Joe Edgin

THIRD EDITION

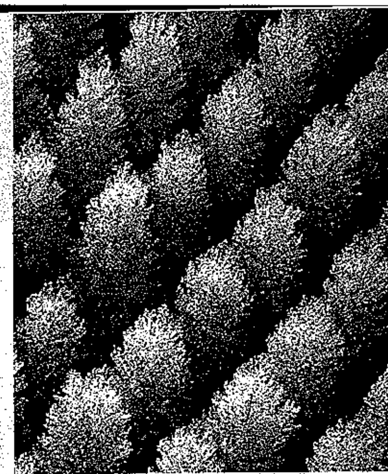
Seeing Through Statistics

Jessica M. Utts
University of California, Davis

THOMSON
BROOKS/COLE

Australia • Canada • Mexico • Singapore • Spain
United Kingdom • United States

CHAPTER 2



Reading the News

Thought Questions

1. Advice columnists sometimes ask readers to write and express their feelings about certain topics. For instance, Ann Landers once asked readers whether they thought engineers made good husbands. Do you think the responses are representative of public opinion? Explain why or why not.
2. Taste tests of new products are often done by having people taste both the new product and an old familiar standard. Do you think the results would be biased if the person handing the products to the respondents knew which was which? Explain why or why not.
3. Nicotine patches attached to the arm of someone who is trying to quit smoking dispense nicotine into the blood. Suppose you read about a study showing that nicotine patches were twice as effective in getting people to quit smoking as "control" patches (made to look like the real thing). Further, suppose you are a smoker trying to quit. What questions would you want answered about how the study was done and its results before you decided whether to try the patches yourself?
4. For a door-to-door survey on opinions about various political issues, do you think it matters who conducts the interviews? Give an example of how it might make a difference.

2.1 The Educated Consumer of Data

Pick up any newspaper or newsmagazine and you are almost certain to find a story containing conclusions based on data. Should you believe what you read? Not always. It depends on how the data were collected, measured, and summarized. In this chapter, we discuss seven critical components of statistical studies. We examine the kinds of questions you should ask before you believe what you read. We go into further detail about these issues in subsequent chapters. The goal in this chapter is to give you an overview of how to be a more educated consumer of the data you encounter in your everyday life.

What Are Data?

In statistical parlance, **data** is a plural word referring to a collection of numbers or other pieces of information to which meaning has been attached. For example, the numbers 1, 3, and 10 are not necessarily data, but they become so when we are told that these were the weight gains in grams of three of the infants in Salk's heartbeat study, discussed in Chapter 1. In Case Study 1.2, the data consisted of two pieces of information measured for each participant: (1) whether they took aspirin or a placebo, and (2) whether they had a heart attack.

Don't Always Believe What You Read

When you read the results of a study in the newspaper, you are rarely presented with the actual data. Someone has usually summarized the information for you, and he or she has probably already drawn conclusions and presented them to you. Don't always believe them. The meaning we can attach to data, and to the resulting conclusions, depends on how well the information was acquired and summarized.

In the remaining chapters of Part 1, we look at proper ways to obtain data. In Part 2, we turn our attention to how it should be summarized. In Part 4, we learn the power as well as the limitations of using the data collected from a sample to make conclusions about the larger population. In this chapter, we address seven features of statistical studies that you should think about when you read a news article. You will begin to be able to think critically and make your own conclusions about what you read.

2.2 Origins of News Stories

Where do news stories originate? How do reporters hear about events and determine that they are newsworthy? For stories based on statistical studies there are several possible sources. The two most common of these sources are also the most common outlets for researchers to present the results of their work: academic conferences and scholarly journals.

Every academic discipline holds conferences, usually annually, in which researchers can share their results with others. Reporters routinely attend these academic conferences and look for interesting news stories. For larger conferences, there is usually a "press room" where researchers can leave press releases for the media. If you pay attention, you will notice that in certain weeks of the year there will be several news stories about studies with related themes. For instance, the American Psychological Association meets in August, and there are generally some news stories emerging from results presented there. The American Association for the Advancement of Science meets in February, and news stories related to various areas of science will appear in the news that week.

One problem with news stories based on conference presentations is that there is unlikely to be a corresponding written report by the researchers, so it is difficult for readers of the news story to obtain further information. News stories based on conference reports generally mention the name and date of the conference as well as the name and institution of the lead researcher, so sometimes it is possible to contact the researcher for further information. Some researchers make conference presentations available on their Web sites.

In contrast, many news stories about statistical studies are based on published articles in scholarly journals. Reporters routinely read these journals when they are published, or they get advance press releases from the journal offices. News stories based on journal articles usually mention the journal and date of publication, so if you are interested in learning more about the study, you can obtain the original journal article. Journal articles are sometimes available on the journal's Web site or on the Web site of the author(s). You can also write to the lead author and request that a "reprint" be sent to you.

As a third source of news stories about statistical studies, some government and private agencies release in-depth research reports. Unlike journal articles, these reports are not necessarily "peer-reviewed" or checked by neutral experts on the topic. An advantage of these reports is that they are not restricted by space limitations imposed by journals and often provide much more in-depth information than do journal articles.

A supplementary source from which news stories may originate is a university media office. Most research universities have an office that provides press releases when faculty members have completed research that may be of interest to the public. The timing of these news releases usually corresponds to a presentation at an academic conference or publication of results in an academic journal, but the news release summarizes the information so that journalists don't have to be as versed in the technical aspects of the research to write a good story. When you read about a study in the news and would like more information, the news office of the lead researcher's institution is a good place to start looking. They may have issued a press release on which the story was based.

News Stories and Original Sources in the Appendix and on the CD

To illustrate how the concepts in this book are used in research and eventually converted into news stories, there is a collection of examples included with this book. In each case, the example includes a story from a newspaper, magazine, or Web site, and these are printed in the Appendix and on the CD accompanying the book. Sometimes there is also a press release. These are provided as an additional "News Story" and included on the CD. Most of the news stories are based on articles from scholarly journals or detailed reports. Many of these articles are printed in full on the CD, labeled as the "Original Source." Throughout this book you will find a CD icon when you need to refer to the material on the CD. By comparing the news story and the original source, you will learn how to evaluate what is reported in the news.

2.3 How to Be a Statistics Sleuth: Seven Critical Components

Reading and interpreting the results of surveys or experiments is not much different from reading and interpreting the results of other events of interest, such as sports competitions or criminal investigations. If you are a sports fan, then you know what information should be included in reports of competitions and you know when crucial information is missing. If you have ever been involved in an event that was later reported in the newspaper, you know that missing information can lead readers to erroneous conclusions.

In this section, you are going to learn what information should be included in news reports of statistical studies. Unfortunately, crucial information is often missing. With some practice, you can learn to figure out what's missing, as well as how to interpret what's reported. You will no longer be at the mercy of someone else's conclusions. You will be able to determine them for yourself. To provide structure to our examination of news reports, let's list Seven Critical Components that determine the soundness of statistical studies. A good news report should provide you with information about all of the components that are relevant to that study.

Component 1: The *source* of the research and of the *funding*.

Component 2: The *researchers* who had *contact* with the participants.

Component 3: The *individuals* or objects studied and how they were *selected*.

Component 4: The exact nature of the *measurements* made or *questions* asked.

Component 5: The *setting* in which the measurements were taken.

Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest.

Component 7: The *extent* or *size* of any claimed effects or differences.

Before delving into some examples, let's examine each component more closely. You will find that most of the problems with studies are easy to identify. Listing these components simply provides a framework for using your common sense.

Component 1: The *source* of the research and of the *funding* Studies are conducted for three major reasons. First, governments and private companies need to have data in order to make wise policy decisions. Information such as unemployment rates and consumer spending patterns are measured for this reason. Second, researchers at universities and other institutions are paid to ask and answer interesting questions about the world around us. The curious questioning and experimentation of such researchers have resulted in many social, medical, and scientific advances. Much of this research is funded by government agencies, such as the National Institutes of Health. Third, companies want to convince consumers that their programs and products work better than the competition's, or special-interest groups want to prove that their point of view is held by the majority.

Unfortunately, it is not always easy to discover who funded research. Many university researchers are now funded by private companies. In her book *Tainted Truth* (1994), Cynthia Crossen warns us:

> *Private companies, meanwhile, have found it both cheaper and more prestigious to retain academic, government, or commercial researchers than to set up in-house operations that some might suspect of fraud. Corporations, litigants, political candidates, trade associations, lobbyists, special interest groups—all can buy research to use as they like.* (p. 19)

If you discover that a study was funded by an organization that would be likely to have a strong preference for a particular outcome, it is especially important to be sure that correct scientific procedures were followed. In other words, be sure the remaining components have sound explanations.

Component 2: The *researchers* who had *contact* with the participants It is important to know who actually had contact with the participants and what message those people conveyed. Participants often give answers or behave in ways to comply with the desires of the researchers. Consider, for example, a study done at a shopping mall to compare a new brand of a certain product to an old familiar brand. Shoppers are asked to taste each brand and state their preference. It is crucial that both the person presenting the two brands and the respondents be kept entirely blind as to which is which until after the preferences have been selected. Any clues might bias the respondent to choose the old familiar brand. Or, if the interviewer is clearly eager to have them choose one brand over the other, the respondents will most likely oblige in order to please. As another example, if you discovered that a study on the prevalence of illegal drug use was conducted by sending uniformed police officers door-to-door, you would probably not have much faith in the results. We will discuss other ways in which researchers influence participants in Chapters 4 and 5.

Component 3: The *individuals* or objects studied and how they were *selected* It is important to know to whom the results can be extended. In general, the results of a study apply only to individuals similar to those in the study. For example, until recently, many medical studies included men only, so the results were of little value to women. When determining who is similar to those in the study, it is also important to know how participants were enlisted for the study. Many studies rely on volunteers recruited through the newspaper, who are usually paid a small amount for their participation. People who would respond to such recruitment efforts may differ in relevant ways from those who would not. Surveys relying on voluntary responses are likely to be biased because only those who feel strongly about the issues are likely to respond. For instance, some Web sites have a "question of the day" to which people are asked to voluntarily respond by clicking on their preferred answer. Only those who have strong opinions are likely to participate, so the results cannot be extended to any larger group.

Component 4: The exact nature of the *measurements* made or *questions* asked As you will see in Chapter 3, precisely defining and measuring most of the things researchers study isn't easy. For example, if you wanted to measure whether people "eat breakfast," how would you do so? What if they just have juice? What if they work until midmorning and then eat a meal that satisfies them until dinner? You need to understand exactly what the various definitions mean when you read about someone else's measurements.

In polls and surveys, the "measurements" are usually answers to specific questions. Both the wording and the ordering of the questions can influence answers. For example, a question about "street people" would probably elicit different responses than a question about "families who have no home." Ideally, you should be given the exact wording that was used in a survey or poll.

Component 5: The *setting* in which the measurements were taken The setting in which measurements were taken includes factors such as when and where they were taken and whether respondents were contacted by phone, mail, or in person. A study can be easily biased by timing. For example, opinions on whether criminals should be locked away for life may change drastically following a highly publicized murder or kidnapping case. If a study is conducted by telephone and calls are made only in the evening, certain groups of people would be excluded, such as those who work the evening shift or who routinely eat dinner in restaurants.

Where the measurements were taken can also influence the results. Questions about sensitive topics, such as sexual behavior or income, might be more readily answered over the phone, where respondents feel more anonymous. Sometimes research is done in a laboratory or university office, and the results may not readily extend to a natural setting. For example, studies of communication between two people are sometimes done by asking them to conduct a conversation in a university office with a tape recorder present. Such conditions almost certainly produce more limited conversation than would occur in a more natural setting.

Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest If two or more groups are being compared on a factor of interest, it is important to consider other ways in which the groups may differ that might influence the comparison. For example, suppose researchers want to know if smoking marijuana is related to academic performance. If the group of people who smoke marijuana has lower test scores than the group of people who don't, researchers may want to conclude that the lower test scores are due to smoking marijuana. Often, however, other disparities in the groups can explain the observed difference just as well. For example, people who smoke marijuana may simply be the type of people who are less motivated to study and thus would score lower on tests whether they smoked or not. Reports of research should include an explanation of any such possible differences that might account for the results. We will explore the issue of these kinds of extraneous factors, and how to control for them, in much more detail in Chapter 5.

Component 7: The *extent* or *size* of any claimed effects or differences Media reports about statistical studies often fail to tell you how large the observed effects were. Without that knowledge, it is hard for you to assess whether you think the results are of any practical importance. For example, if, based on Case Study 1.2, you were told simply that taking aspirin every other day reduced the risk of heart attacks, you would not be able to determine whether it would be worthwhile to take aspirin. You should instead be told that for the men in the study, the rate was reduced from about 17 heart attacks per 1000 participants without aspirin to about 9.4 heart attacks per 1000 with aspirin. Often news reports simply report that a treatment had an effect or that a difference was observed, but don't tell you the size of the difference or effect. We will investigate this issue in great detail in Part 4 of this book.

2.4 Four Hypothetical Examples of Bad Reports

Throughout this book, you will see numerous examples of real studies and news reports. So that you can get some practice finding problems without having to read unnecessarily long news articles, let's examine some hypothetical reports. These are admittedly more problematic than many real reports because they serve to illustrate several difficulties at once.

Hypothetical News Article 1

Study Shows Psychology Majors Are Smarter than Chemistry Majors

A fourth-year psychology student, for her senior thesis, conducted a study to see if students in her major were smarter than those majoring in chemistry. She handed out questionnaires in five advanced psychology classes and five advanced chemistry labs. She asked the students who were in class to record their grade-point averages (GPAs) and their majors. Using the data only from those who were actually majors in these fields in each set of classes, she found that the psychology majors had an average GPA of 3.05, whereas the chemistry majors had an average GPA of only 2.91. The study was conducted last Wednesday, the day before students were home enjoying Thanksgiving dinner.

Read each article and see if your common sense gives you some reasons why the headline is misleading. Then proceed to read the commentary about the Seven Critical Components.

Hypothetical News Article 1: "Study Shows Psychology Majors Are Smarter than Chemistry Majors"

Component 1: The *source* of the research and of the *funding* The study was a senior thesis project conducted by a psychology major. Presumably, it was cheap to run and was paid for by the student. One could argue that she would have a reason to want the results to come out as they did, although with a properly conducted study, the motives of the experimenter should be minimized. As we shall see, there were additional problems with this study.

Component 2: The *researchers* who had *contact* with the participants Presumably, only the student conducting the study had contact with the respondents. Crucial missing information is whether she told them the purpose of the study. Even if she did not tell them, many of the psychology majors may have known her and known what she was doing. Any clues as to desired outcomes on the part of experimenters can bias the results.

Component 3: The *individuals* or objects studied and how they were *selected* The individuals selected are the crux of the problem here. The measurements were taken on advanced psychology and chemistry students, which would have been fine if they had been sampled correctly. However, only those who were in the psychology classes or in the chemistry labs that day were actually measured. Less conscientious students are more likely to leave early before a holiday, but a missed class is probably easier to make up than a missed lab. Therefore, perhaps a larger proportion of the students with low grade-point averages were absent from the psychology classes than from the chemistry labs. Due to the missing students, the investigator's results would overestimate the average GPA for psychology students more so than for chemistry students.

Component 4: The exact nature of the *measurements* made or *questions* asked Students were asked to give a "self-report" of their grade-point averages. A more accurate method would have been to obtain this information from the registrar at the university. Students may not know their exact grade-point average. Also, one group may be more likely to know the exact value than the other. For example, if many of the chemistry majors were planning to apply to medical school in the near future, they may be only too aware of their grades. Further, the headline implies that GPA is a measure of intelligence. Finally, the research assumes that GPA is a standard measure. Perhaps grading is more competitive in the chemistry department.

Component 5: The *setting* in which the measurements were taken Notice that the article specifies that the measurements were taken on the day before a major holiday. Unless the university consisted mainly of commuters, many students may have left early for the holiday, further aggravating the problem that the students with lower grades were more likely to be missing from the psychology classes than from the chemistry labs. Further, because students turned in their questionnaires anonymously, there was presumably no accountability for incorrect answers.

Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest The factor of interest is the student's major, and the two groups being compared are psychology majors and chemistry majors. This component considers whether the students who were interviewed for the study may differ in ways other than their choice of major. It is difficult to know what differences might exist without knowing more about the particular university. For example, because psychology is such a popular major, at some universities students are required to have a certain GPA before they are admitted to the major. A university with a separate premedical major might have the best of the science students enrolled in that major instead of chemistry. Those kinds of extraneous factors would be relevant to interpreting the results of the study.

Component 7: The *extent* or *size* of any claimed effects or differences The news report does present this information, by noting that the average GPAs for the two groups were 3.05 and 2.91. Additional useful information would be to know how many students were included in each of the averages given, what percentage of all students in each major were represented in the sample, and how much variation there was among GPAs within each of the two groups.

Hypothetical News Article 2: "Per Capita Income of U.S. Shrinks Relative to Other Countries"

Component 1: The *source* of the research and of the *funding* We are told nothing except the name of the group that conducted the study, which should be fair warning. Being called "an independent research group" in the story does not mean

Hypothetical News Article 2

Per Capita Income of U.S. Shrinks Relative to Other Countries

An independent research group, the Institute for Foreign Investment, has noted that the per capita income of Americans has been shrinking relative to some other countries. Using per capita income figures from the *World Almanac* and exchange rates from last Friday's financial pages, the organization warned that per capita income for the United States has risen only 10% during the past 5 years, whereas per capita income for certain other countries has risen 50%. The researchers concluded that more foreign investment should be allowed in the United States to bolster the sagging economy.

that it is an unbiased research group. In fact, the last line of the story illustrates the probable motive for their research.

Component 2: The *researchers* who had *contact* with the participants This component is not relevant because there were no participants in the study.

Component 3: The *individuals* or objects studied and how they were *selected* The objects in this study were the countries used for comparison with the United States. We should have been told which countries were used, and why.

Component 4: The exact nature of the *measurements* made or *questions* asked This is the major problem with this study. First, as mentioned, we are not even told which countries were used for comparison. Second, current exchange rates but older per capita income figures were used. If the rate of inflation in a country had recently been very high, so that a large rise in per capita income did not reflect a concomitant rise in spending power, then we should not be surprised to see a large increase in per capita income in terms of actual dollars. In order to make a valid comparison, all figures would have to be adjusted to comparable measures of spending power, taking inflation into account. We will learn how to do that in Chapter 14.

Components 5, 6, and 7: The *setting* in which the measurements were taken. *Differences* in the groups being compared, *in addition* to the factor of interest. The *extent* or *size* of any claimed effects or differences These issues are not relevant here, except as they have already been discussed. For example, although the size of the difference between the United States and the other countries is reported, it is meaningless without an inflation adjustment.

Hypothetical News Article 3

Researchers Find Drug to Cure Excessive Barking in Dogs

Barking dogs can be a real problem, as anyone who has been kept awake at night by the barking of a neighbor's canine companion will know. Researchers at a local university have tested a new drug that they hope will put all concerned to rest. Twenty dog owners responded to a newspaper article asking for volunteers with problem barking dogs to participate in a study. The dogs were randomly assigned to two groups. One group of dogs was given the drug, administered as a shot, and the other dogs were not. Both groups were kept overnight at the research facility and frequency of barking was observed. The researchers deliberately tried to provoke the dogs into barking by doing things like ringing the doorbell of the facility and having a mail carrier walk up to the door. The two groups were treated on separate weekends because the facility was only large enough to hold ten dogs. The researchers left a tape recorder running and measured the amount of time during which any barking was heard. The dogs who had been given the drug spent only half as much time barking as did the dogs in the control group.

Hypothetical News Article 3: "Researchers Find Drug to Cure Excessive Barking in Dogs"

Component 1: The *source* of the research and of the *funding* We are not told why this study was conducted. Presumably it was because the researchers were interested in helping to solve a societal problem, but perhaps not. It is not uncommon for drug companies to fund research to test a new product or a new use for a current product. If that were the case, the researchers would have added incentive for the results to come out favorable to the drug. If everything were done correctly, such an incentive wouldn't be a major factor; however, when research is funded by a private source, that information should be announced when the results are announced.

Component 2: The *researchers* who had *contact* with the participants We are not given any information about who actually had contact with the dogs. One important question is whether the same handlers were used with both groups of dogs. If not, the difference in handlers could explain the results. Further, we are not told whether the dogs were primarily left alone or were attended most of the time. If researchers were present most of the time, their behavior toward the dogs could have had a major impact on the amount of barking.

Component 3: The *individuals* or objects studied and how they were *selected* We are told that the study used dogs whose owners volunteered them as problem dogs for the study. Although the report does not mention payment, it is quite common for volunteers to receive monetary compensation for their participation. The volunteers presumably lived in the area of the university. The dog owners had to be willing to be separated from their pets for the weekend. These and other factors mean that the owners and dogs who participated may differ from the general population. Further, the initial reasons for the problem behavior may vary from one participant to the next, yet the dogs were measured together. Therefore, there is no way to ascertain if, for example, dogs who bark only because they are lonely would be helped. In any case, we cannot extend the results of this study to conclude that the drug would work similarly on all dogs or even on all problem dogs. Because the dogs were randomly assigned to the two groups—and if there were no other problems—we would be able to extend the results to all dogs similar to those who participated.

Component 4: The exact nature of the *measurements* made or *questions* asked The researchers measured each group of dogs as a group, by listening to a tape and recording the amount of time during which there was any barking. Because dogs are quite responsive to group behavior, one barking dog could set the whole group barking for a long time. Therefore, just one particularly obnoxious dog in the control group alone could explain the results. It would have been better to separate the dogs and measure each one individually.

Component 5: The *setting* in which the measurements were taken The groups were measured on separate weekends. This creates another problem. First, the researchers knew which group was which and may have unconsciously provoked the control group slightly more than the group receiving the drug. Further, conditions differed over the two weeks. Perhaps it was sunny one weekend and raining the next, or there were other subtle differences, such as more traffic one weekend than the next, small planes overhead, and so on. All of these could change the behavior of the dogs but might go unnoticed or unreported by the experimenters.

The measurements were also taken outside of the dogs' natural environments. The dogs in the experimental group in particular would have reason to be upset because they were first given a shot and then put together with nine other dogs in the research facility. It would have been better to put them back into their natural environment because that's where the problem barking was known to occur.

Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest The dogs were randomly assigned to the two groups (drug or no drug), which should have minimized overall differences in size, temperament, and so on for the dogs in the two groups. However, differences were induced between the two groups by the way the experiment was conducted. Recall that the groups were measured on different weekends—this could have created the difference in behavior. Also, the drug-treated dogs were given a shot to administer the drug, whereas the control group was given no shot. It could be that the very act of getting a shot made the drug group lethargic. A better design would have been to administer a placebo shot—that is, a shot with an inert substance—to the control group.

Hypothetical News Article 4

Survey Finds Most Women Unhappy in Their Choice of Husbands

A popular women's magazine, in a survey of its subscribers, found that over 90% of them are unhappy in their choice of whom they married. Copies of the survey were mailed to the magazine's 100,000 subscribers. Surveys were returned by 5000 readers. Of those responding, 4520, or slightly over 90%, answered no to the question: "If you had it to do over again, would you marry the same man?" To keep the survey simple so that people would return it, only two other questions were asked. The second question was, "Do you think being married is better than being single?" Despite their unhappiness with their choice of spouse, 70% answered yes to this. The final question, "Do you think you will outlive your husband?" received a yes answer from 80% of the respondents. Because women generally live longer than men, and tend to marry men somewhat older than themselves, this response was not surprising. The magazine editors were at a loss to explain the huge proportion of women who would choose differently. The editor could only speculate: "I guess finding Mr. Right is much harder than anyone realized."

Component 7: The *extent* or *size* of any claimed effects or differences We are told only that the treated group barked half as much as the control group. We are not told how much time either group spent barking. If one group barked 8 hours a day but the other group only 4 hours a day, that would not be a satisfactory solution to the problem of barking dogs.

Hypothetical News Article 4: "Survey Finds Most Women Unhappy in Their Choice of Husbands"

Components 1 through 7 We don't even need to consider the details of this study because it contains a fatal flaw from the outset. The survey is an example of what is called a "volunteer sample" or a "self-selected sample." Of the 100,000 who received the survey, only 5% responded. The people who are most likely to respond to such a survey are those who have a strong emotional response to the question. In this case, it would be women who are unhappy with their current situation who would probably respond. Notice that the other two questions are more general and thus not likely to arouse much emotion either way. Thus, it is the strong reaction to the first question that would drive people to respond. The results would certainly not be representative of "most women" or even of most subscribers to the magazine.

CASE STUDY 2.1 Who Suffers from Hangovers?

SOURCE: News Story 2 in the Appendix and Original Source 2 on the CD.

Read News Story 2 in the Appendix, "Research shows women harder hit by hangovers" and access the original source of the story on the CD, the journal article "Development and Initial Validation of the Hangover Symptoms Scale: Prevalence and Correlates of Hangover Symptoms in College Students." Let's examine the seven critical components based on the news story and, where necessary, additional information provided in the journal article.

Component 1: The *source* of the research and of the *funding* The news story covers this aspect well. The researchers were "a team at the University of Missouri-Columbia" and the study was "supported by the National Institutes of Health."

Component 2: The *researchers* who had *contact* with the participants This aspect of the study is not clear from the news article, which simply mentions, "The researchers asked 1,230 drinking college students . . . ." However, the journal article says that the participants were enrolled in Introduction to Psychology courses and were asked to fill out a questionnaire. So it can be assumed that professors or research assistants in psychology had contact with the participants.

Component 3: The *individuals* or objects studied and how they were *selected* The news story describes the participants as "1,230 drinking college students, only 5 percent of whom were of legal drinking age." The journal article provides much more information, including the important fact that the participants were all enrolled in introductory psychology classes and were participating in the research to fulfill a requirement for the course. The reader must decide whether this group of participants is likely to be representative of all drinking college students, or some larger population, for severity of hangover symptoms. The journal article also provides information on the sex, ethnicity, and age of participants.

Component 4: The exact nature of the *measurements* made or *questions* asked The news story provides some detail about what was asked, noting that the participants were asked *"to describe how often they experienced any of 13 symptoms after drinking. The symptoms ranged from headaches and vomiting to feeling weak and unable to concentrate."* The journal article again provides much more detail, listing the 13 symptoms and explaining that participants were asked to indicate how often they were experienced on a 5-point scale (p. 1444 of the journal article). Further, participants were asked to provide a "hangover count" in which they noted how many times they had experienced at least one of the 13 symptoms in the past year, using a 5-point scale. This scale ranged from "never" to "52 times or more." Additional questions were asked about alcoholism in the participant's family and early experience with alcohol. Detailed information about all of these questions is included in the journal article.

Component 5: The *setting* in which the measurements were taken This information is not provided explicitly, but it can be assumed that measurements were taken in the Psychology Department at the University of Columbia-Missouri. One missing fact that may be helpful in interpreting the results is if the questions were administered to a large group of students at once, or individually, and whether students could be identified when the researchers read their responses.

Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest The purpose of the research was to develop and test a "Hangover Symptoms Scale" but two interesting differences in groups emerged when the researchers made comparisons. The groups being compared in the first instance were males and females; thus, Male/Female was the factor of interest. The researchers found that females suffered more from hangovers. This component is asking if there may be other differences between males and females, other than "Male" and "Female" that could help account for the difference. One possibility mentioned in the news article is body weight. Males tend to weigh more than females on average. An interesting question, not answered by the research, is if a group of males and females of the same weight, say 130 pounds, were to consume the same amount of alcohol, would the females suffer more hangover symptoms? The difference in weight between the two groups is in addition to the factor of interest, which is Male/Female. It may be the weight difference, and not the sex difference, that accounts for the difference in hangover severity.

The other comparison mentioned in the news article is between students who had alcohol-related problems, or whose biological parents had such problems, and students who did not have that history. In this case, the alcohol-related problems (of the student or parents) is the factor of interest. However, you can probably think of other differences in the two groups (those with problems and those without) that may help account for the difference in hangover severity between the two groups. For instance, students with a history of problems may not have as healthful diets in the past or present as students without such problems, and that may contribute to hangover severity. So the comparison of interest, between those with an alcohol problem in their background and those without, may be complicated by other differences in these two groups.

Component 7: The *extent* or *size* of any claimed effects or differences The news story does not report how much difference in hangover severity was found between men and women, or between those with and without a history of alcohol problems. Reading the journal article may explain why this is so—the article itself does not report a simple difference. In fact, simple comparisons don't yield much difference; for instance, 11% of men and 14% of women never experienced any hangover symptoms in the previous year. Differences only emerged when complicating factors such as amount of alcohol consumed were factored in. The researchers report, "After controlling for the frequency of drinking and getting drunk and for the typical quantity of alcohol consumed when drinking, women were significantly more likely than men to experience at least one of the hangover symptoms" (p. 1446). The article does not elaborate, such as explaining what would be the difference for a male and female who drank the same amount and equally often.

2.5 Planning Your Own Study: Defining the Components in Advance

Although you may never have to design your own survey or experiment, it will help you understand how difficult it can be if we illustrate the Seven Critical Components for a very simple hypothetical study you might want to conduct. Suppose you are interested in determining which of three local supermarkets has the best prices so you can decide where to shop. Because you obviously can't record and summarize the prices for all available items, you would have to use some sort of sample.

To obtain meaningful data, you would need to make many decisions. Some of the components need to be reworded because they are being answered in advance of the study, and obviously not all of the components are relevant for this simple example. However, by going through them for such a simple case, you can see how many ambiguities and decisions can arise when designing a study.

Component 1: The *source* of the research and of the *funding* Presumably you would be funding the study yourself, but before you start you need to decide why you are doing the study. Are you only interested in items you routinely buy, or are you interested in comparing the stores on the multitude of possible items?

Component 2: The *researchers* who had *contact* with the participants In this example, the question would be who is going to visit the stores and record the prices. Will you personally visit each store and record the prices? Will you send friends to two of the stores and visit the third yourself? If you use other people, you would need to train them so there would be no ambiguities.

Component 3: The *individuals* or objects studied and how they were *selected* In this case, the "objects studied" are items in the grocery store. The correct question is, "On what items should prices be recorded?" Do you want to use exactly the same items at all stores? What if one store offers its own brand but another only offers name brands? Do you want to choose a representative sampling of items you are likely to buy or choose from all possible items? Do you want to include nonfood items? How many items should you include? How should you choose which ones to select? If you are simply trying to minimize your own shopping bill, it is probably best to list the 20 or 30 items you buy most often. However, if you are interested in sharing your results with others, you might prefer to choose a representative sample of items from a long list of possibilities.

Component 4: The exact nature of the *measurements* made or *questions* asked You may think that the cost of an item in a supermarket is a well-defined measurement. But if a store is having a sale on a particular item on your list, should you use the sale price or the regular price? Should you use the price of the smallest possible size of the product? The largest? What if a store always has a sale on one brand or another of something, such as laundry soap, and you don't really care which brand you buy? Should you then record the price of the brand on sale that week? Should you record the prices listed on the shelves or actually purchase the items and see if the prices listed were accurate?

Component 5: The *setting* in which the measurements were taken When will you conduct the study? Supermarkets in university towns may offer sale prices on items typically bought by students at certain times of the year—for example, just after students have returned from vacation. Many stores also offer sale items related to certain holidays, such as ham or turkey just before Christmas or eggs just before Easter. Should you take that kind of timing into account?

Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest The groups being compared are the groups of items from the three stores. There should be no additional differences related to the direct costs of the items. However, if you were conducting the study in order to minimize your shopping costs, you might ask if there are hidden costs for shopping at one store versus another. For example, do you always have to wait in line at one store and not at another, and should you therefore put a value on your time? Does one store make mistakes at the cash register more often than another? Does one store charge a higher fee to use your cash card for payment? Does it cost more to drive to one store than another?

Component 7: The *extent* or *size* of any claimed effects or differences This component should enter into your decision about where to shop after you have finished the study. Even if you find that items in one store cost less than in another, the amount of the difference may not convince you to shop there. You would probably want to figure out approximately how much shopping in a particular store would save you over the course of a year. You can see why knowing the amount of a difference found in a study is an important component for using that study to make future decisions.

CASE STUDY 2.2 Brooks Shoes Brings Flawed Study to Court

SOURCE: Gastwirth (1988) pp. 517–520.

In 1981, Brooks Shoe Manufacturing Company sued Suave Shoe Corporation for manufacturing shoes incorporating a "V" design used in Brooks's athletic shoes. Brooks claimed that the design was an unregistered trademark that people used to identify Brooks shoes. According to Gastwirth (1988, p. 517), it was the role of the court to determine "the distinctiveness or strength of the mark as well as its possible secondary meaning (similarity of product or mark might confuse prospective purchasers of the source of the item)."

To show that the design had "secondary meaning" to buyers, Brooks conducted a survey of 121 spectators and participants at three track meets. Interviewers approached people and asked them a series of questions that included showing them a Brooks shoe with the name masked and asking them to identify it. Of those surveyed, 71% were able to identify it as a Brooks shoe, and 33% of those people said it was because they recognized the "V." When shown a Suave shoe, 39% of them thought it was a Brooks shoe, with 48% of those people saying it was because of the "V" design on the Suave shoe. Brooks Company argued that this was sufficient evidence that people might be confused and think Suave shoes were manufactured by Brooks.

Suave had a statistician as an expert witness who pointed out a number of flaws in the Brooks survey. Let's examine them using the Seven Critical Components as a guide. First, the survey was funded and conducted by Brooks, and the company's lawyer was instrumental in designing it. Second, the court determined that the interviewers who had contact with the respondents were inadequately trained in how to conduct an unbiased survey. Third, the individuals asked were not selected to be representative of the general public in the area (Baltimore/Washington, D.C.). For example, 78% had some college education, compared with 18.4% in Baltimore and 37.7% in Washington, D.C. Further, the settings for the interviews were track meets, where people were likely to be more familiar with athletic shoes. The questions asked were biased. For example, the exact wording used when a person was handed the shoes was: "I am going to hand you a shoe. Please tell me what brand you think it is." The way the question is framed would presumably lead respondents to think the shoe has a well-known brand name. Later in the questioning, respondents were asked, "How long have you known about Brooks Running Shoes?" Because of the setting, respondents could have informed others at the track meet that Brooks was probably conducting the survey, and those informed could have subsequently been interviewed.

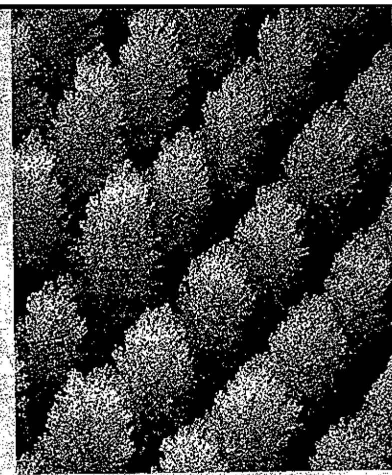
Suave introduced its own survey conducted on 404 respondents properly sampled from the population of all people who had purchased any type of athletic shoe during the previous year. Of those, only 2.7% recognized a Brooks shoe on the basis of the "V" design. The combination of the poor survey methods by Brooks and the proper survey by Suave convinced the court that the public did not make enough of an association between Brooks and the "V" design to allow Brooks to claim legal rights to the design.

Exercises

Asterisked (*) exercises are included in the Solutions at the back of the book.

1. Suppose that a television network wants to know how daytime television viewers feel about a new soap opera the network is broadcasting. A staff member suggests that just after the show ends they give two phone numbers, one for viewers to call if they like the show and the other to call if they don't. Give two reasons why this method would not produce the desired information. (*Hint:* The network is interested in all daytime television viewers. Who is likely to be watching just after the show, and who is likely to call in?)
2. The April 24, 1997, issue of "UCDavis Lifestyle Newstips" reported that a professor of veterinary medicine was conducting a study to see if a drug called clomipramine, an anti-anxiety medication used for humans, could reduce "canine aggression toward family members." The newsletter said, "Dogs demon-

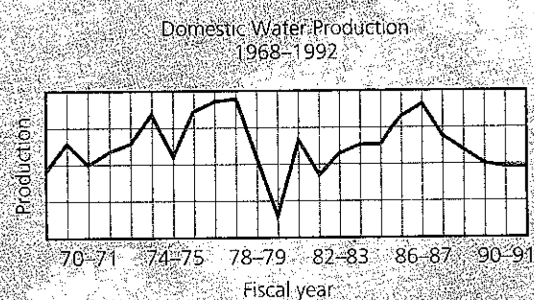
CHAPTER 9



Plots, Graphs, and Pictures

Thought Questions

1. You have seen pie charts and bar graphs and should have some rudimentary idea of how to construct them. Suppose you have been keeping track of your living expenses and find that you spend 50% of your money on rent, 25% on food, and 25% on other expenses. Draw a pie chart and a bar graph to depict this information. Discuss which is more visually appealing and useful.
2. Here is an example of a plot that has some problems. Give two reasons why this is not a good plot.



3. Suppose you had a set of data representing two measurement variables—namely, height and weight—for each of 100 people. How could you put that information into a plot, graph, or picture that illustrated the relationship between the two measurements for each person?
4. Suppose you own a company that produces candy bars and you want to display two graphs. One graph is for customers and shows the price of a candy bar for each of the past 10 years. The other graph is for stockholders and shows the amount the company was worth for each of the past 10 years. You decide to adjust the dollar amounts in one graph for inflation but to use the actual dollar amounts in the other graph. If you were trying to present the most favorable story in each case, which graph would be adjusted for inflation? Explain.

9.1 Well-Designed Statistical Pictures

There are many ways to present data in pictures. The most common are plots and graphs, but sometimes a unique picture is used to fit a particular situation. The purpose of a plot, graph, or picture of data is to give you a visual summary that is more informative than simply looking at a collection of numbers. Done well, a picture can quickly convey a message that would take you longer to find if you had to study the data on your own. Done poorly, a picture can mislead all but the most observant of readers. Here are some basic characteristics that all plots, graphs, and pictures should exhibit:

1. The data should stand out clearly from the background.
2. There should be clear labeling that indicates
   a. the title or purpose of the picture.
   b. what each of the axes, bars, pie segments, and so on, denotes.
   c. the scale of each axis, including starting points.
3. A source should be given for the data.
4. There should be as little "chart junk"—that is, extraneous material—in the picture as possible.

9.2 Pictures of Categorical Data

Categorical data are easy to represent with pictures. The most frequent use of such data is to determine how the whole divides into categories, and pictures are useful in expressing that information. Let's look at three common types of pictures for categorical data and their uses.

Pie Charts

Pie charts are useful when only one categorical variable is measured. Pie charts show what percentage of the whole falls into each category. They are simple to understand, and they convey information about the relative size of groups more readily than a table. Figure 9.1 shows a pie chart that represents the percentage of Caucasian American children who have various hair colors.

Bar Graphs

Bar graphs also show percentages or frequencies in various categories, but they can be used to represent two or three categorical variables simultaneously. One categorical variable is used to label the horizontal axis. Within each of the categories along that axis, a bar is drawn to represent each category of the second variable. Frequencies or percentages are shown on the vertical axis. A third variable can be included

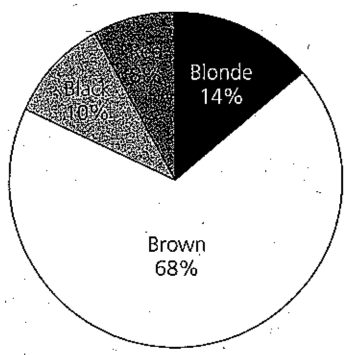


Figure 9.1
Pie chart of hair colors of Caucasian American children
Source: Krantz, 1992, p. 188.

if the graph has only two categories by using percentages on the vertical axis. One category is shown, and the other is implied by the fact that the total must be 100%.

For example, Figure 9.2 illustrates employment trends for men and women across decades. The year in which the information was collected is one categorical variable, represented by the horizontal axis. In each year, people were categorized according to two additional variables: whether they were in the labor force and whether they were male or female. Separate bars are drawn for males and females, and the percentage in the labor force determines the heights of the bars. It is implicit that the remainder were not in the labor force. Respondents were part of the Bureau of Labor Statistics' Current Population Survey, the large monthly survey used to determine unemployment rates.

The decision about which variable occupies which position should be made to better convey visually the purpose for the graph. The purpose of the graph in Figure 9.2 is to illustrate that the percentage of women in the labor force has increased

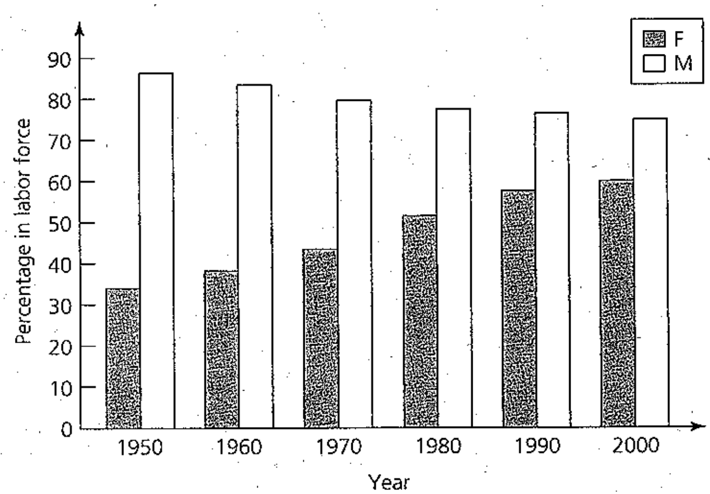


Figure 9.2
Percentage of males and females 16 and over in the labor force
Source: Based on data from U.S. Dept. of Labor, Bureau of Labor Statistics, *Current Population Survey.*

since 1950, whereas the percentage of men has decreased slightly, resulting in the two percentages coming closer together. The gap in 1950 was 53 percentage points, but by 2000 it was less than 15 percentage points, as is illustrated by the graph.

Bar graphs are not always as visually appealing as pie charts, but they are much more versatile. They can also be used to represent actual frequencies instead of percentages and to represent proportions that are not required to sum to 100%.

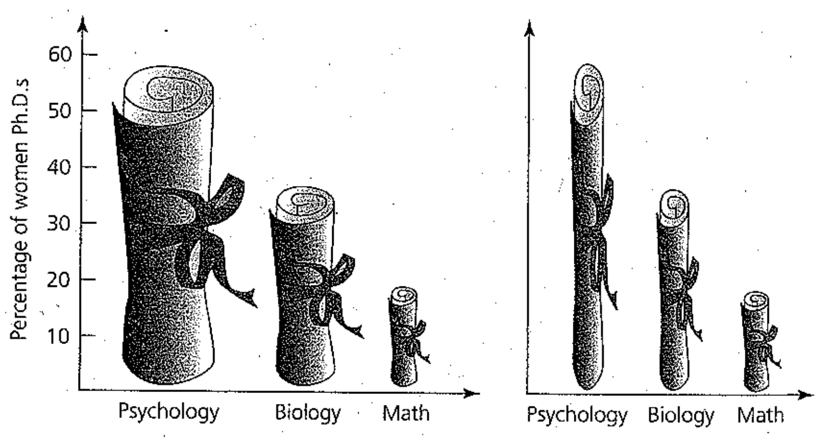
Pictograms

A **pictogram** is like a bar graph except that it uses pictures related to the topic of the graph. Figure 9.3 shows a pictogram illustrating the proportion of Ph.D.s earned by women in three fields—psychology (58%), biology (37%), and mathematics (18%)—as reported in *Science* (16 April, 1993, 260, p. 409). Notice that in place of bars, the graph uses pictures of diplomas.

It is easy to be misled by pictograms. The pictogram on the left shows the diplomas using realistic dimensions. However, it is misleading because the eye tends to focus on the *area* of the diploma rather than just its height. The heights of the three diplomas reach the correct proportions, with heights of 58%, 37%, and 18%, so the height of the one for psychology Ph.D.s is just over three times the height of the one for math Ph.D.s. However, in keeping the proportions realistic, the area of the diploma for psychology is about nine times the area of the one for math, leading the eye to inflate the difference.

The pictogram on the right is drawn by keeping the width of the diplomas the same for each field. The picture is visually more accurate, but it is less appealing because the diplomas are consequently quite distorted in appearance. When you see a pictogram, be careful to interpret the information correctly and not to let your eye mislead you.

Figure 9.3
Two pictograms showing percentages of Ph.D.s earned by women
Source: Alper, 16 April 1993, p. 409.



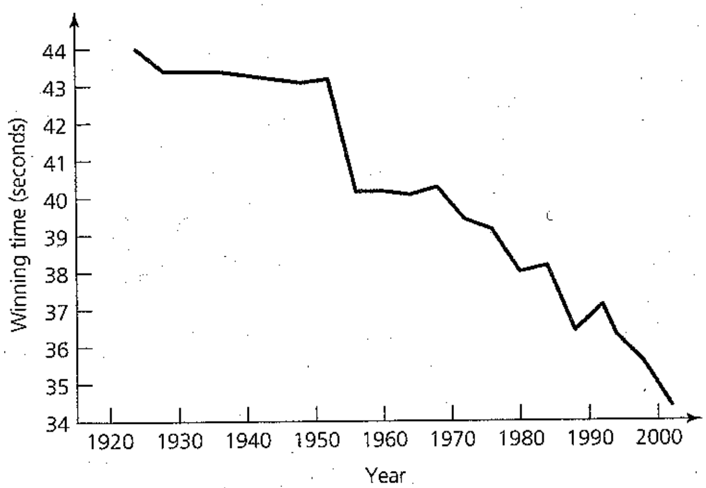


Figure 9.4
Line graph displaying winning time versus year for men's 500-meter Olympic speed skating
Source: http://sportsillustrated.cnn.com

9.3 Pictures of Measurement Variables

Measurement variables can be illustrated with graphs in numerous ways. We saw two ways to illustrate a single measurement variable in Chapter 7—namely, stem-plots and histograms. Graphs are most useful for displaying the relationship between two measurement variables or for displaying how a measurement variable changes over time. Two common types of displays for measurement variables are illustrated in Figures 9.4 and 9.5.

Line Graphs

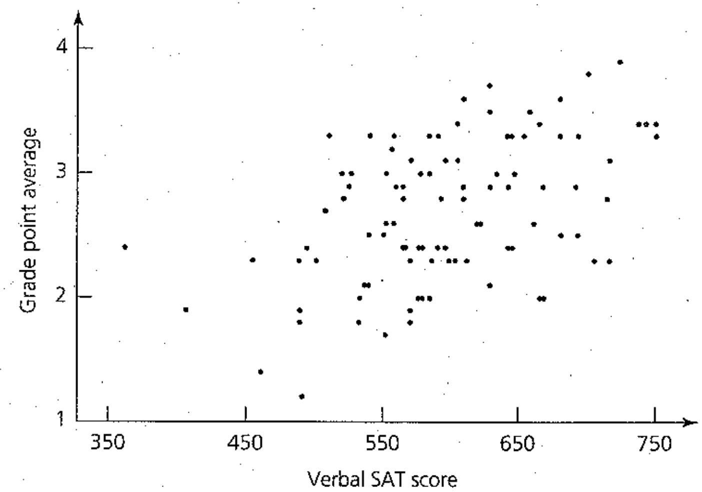
Figure 9.4 is an example of a **line graph** displayed over time. It shows the winning times for the men's 500-meter speed skating event in the Winter Olympics from 1924 to 2002. Notice the distinct downward trend, with only a few upturns over the years. There was a large drop between 1952 and 1956, followed by a period of relative stability. These patterns are much easier to detect with a picture than they would be by scanning a list of winning times.

Scatterplots

Figure 9.5 is an example of a **scatterplot.** Scatterplots are useful for displaying the relationship between two measurement variables. Each dot on the plot represents one individual, unless two or more individuals have the same data, in which case only one point is plotted at that location. The plot in Figure 9.5 shows the grade point averages (GPAs) and verbal scholastic achievement test (SAT) scores for a sample of 100 students at a university in the northeastern United States.

Although a scatterplot can be more difficult to read than a line graph, it displays more information. It shows outliers, as well as the degree of variability that exists for

Figure 9.5
Scatterplot of grade point average versus verbal SAT score
Source: Ryan, Joiner, and Ryan, 1985, pp. 309–312.



one variable at each location of the other variable. In Figure 9.5, we can see an increasing trend toward higher GPAs with higher SAT scores, but we can also still see substantial variability in GPAs at each level of verbal SAT scores. A scatterplot is definitely more useful than the raw data. Simply looking at a list of the 100 pairs of GPAs and SAT scores, we would find it difficult to detect the trend that is so obvious in the scatterplot.

9.4 Difficulties and Disasters in Plots, Graphs, and Pictures

A number of common mistakes appear in plots and graphs that may mislead readers. If you are aware of them and watch for them, you will substantially reduce your chances of misreading a statistical picture.

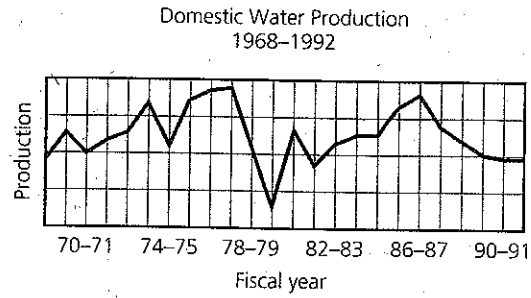
The most common problems in plots, graphs, and pictures are

1. No labeling on one or more axes
2. Not starting at zero as a way to exaggerate trends
3. Change(s) in labeling on one or more axes
4. Misleading units of measurement
5. Using poor information

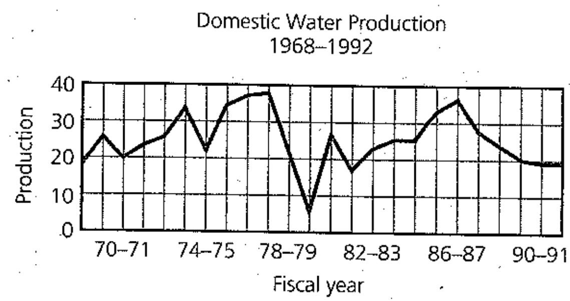
Figure 9.6

Example of a graph with no labeling (a) and possible interpretations (b and c)

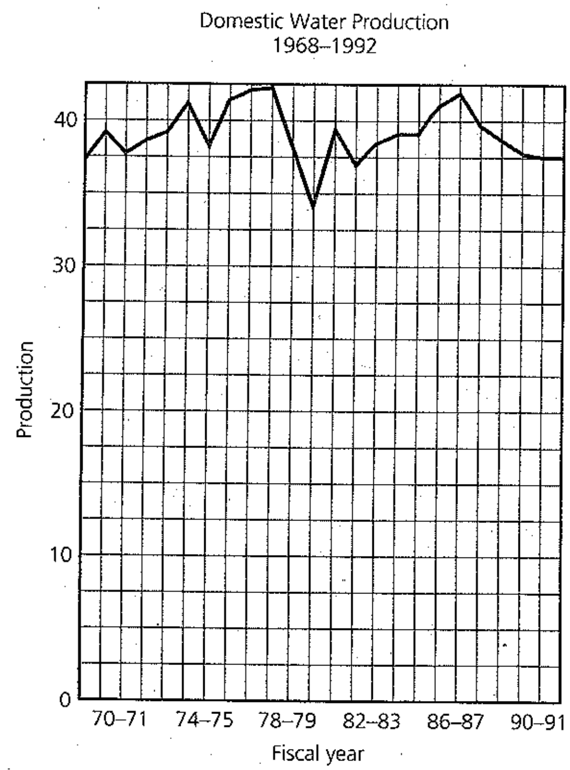
Source: Insert in the *California Aggie* (UC Davis), 30 May 1993.



(a) Actual graph



(b) Axis in "actual graph" starts at zero



(c) Axis in "actual graph" does not start at zero as a way to exaggerate trends

No Labeling on One or More Axes

You should always look at the axes in a picture to make sure they are labeled. Figure 9.6a gives an example of a plot for which the units were *not* labeled on the vertical axis. The plot appeared in a newspaper insert titled, "May 1993: Water awareness month." When there is no information about the units used on one of the axes, the plot cannot be interpreted. To see this, consider Figure 9.6b and c, displaying two different scenarios that could have produced the actual graph in Figure 9.6a. In Figure 9.6b, the vertical axis starts at zero for the existing plot. In Figure 9.6c, the vertical axis for the original plot starts at 30 and stops at 40, so what appears to be a large drop in 1979 in the other two graphs is only a minor fluctuation. We do not know which of these scenarios is closer to the truth, yet you can see that the two possibilities represent substantially different situations.

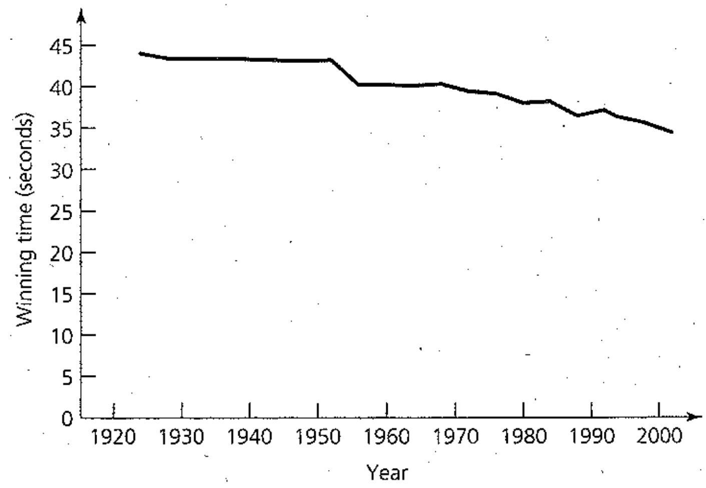


Figure 9.7
An example of the change in perception when axes start at zero

Not Starting at Zero

Often, even when the axes are labeled, the scale of one or both of the axes does not start at zero, and the reader may not notice that fact. A common ploy is to present an increasing or decreasing trend over time on a graph that does not start at zero. As we saw for the example in Figure 9.6, what appears to be a substantial change may actually represent quite a modest change. Always make it a habit to check the numbers on the axes to see where they start.

Figure 9.7 shows what the line graph of winning times for the Olympic speed skating data in Figure 9.4 would have looked like if the vertical axis had started at zero. Notice that the drop in winning times over the years does not look nearly as dramatic as it did in Figure 9.4. Be very careful about this form of potential deception if someone is presenting a graph to display growth in sales of a product, a drop in interest rates, and so on. Be sure to look at the labeling, especially on the vertical axis.

Despite this, be aware that for some graphs it makes sense to start the units on the axes at values different from zero. A good example is the scatterplot of GPAs versus SAT scores in Figure 9.5. It would make no sense to start the horizontal axis (SAT scores) at zero because the range of interest is from about 350 to 800. It is the responsibility of the reader to notice the units. Never assume a graph starts at zero without checking the labeling.

Changes in Labeling on One or More Axes

Figure 9.8 shows an example of a graph where a cursory look would lead one to think the vertical axis starts at zero. However, notice the white horizontal bar just above the bottom of the graph, in which the vertical bars are broken. That indicates a gap in the vertical axis. In fact, you can see that the bottom of the graph actually corresponds to about 4.0%. It would have been more informative if the graph had simply been labeled as such, without the break.

Figure 9.8
A bar graph with gap in labeling
Source: Davis (CA) *Enterprise,* 4 March 1994, p. A-7.

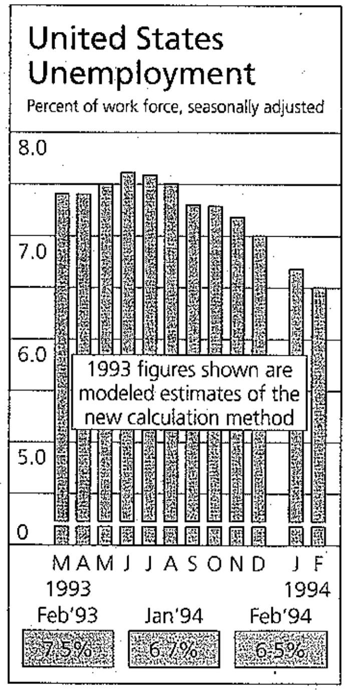


Figure 9.9 shows a much more egregious example of changes in labeling. Notice that the horizontal axis does not maintain consistent distances between years and that varying numbers of years are represented by each of the bars. The distance between the first and second bars on the left is 8 years, whereas the 5 bars farthest to the right each represent a single year. This is an extremely misleading graph.

Misleading Units of Measurement

The units shown on a graph can be different from those that the reader would consider important. For example, Figure 9.10 shows a graph with the heading, "Rising Postal Rates." It accurately represents how the cost of a first-class stamp rose from 1971 to 1991. However, notice that the fine print at the bottom reads, "In 1971 dollars, the price of a 32-cent stamp in February 1995 would be 8.4 cents." A more truthful picture would show the changing price of a first-class stamp adjusted for inflation. As the footnote implies, such a graph would show little or no rise in postal rates as a function of the worth of a dollar.

Using Poor Information

A picture can only be as accurate as the information that was used to design it. All of the cautions about interpreting the collection of information given in Part 1 of this

Figure 9.9
The distance between successive bars keeps changing.
Source: Washington Post graph reprinted in Wainer, 1984.

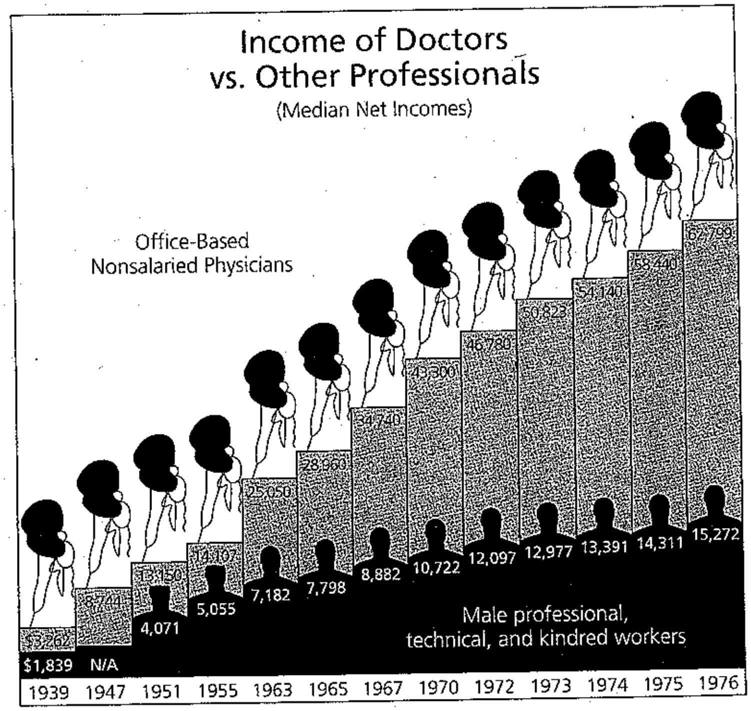
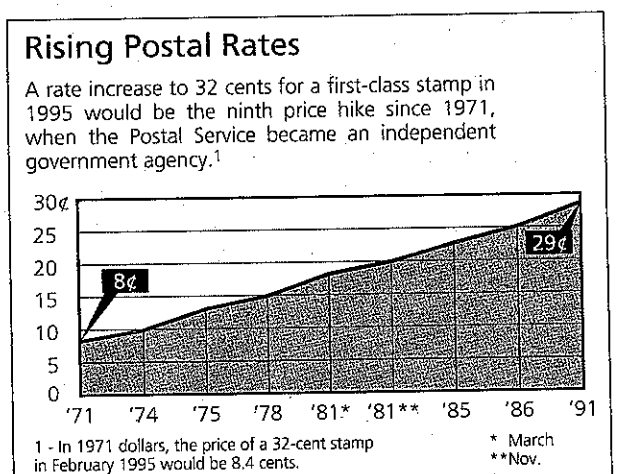


Figure 9.10
A graph using misleading units
Source: USA Today, 7 March 1994, p. 13A.



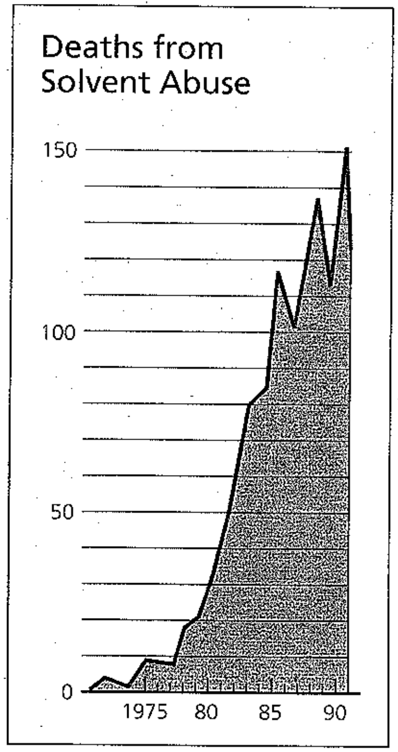


Figure 9.11
A graph based on poor information
Source: The Independent on Sunday (London), 13 March 1994.

book apply to graphs and plots as well. You should always be told the source of information presented in a picture, and an accompanying article should give you as much information as necessary to determine the worth of that information.

Figure 9.11 shows a graph that appeared in the London newspaper the *Independent on Sunday* on March 13, 1994. The accompanying article was titled, "Sniffers Quit Glue for More Lethal Solvents." The graph appears to show that very few deaths occurred in Britain from solvent abuse before the late 1970s. However, the accompanying article includes the following quote, made by a research fellow at the unit where the statistics are kept: "It's only since we have started collecting accurate data since 1982 that we have begun to discover the real scale of the problem" (p. 5). In other words, the article indicates that the information used to create the graph is not at all accurate until at least 1982. Therefore, the apparent sharp increase in deaths linked to solvent abuse around that time period is likely to have been simply a sharp increase in deaths reported and classified.

Don't forget that a statistical picture isn't worth much if the data can't be trusted. Once again, you should familiarize yourself to the extent possible with the Seven Critical Components listed in Chapter 2 (pp. 18–19).