the boxplot leaves out the highest 5% in each group. Based on the plot, compare the distributions of income for these two levels of education. Comment on both center and spread.

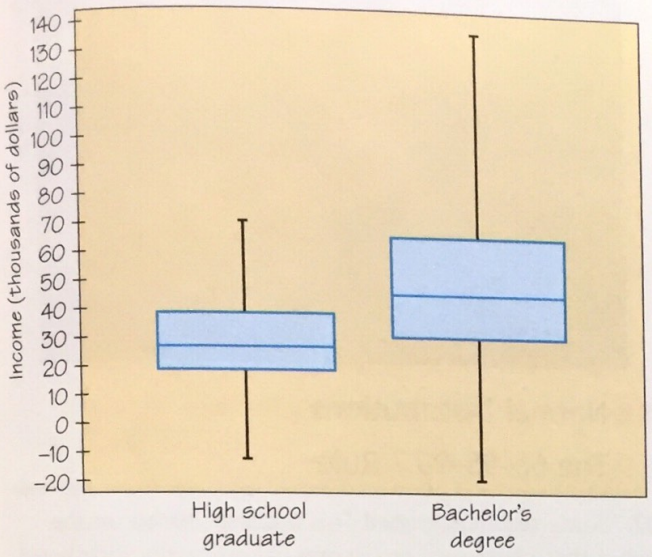

FIGURE 5.21 Boxplots comparing the incomes (in thousands of dollars) of people aged 25 to 64 years who worked full time for two levels of education. Because the highest incomes in any large group are very high indeed, the plot omits the top 5% of incomes in each group.

26. The data for Figure 5.21 include the incomes of 14,959 people whose highest level of education is a bachelor's degree.

(a) What is the position of the median in the ordered list of incomes (1 to 14,959)? From the boxplot, about what is the median income of people with a bachelor's degree?

(b) What is the position of the first and third quartiles in the ordered list of incomes for these people? About what are the numerical values of Q_1 and Q_3?

■ **27.** How much oil the wells in a given field will ultimately produce is key information in deciding whether to drill more wells. Here are the estimated total amounts of oil recovered from 64 wells in the Devonian Richmond Dolomite area of the Michigan basin, in thousands of barrels. [J. Marcus Jobe and Hutch Jobe, A statistical approach for additional infill development, *Energy Exploration and Exploitation*, 18 (2000): 89–103.]

2.0	18.5	34.6	47.6	69.5
2.5	20.1	34.6	49.4	69.8
3.0	21.3	35.1	50.4	79.5
7.1	21.7	36.6	51.9	81.1
10.1	24.9	37.0	53.2	82.2
10.3	26.9	37.7	54.2	92.2
12.0	28.3	37.9	56.4	97.7
12.1	29.1	38.6	57.4	103.1
12.9	30.5	42.7	58.8	118.2
14.7	31.4	43.4	61.4	156.5
14.8	32.5	44.5	63.1	196.0
17.6	32.9	44.9	64.9	204.9
18.0	33.7	46.4	65.6	

(a) Make a histogram and describe its main features.

(b) Find the mean and median of the amounts recovered. Explain how the relationship between the mean and the median reflects the shape of the distribution.

(c) Give the five-number summary and explain briefly how it reflects the shape of the distribution.

■ **28.** Look at the histogram of lengths of words in Shakespeare's plays, Figure 5.18. The heights of the bars tell us what percent of words have each length. (Analysis of writing tendencies can help determine authorship of a newly-discovered manuscript.) The median length is the middle, the length with half of all words shorter and half longer. What is the median length of words used by Shakespeare? Similarly, what are the quartiles? Give the five-number summary for Shakespeare's word lengths.

■ **29.** A common criterion for identifying an outlier in a set of data is if an observation falls more than $1.5 \times IQR$ above the third quartile or below the first quartile. (IQR stands for the interquartile range, which is the difference between the quartiles: $Q_3 - Q_1$.)

So which states are suspected outliers in the distribution of percent of Hispanics among adult residents, Table 5.1?

5.7 Describing Spread: The Standard Deviation

30. Do you think the standard deviation of the tuition and fees of the public colleges in Massachusetts is likely to be bigger or smaller than the standard deviation for the private colleges? Why?

31. Many standard statistical methods are intended for use with distributions that are symmetric and have no outliers. These methods start with the mean and standard deviation, $\bar{x}$ and s. An example of scientific data for which standard methods should work well are Cavendish's measurements of the density of the earth in Exercise 10.

(a) Summarize this data set by giving $\bar{x}$ and s.

(b) Find the median. Is the median quite close to the mean, as we expect it to be for symmetric distributions?

32. The level of various substances in the blood influences our health. Here are measurements of the level of phosphate in the blood of a patient, in milligrams of phosphate per deciliter of blood, made on six consecutive visits to a clinic.

5.6 5.2 4.6 4.9 5.7 6.4

(a) Find the mean.

(b) Find the standard deviation.

33. The mean $\bar{x}$ and standard deviation s measure center and spread but are not a complete description of a