With two or more data sets consisting of observations on the same variable (for example, fuel efficiencies for four types of car or weight gains for a control group and a treatment group), **comparative boxplots** (more than one boxplot drawn using the same scale) can tell us a lot about similarities and differences between the data sets.

EXAMPLE 4.13   NBA Salaries Revisited

The 2009–2010 salaries of NBA players published on the web site hoopshype.com were used to construct the comparative boxplot of the salary data for five teams shown in Figure 4.11.
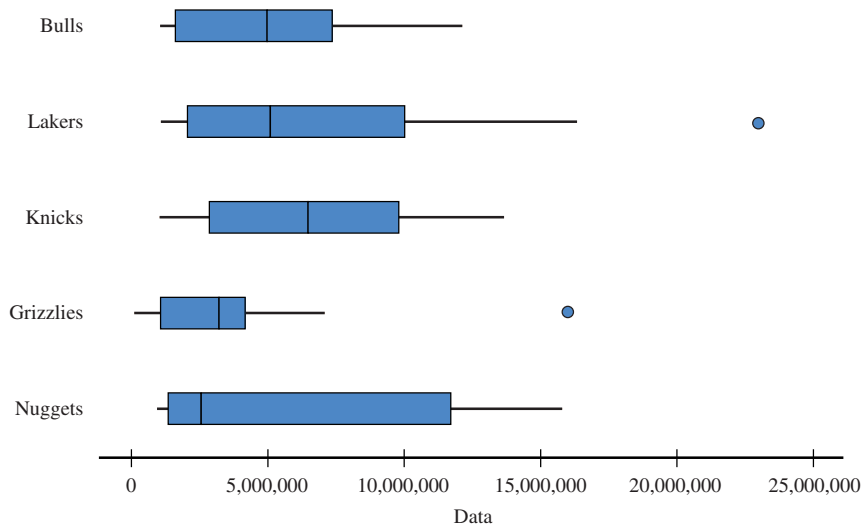

FIGURE 4.11

Comparative boxplot for salaries for five NBA teams.

The comparative boxplot reveals some interesting similarities and differences in the salary distributions of the five teams. The minimum salary is lower for the Grizzlies, but is about the same for the other four teams. The median salary was lowest for the Nuggets—in fact the median for the Nuggets is about the same as the lower quartile for the Knicks and the Lakers, indicating that half of the players on the Nuggets have salaries less than about $2.5 million, whereas only about 25% of the Knicks and the Lakers have salaries less than about $2.5 million. The Lakers had the player with by far the highest salary. The Grizzlies and the Lakers were the only teams that had any salary outliers. With the exception of one highly paid player, salaries for players on the Grizzlies team were noticeably lower than for the other four teams.

EXERCISES 4.32 - 4.37

4.32   Based on a large national sample of working adults, the U.S. Census Bureau reports the following information on travel time to work for those who do not work at home:

lower quartile = 7 minutes
median = 18 minutes
upper quartile = 31 minutes

Bold exercises answered in back   ● Data set available online   ✦ Video Solution available

Also given was the mean travel time, which was reported as 22.4 minutes.

a. Is the travel time distribution more likely to be approximately symmetric, positively skewed, or negatively skewed? Explain your reasoning based on the given summary quantities.

b. Suppose that the minimum travel time was 1 minute and that the maximum travel time in the sample was 205 minutes. Construct a skeletal boxplot for the travel time data.

c. Were there any mild or extreme outliers in the data set? How can you tell?

4.33 ● The report "Who Moves? Who Stays Put? Where's Home?" (*Pew Social and Demographic Trends,* December 17, 2008) gave the accompanying data for the 50 U.S. states on the percentage of the population that was born in the state and is still living there. The data values have been arranged in order from largest to smallest.

75.8 71.4 69.6 69.0 68.6 67.5 66.7 66.3 66.1 66.0 66.0
65.1 64.4 64.3 63.8 63.7 62.8 62.6 61.9 61.9 61.5 61.1
59.2 59.0 58.7 57.3 57.1 55.6 55.6 55.5 55.3 54.9 54.7
54.5 54.0 54.0 53.9 53.5 52.8 52.5 50.2 50.2 48.9 48.7
48.6 47.1 43.4 40.4 35.7 28.2

a. Find the values of the median, the lower quartile, and the upper quartile.

b. The two smallest values in the data set are 28.2 (Alaska) and 35.7 (Wyoming). Are these two states outliers?

c. Construct a boxplot for this data set and comment on the interesting features of the plot.

4.34 ● The National Climate Data Center gave the accompanying annual rainfall (in inches) for Medford, Oregon, from 1950 to 2008 (www.ncdc.noaa.gov/oa/climate/research/cag3/city.html):

28.84 20.15 18.88 25.72 16.42 20.18 28.96 20.72 23.58
10.62 20.85 19.86 23.34 19.08 29.23 18.32 21.27 18.93
15.47 20.68 23.43 19.55 20.82 19.04 18.77 19.63 12.39
22.39 15.95 20.46 16.05 22.08 19.44 30.38 18.79 10.89
17.25 14.95 13.86 15.30 13.71 14.68 15.16 16.77 12.33
21.93 31.57 18.13 28.87 16.69 18.81 15.15 18.16 19.99
19.00 23.97 21.99 17.25 14.07

a. Compute the quartiles and the interquartile range.

b. Are there outliers in this data set? If so, which observations are mild outliers? Which are extreme outliers?

c. Draw a boxplot for this data set that shows outliers.

4.35 ● The accompanying data on annual maximum wind speed (in meters per second) in Hong Kong for each year in a 45-year period were given in an article that appeared in the journal *Renewable Energy* (March 2007). Use the annual maximum wind speed data to construct a boxplot. Is the boxplot approximately symmetric?

30.3 39.0 33.9 38.6 44.6 31.4 26.7 51.9 31.9
27.2 52.9 45.8 63.3 36.0 64.0 31.4 42.2 41.1
37.0 34.4 35.5 62.2 30.3 40.0 36.0 39.4 34.4
28.3 39.1 55.0 35.0 28.8 25.7 62.7 32.4 31.9
37.5 31.5 32.0 35.5 37.5 41.0 37.5 48.6 28.1

4.36 ● Fiber content (in grams per serving) and sugar content (in grams per serving) for 18 high fiber cereals (www.consumerreports.com) are shown below.

Fiber Content

7 10 10 7 8 7 12 12 8
13 10 8 12 7 14 7 8 8

Sugar Content

11 6 14 13 0 18 9 10 19
6 10 17 10 10 0 9 5 11

a. Find the median, quartiles, and interquartile range for the fiber content data set.

b. Find the median, quartiles, and interquartile range for the sugar content data set.

c. Are there any outliers in the sugar content data set?

d. Explain why the minimum value for the fiber content data set and the lower quartile for the fiber content data set are equal.

e. Construct a comparative boxplot and use it to comment on the differences and similarities in the fiber and sugar distributions.

4.37 ● Shown here are the number of auto accidents per year for every 1000 people in each of 40 occupations (*Knight Ridder Tribune,* June 19, 2004):

Occupation	Accidents per 1000	Occupation	Accidents per 1000
Student	152	Social worker	98
Physician	109	Manual laborer	96
Lawyer	106	Analyst	95
Architect	105	Engineer	94
Real estate broker	102	Consultant	94
Enlisted military	99	Sales	93

(continued)

Bold exercises answered in back ● Data set available online ✦ Video Solution available