although we cannot know exactly how often. Note, too, that individual coroners may assess cases differently; we might imagine one coroner who is relatively willing to label deaths sui-

cides, and another who is very reluctant to do so. Presented with the same set of cases, the first coroner might find many more suicides than the second.[11]

It is important to appreciate that coroners view their task as classifying individual deaths, as giving each one an appropriate label, rather than as compiling statistics for suicide rates. Whatever statistical reports come out of coroners' offices (say, total number of suicides in the jurisdiction during the past year) are by-products of their real work (classifying individual deaths). That is, coroners are probably more concerned with being able to justify their decisions in individual cases than they are with whatever overall statistics emerge from those decisions.

The example of suicide records reveals that all official statistics are products—and often by-products—of decisions by various officials: not just coroners, but also the humble clerks who fill out and file forms, the exalted supervisors who prepare summary reports, and so on. These people make choices (and sometimes errors) that shape whatever statistics finally emerge from their organization or agency, and the organization provides a context for those choices. For example, the law requires coroners to choose among a specified set of causes for death: homicide, suicide, accident, natural causes, and so on. That list of causes reflects our culture. Thus, our laws do not allow coroners to list "witchcraft" as a cause of death, although that might be considered a reasonable choice in other societies. We can imagine different laws that would give coroners different arrays of choices: perhaps there might be no category for suicide; perhaps people who kill themselves might be considered ill, and their

deaths listed as occurring from natural causes; or perhaps suicides might be grouped with homicides in a single category of deaths caused by humans. In other words, official statistics reflect what sociologists call *organizational practices*—the organization's culture and structure shape officials' actions, and those actions determine whatever statistics finally emerge.

Now consider an even more complicated example. Police officers have a complex job; they must maintain order, enforce the law, and assist citizens in a variety of ways. Unlike the coroner who faces a relatively short list of choices in assigning cause of death, the police have to make all sorts of decisions. For example, police responding to a call about a domestic dispute (say, a fight between husband and wife) have several, relatively ill-defined options. Perhaps they should arrest someone; perhaps the wife wants her husband arrested—or perhaps she says she does not want that to happen; perhaps the officers ought to encourage the couple to separate for the night; perhaps they ought to offer to take the wife to a women's shelter; perhaps they ought to try talking to the couple to calm them down; perhaps they find that talking doesn't work, and then pick arrest or a shelter as a second choice; perhaps they decide that the dispute has already been settled, or that there is really nothing wrong. Police must make decisions about how to respond in such cases, and some—but probably not all—of those choices will be reflected in official statistics. If officers make an arrest, the incident will be recorded in arrest statistics, but if the officers decide to deal with the incident informally (by talking with the couple until they calm down), there may be no statistical record of what happens. The choices officers

make depend on many factors. If the domestic dispute call comes near the end of the officers' shift, they may favor quick solutions. If their department has a new policy to crack down on domestic disputes, officers will be more likely to make arrests. All these decisions, each shaped by various considerations, will affect whatever statistics eventually summarize the officers' actions.[12]

Like our earlier examples of marriage records and coroners labeling suicides, the example of police officers dealing with domestic disputes reveals that officials make decisions (relatively straightforward for marriage records, more complicated for coroners, and far less clear-cut in the case of the police), that official statistics are by-products of those decisions (police officers probably give even less thought than coroners to the statistical outcomes of their decisions), and that organizational practices form the context for those decisions (while there may be relatively little variation in how marriage records are kept, organizational practices likely differ more among coroners' offices, and there is great variation in how police deal with their complex decisions, with differences among departments, precincts, officers, and so on). In short, even official statistics are social products, shaped by the people and organizations that create them.

THINKING ABOUT STATISTICS AS SOCIAL PRODUCTS

The lesson should be clear: statistics—even official statistics such as crime rates, unemployment rates, and census counts—

are products of social activity. We sometimes talk about statistics as though they are facts that simply exist, like rocks, completely independent of people, and that people gather statistics much as rock collectors pick up stones. This is wrong. All statistics are created through people's actions: people have to decide what to count and how to count it, people have to do the counting and the other calculations, and people have to interpret the resulting statistics, to decide what the numbers mean. All statistics are social products, the results of people's efforts.

Once we understand this, it becomes clear that we should not simply accept statistics by uncritically treating numbers as true or factual. If people create statistics, then those numbers need to be assessed, evaluated. Some statistics are pretty good; they reflect people's best efforts to measure social problems carefully, accurately, and objectively. But other numbers are bad statistics—figures that may be wrong, even wildly wrong. We need to be able to sort out the good statistics from the bad. There are three basic questions that deserve to be asked whenever we encounter a new statistic.

1. *Who created this statistic?* Every statistic has its authors, its creators. Sometimes a number comes from a particular individual. On other occasions, large organizations (such as the Bureau of the Census) claim authorship (although each statistic undoubtedly reflects the work of particular people within the organization).

In asking who the creators are, we ought to be less concerned with the names of the particular individuals who produced a number than with their part in the public drama about statistics.

Does a particular statistic come from activists, who are striving to draw attention to and arouse concern about a social problem? Is the number being reported by the media in an effort to prove that this problem is newsworthy? Or does the figure come from officials, bureaucrats who routinely keep track of some social phenomenon, and who may not have much stake in what the numbers show?

2. *Why was this statistic created?* The identities of the people who create statistics are often clues to their motives. In general, activists seek to promote their causes, to draw attention to social problems. Therefore, we can suspect that they will favor large numbers, be more likely to produce them and less likely to view them critically. When reformers cry out that there are many prostitutes or homeless individuals, we need to recognize that their cause might seem less compelling if their numbers were smaller. On the other hand, note that other people may favor lower numbers. Remember that New York police officials produced figures showing that there were very few prostitutes in the city as evidence they were doing a good job. We need to be aware that the people who produce statistics often care what the numbers show, they use numbers as tools of persuasion.

3. *How was this statistic created?* We should not discount a statistic simply because its creators have a point of view, because they view a social problem as more or less serious. Rather, we need to ask how they arrived at the statistic. All statistics are imperfect, but some are far less perfect than others. There is a big difference between a number produced by a wild guess, and one generated through carefully designed research. This is the key

question. Once we understand that all social statistics are created by someone, and that everyone who creates social statistics wants to prove something (even if that is only that they are careful, reliable, and unbiased), it becomes clear that the methods of creating statistics are key. The remainder of this book focuses on this third question.

PLAN OF THE BOOK

The following chapters discuss some of the most common and important problems with the creation and interpretation of social statistics. Chapter 2 examines four basic sources of bad statistics: bad guesses, deceptive definitions, confusing questions, and biased samples. Chapter 3 looks at mutant statistics, at ways even good statistics can be mangled, misused, and misunderstood. Chapter 4 discusses the logic of statistical comparison and explores some of the most common errors in comparing two or more time periods, places, groups, or social problems. Chapter 5 considers debates over statistics. Finally, chapter 6 examines three general approaches to thinking about statistics.