human error, and O denotes other causes (based on data from the Federal Railroad Administration).

T T T E E H H H H H O O H H H E E T T T E T H O T

T T T T T T H T T H E E T T E E T T T H T T O O O

31. Fatal Plane Crashes Among fatal plane crashes that occurred during the past 60 years, 650 were due to pilot error, 78 were due to other human error, 156 were due to weather, 286 were due to mechanical problems, and 117 were due to sabotage. (These results are based on data from PlaneCrashInfo.com, and they do not include private aircraft, military aircraft, aircraft carrying 10 or fewer people, or helicopters.) What is the most serious threat to aviation safety, and can anything be done about it?

32. California Lottery The digits drawn in one month for the California Daily 4 lottery were recorded. The digits 0 through 9 had these frequencies: 20, 10, 12, 12, 8 11, 9, 10, 9, 19. Do the digits appear to be selected with a process that is functioning correctly? Why are these digits *categorical* data?

2-2 Beyond the Basics

33. Interpreting Effects of Outliers Refer to Data Set 22 in Appendix B for the axial loads of aluminum cans that are 0.0111 in. thick. The load of 504 lb is an *outlier* because it is very far away from all of the other values. Construct a frequency distribution that includes the value of 504 lb, and then construct another frequency distribution with the value of 504 lb excluded. In both cases, start the first class at 200 lb and use a class width of 20 lb. State a generalization about the effect of an outlier on a frequency distribution.

34. Number of Classes According to what is known as Sturges' guideline, the ideal number of classes for a frequency distribution can be approximated by $1 + (\log n)/(\log 2)$ where n is the number of data values. Use this guideline to complete the table in the margin.

Table for Exercise 34

Number of Data Values	Ideal Number of Classes
16–22	5
23–45	6
?	7
?	8
?	9
?	10
?	11
?	12

2-3 Histograms

Key Concept While a frequency distribution is a useful tool for summarizing data and investigating the distribution of data, an even better tool is a *histogram*, which consists of a graph that is easier to interpret than a table of numbers.

DEFINITION A **histogram** is a graph consisting of bars of equal width drawn adjacent to each other (unless there are gaps in the data). The horizontal scale represents classes of quantitative data values and the vertical scale represents frequencies. The heights of the bars correspond to the frequency values.

A histogram is basically a graph of a frequency distribution. For example, Figure 2-2 shows the Minitab-generated histogram corresponding to the frequency distribution given in Table 2-2 on page 45.

Class frequencies should be used for the vertical scale and that scale should be labeled as in Figure 2-2. The bar locations on the horizontal scale are usually labeled with one of the following: (1) class boundaries (as shown in Figure 2-2), (2) class midpoints, or (3) lower class limits. The first and second options are technically correct, while the third option introduces a small error. It is often easier for us mere mortals to use class midpoints for the horizontal scale. Histograms can usually be generated using technology.