

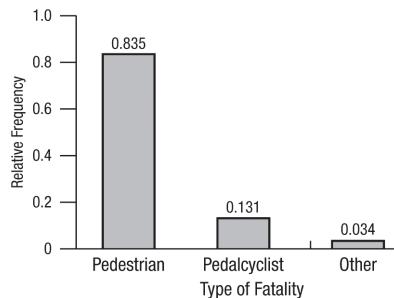
## Chapter 2 – Displaying and Describing Data

### Section 2.1

#### 1. Automobile fatalities.

Subcompact and Mini	0.2658
Compact	0.2084
Intermediate	0.3006
Full	0.2069
Unknown	0.0183

#### 2. Nonoccupant fatalities.



#### 3. Movie genres.

- a) A pie chart seems appropriate for the movie genre data. Each movie has only one genre, and the list of all movies constitute a “whole”.
- b) “Other” is the least common genre. It has the smallest region in the chart.

#### 4. Movie ratings.

- a) A pie chart seems appropriate for the movie ratings data. Each movie has only one rating, and the list of all movies constitute a “whole”.
- b) “R” is the most common rating. It has the largest region in the chart.

#### 5. Movie ratings.

- i) C              ii) A              iii) D              iv) B

#### 6. Marriage in decline.

- i) D              ii) A              iii) C              iv) B

### Section 2.2

#### 7. Traffic Fatalities 2013.

- a) The gaps in the histogram for *Year* indicate that we do not have data for those years. This data set contains two variables for each case, and a histogram of the years doesn’t give us much useful information.
- b) All of the bars in the *Year* histogram are the same height because each year only appears once in the data set.
- c) The distribution of passenger car fatalities has between 17,500 and 25,000 traffic fatalities per year in most years. There were also several years—possibly a second mode—with between 10,000 and 12,500 traffic fatalities.

## **8 Part I Exploring and Understanding Data**

### **8. Traffic fatalities 2013 again.**

- a) Two years are included in the highlighted bar.
- b) The years are 1975 and 1980.
- c) There have been fewer traffic deaths more recently than there were in 1975 and 1980.

### **9. How big is your bicep?**

The distribution of the bicep measurements of 250 men is unimodal and symmetric. Based on the height of the tallest points, about 85 of these 250 men have biceps close to 13 inches around. Most are between 12 and 15 inches around. But there are two as small as 10 inches and several that are 16 inches.

### **10. How big is your bicep in cm?**

Yes, the dotplots look different. The plot based on inches has fewer values on the  $x$ -axis, so it shows less detail. This plot gives a better picture of the distribution.

### **11. E-mails.**

The distribution of the number of emails received from each student by a professor in a large introductory statistics class during an entire term is skewed to the right, with the number of emails ranging from 1 to 21 emails. The distribution is centered at about 2 emails, with many students only sending 1 email. There is one outlier in the distribution, a student who sent 21 emails. The next highest number of emails sent was only 8.

### **12. Adoptions.**

- a) The distributions of the number of adoptions and population are skewed to the right. Most states have smaller populations and fewer adoptions, but some big states have substantially more of each.
- b) States with higher populations are likely to have more adoptions.
- c) Report the number of adoptions per 100,000 people.

## **Section 2.3**

### **13. Biceps revisited.**

The distribution of the bicep measurements of 250 men is mound-shaped and roughly symmetric.

### **14. E-mails II.**

The distribution of e-mails sent by students is skewed to the right.

### **15. Life expectancy.**

- a) The distribution of life expectancies at birth in 190 countries is skewed to the left.
- b) The distribution of life expectancies at birth in 190 countries has one mode, at about 74 to 76 years. The fluctuations from bar to bar don't seem to rise to the level of defining additional modes, although opinions can differ.

### **16. Shoe sizes.**

- a) The distribution is roughly symmetric and possibly bimodal.
- b) There seem to be two modes, one between sizes 38 and 40 and another between sizes 44 and 46. This could be due to having data for both men and women. The lower mode might be for women and the upper for men.

## **Section 2.4**

### **17. Life expectancy II.**

- a) The distribution of life expectancies at birth in 190 countries is skewed to the left, so the median is expected to be larger than the mean. The mean life expectancy is pulled down toward the tail of the distribution.

**17.** (continued)

- b)** Since the distribution of life expectancies at birth in 190 countries is skewed to the left, the median is the better choice for reporting the center of the distribution. The median is resistant to the skewed shape of the distribution, so it is a better choice for most summaries.

**18. Adoptions II.**

- a)** The distribution of number of adoptions is skewed to the right, so the mean is expected to be larger than the median. The mean life expectancy is pulled upwards toward the tail of the distribution.
- b)** Since the distribution of number of adoptions is skewed to the right, the median is the better choice for reporting the center of the distribution. The median is resistant to the skewed shape of the distribution, so it is a better choice for most summaries.

**19. How big is your bicep II?**

Because the distribution of bicep circumferences is unimodal and symmetric, the mean and the median should be very similar. The usual choice is to report the mean or to report both.

**20. Shoe sizes II.**

Because the distribution of shoe sizes has two modes, the mean and median are not helpful in reporting the story that the data tell. It is better to report the locations of the two modes.

**Section 2.5**

**21. Life expectancy III.**

- a)** We should report the IQR.
- b)** Since the distribution of life expectancies at birth in 190 countries is skewed to the left, the better measure of spread is the IQR. The skewness of the distribution inflates the standard deviation.

**22. Adoptions III.**

- a)** We should report the IQR.
- b)** Since the distribution of number of adoptions is skewed to the left, the better measure of spread is the IQR. The skewness of the distribution inflates the standard deviation.

**23. How big is your bicep III?**

Because the distribution of bicep circumferences is unimodal and roughly symmetric, we should report the standard deviation. The standard deviation is generally more useful whenever it is appropriate. However, it would not be strictly wrong to use the IQR. We just prefer the standard deviation.

**24. Shoe sizes III.**

The data combine shoe sizes for men and for women. It isn't appropriate to summarize them as if they were a coherent collection of values.

**Chapter Exercises**

**25. Graphs in the news.**

Answers will vary.

**29. Histogram.**

Answers will vary.

**26. Graphs in the news II.**

Answers will vary.

**30. Not a histogram.**

Answers will vary.

**27. Tables in the news.**

Answers will vary.

**31. Centers in the news.**

Answers will vary.

**28. Tables in the news II.**

Answers will vary.

**32. Spreads in the news.**

Answers will vary.

## **10 Part I Exploring and Understanding Data**

### **33. Thinking about shape.**

- a) The distribution of the number of speeding tickets each student in the senior class of a college has ever had is likely to be unimodal and skewed to the right. Most students will have very few speeding tickets (maybe 0 or 1), but a small percentage of students will likely have comparatively many (3 or more?) tickets.
- b) The distribution of player's scores at the U.S. Open Golf Tournament would most likely be unimodal and slightly skewed to the right. The best golf players in the game will likely have around the same average score, but some golfers might be off their game and score 15 strokes above the mean. (Remember that high scores are undesirable in the game of golf!)
- c) The weights of female babies in a particular hospital over the course of a year will likely have a distribution that is unimodal and symmetric. Most newborns have about the same weight, with some babies weighing more and less than this average. There may be slight skew to the left, since there seems to be a greater likelihood of premature birth (and low birth weight) than post-term birth (and high birth weight).
- d) The distribution of the length of the average hair on the heads of students in a large class would likely be bimodal and skewed to the right. The average hair length of the males would be at one mode, and the average hair length of the females would be at the other mode, since women typically have longer hair than men. The distribution would be skewed to the right, since it is not possible to have hair length less than zero, but it is possible to have a variety of lengths of longer hair.

### **34. More shapes.**

- a) The distribution of the ages of people at a Little league game would likely be bimodal because you have both players and parents. The distribution may also be skewed to the right, since parents' ages can be higher than the mean more easily than lower.
- b) The distribution of the number of siblings of people in your class would likely be unimodal and skewed to the right. There are probably many students with 0 or 1 sibling and some with 2 or more.
- c) The distribution of pulse rates of college-age males would likely be unimodal and symmetric. It is unusual to have either very high or low pulse rates.
- d) The distribution of the number of times each face of a die shows in 100 tosses would likely be uniform. Each face of the die has the same chance of coming up, so the number of times should be about the same for each face.

### **35. Movie genres again.**

- a) Thriller/Suspense has a higher bar than Adventure, so it is the more common genre.
- b) It is easy to tell from either chart; sometimes differences are easier to see on the bar chart because slices of the pie chart look too similar in size.

### **36. Movie ratings again.**

- a) NC-17 has the lowest bar, so it is the least common rating.
- b) It is easy to tell from either chart; sometimes differences are easier to see on the bar chart because slices of the pie chart look too similar in size.

### **37. Magnet Schools.**

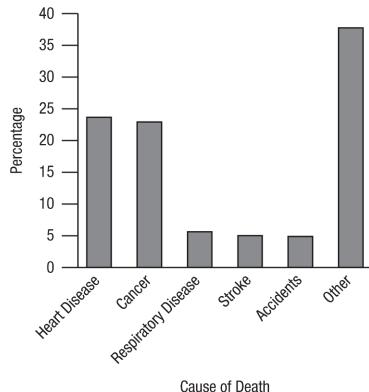
There were 1755 qualified applicants for the Houston Independent School District's magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away for lack of space.

### **38. Magnet schools again.**

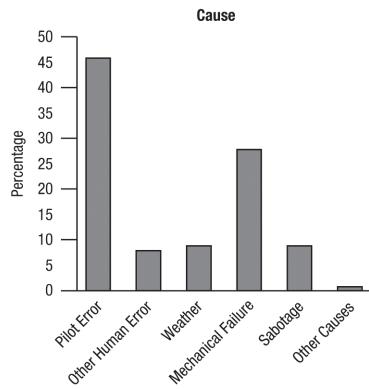
Of the 1755 students who applied for admission to the magnet schools program, 29.5% were black or Hispanic, 16.6% were Asian, and 53.9% were white.

**39. Causes of death 2014.**

- a) Yes, it is reasonable to assume that heart or lung diseases caused approximately 29% of U.S. deaths in 2014, since there is no possibility for overlap. Each person could only have one cause of death.
- b) Since the percentages listed add up to 61.9%, other causes must account for 38.1% of US deaths.
- c) A bar chart is a good choice (with the inclusion of the “Other” category). Since causes of US deaths represent parts of a whole, a pie chart would also be a good display.

**40. Plane crashes.**

- a) Yes, it is reasonable to assume that weather or mechanical failures caused approximately 37% of recent plane crashes, since there is no possibility for overlap. Each crash was assigned to only one cause category.
- b) The data are given to the nearest whole percent. If several of the categories were rounded up, an extra percentage point could be gained.
- c) A histogram is shown below. A pie chart would also be an appropriate display.

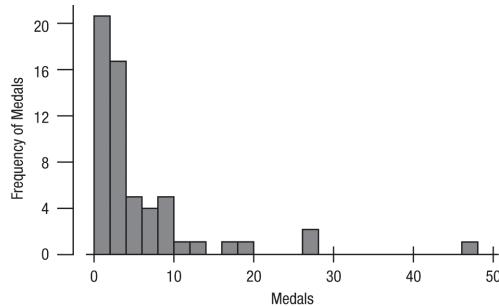
**41. Movie genres once more.**

- a) There are too many categories to construct an appropriate display. In a bar chart, there are too many bars. In a pie chart, there are too many slices. In each case, we run into difficulty trying to display genres that only represented a few movies.
- b) The creators of the bar chart included a category called “Other” for many of the genres that only occurred a few times.

## 12 Part I Exploring and Understanding Data

### 42. Summer Olympics 2016.

- a) There are too many categories to make a meaningful bar chart or pie chart by country.
- b) It would be better to make a histogram of total medals, as shown below.



### 43. Global warming.

Perhaps the most obvious error is that the percentages in the pie chart add up to 141%, when they should, of course, add up to 100%. This means that survey respondents were allowed to choose more than one response, so a pie chart is not an appropriate display. Furthermore, the three-dimensional perspective view distorts the regions in the graph, violating the area principle. The regions corresponding to “Could reduce global warming but unsure if we will” and “Could reduce global warming but people aren’t willing to so we won’t” look roughly the same size, but at 46% and 30% of respondents, respectively, they should have very different sizes. Always use simple, two-dimensional graphs. Additionally, the graph does not include a title.

### 44. Modalities.

- a) The bars have false depth, which can be misleading. This is a bar chart, so the bars should have space between them. From a simple design standpoint, running the labels on the bars one way and the vertical axis labels the other way is awkward.
- b) The percentages sum to 100%. This is unlikely if the respondents were asked to name three methods each. For example, it would be possible for 80% of respondents to use ice at some time and another 75% to use electric stimulation. This is a case where summing to 100% seems wrong rather than correct.

### 45. Cereals.

- a) The distribution of the carbohydrate content of breakfast cereals is bimodal, with a cluster of cereals with carbohydrate content around 13 grams of carbs and another cluster of cereals around 22 grams of carbs. The lower cluster shows a bit of skew to the left. Most cereals in the lower cluster have between 10 and 20 grams of carbs. The upper cluster is symmetric, with cereals in the cluster having between 20 and 24 grams of carbs.
- b) The cereals with the highest carbohydrate content are Corn Chex, Corn Flakes, Cream of Wheat (Quick), Crispix, Just Right Fruit & Nut, Kix, Nutri-Grain Almond-Raisin, Product 19, Rice Chex, Rice Krispies, Shredded Wheat ‘n’ Bran, Shredded Wheat Spoon Size, Total Corn Flakes, and Triples.

### 46. Run times.

The distribution of run times is skewed to the right and centered at around 30 to 31 minutes, with most observations between 29 and 32 minutes. The skewness is because it is easier to run much slower than usual but harder to run much faster than usual. Also, over a 10-year period, the runner may have slowed down due to injury or aging.

### 47. Heart attack stays.

- a) The distribution of length of stays is skewed to the right, so the mean is larger than the median.
- b) The distribution of the length of hospital stays of female heart attack patients is bimodal and skewed to the right, with stays ranging from 1 day to 36 days. The distribution is centered around 8 days, with the majority of the hospital stays lasting between 1 and 15 days. There are a relatively few hospital stays longer than 27 days. Many patients have a stay of only one day, possibly because the patient died.

47. (continued)

- c) The median and IQR would be used to summarize the distribution of hospital stays, since the distribution is strongly skewed.

48. Bird species 2013.

a)

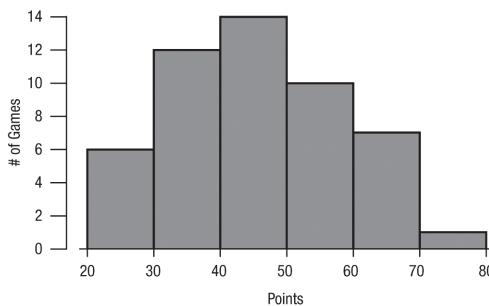
8	2368
9	78
10	1156
11	8
12	468
13	136
14	
15	0
16	6
17	
18	4

8|2 = 82

- b) The distribution of the number of bird species is mostly symmetric, with a few values trailing off to the high end. It could be described as skewed right.

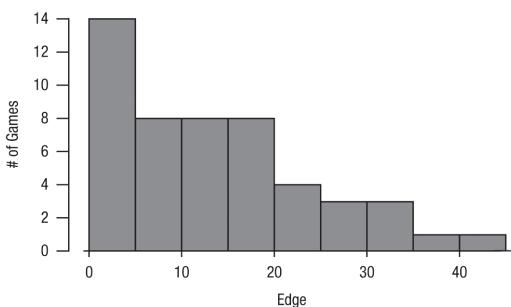
49. Super Bowl points 2016.

- a) The median number of points scored in the first 50 Super Bowl games is 46 points.
- b) The first quartile (Q1) of the number of points scored in the first 50 Super Bowl games is 37 points. The third quartile (Q3) is 55 points.
- c) In the first 50 Super Bowl games, the lowest number of points scored was 21, and the highest number of points scored was 75. The median number of points scored was 46, and the middle 50% of Super Bowls has between 37 and 55 points scored, making the IQR 18 points.



50. Super Bowl edge 2016.

- a) The median margin is 12.5 points.
- b) The first quartile is 4 and the third quartile is 19.
- c) The distribution of margins is strongly skewed to the right, with a median of 12.5 points. The minimum is 1 and the maximum is 45. The IQR is 15 points.



## 14 Part I Exploring and Understanding Data

### 51. Test scores, large class.

- a) The distribution of Calculus test scores is bimodal with one mode at about 62 and one at about 78. The higher mode might be math majors, and the lower mode might be non-math majors.
- b) Because the distribution of Calculus test scores is bimodal, neither the mean nor the median tells much about a typical score. We should attempt to learn if another variable (such as whether or not the student is a math major) can account for the bimodal character of the distribution.

### 52. Test scores, small class.

The distribution of Calculus test scores appears to be bimodal, with one mode at about 62 and one at about 78. The higher mode might be math majors, and the lower mode might be non-math majors. Because there are so few values in the middle bins, it is not as clear that this distribution is actually bimodal.

### 53. Mistake.

- a) As long as the boss's true salary of \$200,000 is still above the median, the median will be correct. The mean will be too large, since the total of all the salaries will decrease by  $\$2,000,000 - \$200,000 = \$1,800,000$ , once the mistake is corrected.
- b) The range will likely be too large. The boss's salary is probably the maximum, and a lower maximum would lead to a smaller range. The IQR will likely be unaffected, since the new maximum has no effect on the quartiles. The standard deviation will be too large, because the \$2,000,000 salary will have a large squared deviation from the mean.

### 54. Sick days.

- a) There may be high outliers, or the distribution of sick days may be skewed to the right. The company is using the mean, which is higher because of these features of the distribution. The union negotiators are quoting the median.
- b) These choices make sense: The company cares about the number of days lost to sickness, so the mean (or total) is what matters. The union looks at individual workers and sees that the typical (median) worker doesn't miss many days for sickness.

### 55. Floods 2015.

- a) The mean annual number of deaths from floods is 81.95.
- b) In order to find the median and the quartiles, the list must be ordered.  
29 38 38 43 48 49 56 68 76 80 82 82 82 86 87 103 113 118 131 136 176  
The median annual number of deaths from floods is 82, Q1 is 49 deaths, and Q3 is 103 deaths.  
(Some statisticians consider the median to be separate from both the lower and upper halves of the ordered list when the list contains an odd number of elements. This changes the position of the quartiles slightly. If median is excluded, Q1 = 48.5 and Q3 = 108. In practice, it rarely matters, since these measures of position are best for large data sets.)
- c) The range of the distribution of deaths is  $\text{Max} - \text{Min} = 176 - 29 = 147$  deaths. The IQR =  $Q3 - Q1 = 103 - 49 = 54$  deaths. (Or, the IQR =  $108 - 48.5 = 59.5$  deaths, if the median is excluded from both halves of the ordered list.)

### 56. Tornadoes 2015.

- a) The mean annual number of deaths from tornadoes is 81.95.
- b) In order to find the median and the quartiles, the list must be ordered.  
21 25 30 35 36 38 40 41 45 47 54 55 55 67 67 70 81 94 126 130 553  
The median annual number of deaths from tornadoes is 54, Q1 is 38 deaths, and Q3 is 70 deaths.  
(Some statisticians consider the median to be separate from both the lower and upper halves of the ordered list when the list contains an odd number of elements. This changes the position of the quartiles slightly. If median is excluded, Q1 is 37 and Q3 is 75.5. In practice, it rarely matters, since these measures of position are best for large data sets.)

**56.** (continued)

- c) The range of the distribution of deaths is  $\text{Max} - \text{Min} = 553 - 21 = 532$  deaths. The IQR =  $Q_3 - Q_1 = 70 - 38 = 32$  deaths. (Or, the IQR =  $75.5 - 37 = 38.5$  deaths, if the median is excluded from both halves of the ordered list.)

**57. Floods 2105 II.**

The distribution of deaths from floods is slightly skewed to the right and bimodal. There is one mode at about 40 deaths and one at about 80 deaths. There is one extreme value at 180 deaths.

**58. Tornadoes 2015 II.**

The distribution of deaths from tornadoes is slightly skewed to the right, with one extreme outlier at 553. The median is 54 deaths, and the IQR is 32 deaths.

**59. Pizza prices.**

The mean and standard deviation would be used to summarize the distribution of pizza prices, since the distribution is unimodal and symmetric.

**60. Neck size.**

The mean and standard deviation would be used to summarize the distribution of neck sizes, since the distribution is unimodal and symmetric.

**61. Pizza prices again.**

- a) The mean pizza price is closest to \$2.60. That's the balancing point of the histogram.
- b) The standard deviation in pizza prices is closest to \$0.15, since that is the typical distance to the mean. There are no pizza prices as far as \$0.50 or \$1.00.

**62. Neck sizes again.**

- a) The mean neck size is closest to 15 inches. That's the balancing point of the histogram.
- b) The standard deviation in neck sizes is closest to 1 inch, because a typical value lies about 1 inch from the mean. There are few points as far as 3 inches from the mean and none as far as 5 inches, so those are too large to be the standard deviation.

**63. Movie lengths 2010.**

- a) A typical movie would be around 105 minutes long. This is near the center of the unimodal and slightly skewed histogram, with the outlier set aside.
- b) You would be surprised to find that your movie ran for 150 minutes. Only 3 movies (2%) ran that long.
- c) The mean run time would probably be higher, since the distribution of run times is skewed to the right, and also has a high outlier. The mean is pulled towards this tail, while the median is more resistant. However, it is difficult to predict what the effect of the low outlier might be from just looking at the histogram.

**64. Golf drives 2015.**

- a) A typical drive length would be around 290 yards. This is near the center of the unimodal and reasonably symmetric histogram.
- b) About 25% of professional male golfers drive, on average, 280 yards or less.
- c) The mean of the distribution is 288.69 yards. Any estimate around 290 is a good guess.
- d) The median should be close to the mean because this distribution is symmetric. (In fact, it is 288.7, so they agree to within inches.)

## 16 Part I Exploring and Understanding Data

### 65. Movie lengths II 2010.

- a) i) The distribution of movie running times is fairly consistent, with the middle 50% of running times between 98 and 116 minutes. The interquartile range is 18 minutes.
- ii) The standard deviation of the distribution of movie running times is 16.6 minutes, which indicates that movies typically varied from the mean running time by 16.6 minutes.
- b) Since the distribution of movie running times is skewed to the right and contains an outlier, the standard deviation is a poor choice of numerical summary for the spread since it may have an inflated value. The interquartile range is better, since it is resistant to outliers.

### 66. Golf drives II 2015.

- a) i) The distribution of drive lengths is fairly consistent, with the middle 50% of golfers had driving averages between 282 and 294.5 yards.
- ii) The standard deviation of the distribution of drive lengths is 9.8 yards, which indicates that drive lengths typically varied from the mean drive length by 9.8 yards.
- b) The distribution of drive lengths is symmetric, so either is appropriate.

### 67. Movie earnings 2015.

The industry publication is using the median, while the watchdog group is using the mean. It is likely that the mean is pulled higher by a few very expensive movies.

### 68. Cold weather.

- a) The mean low temperature would probably be lower, since the distribution of low temperature is skewed to the left. The mean is pulled towards this tail, while the median is more resistant.
- b) The standard deviation will be higher since the data may have an outlier. The IQR won't change, since it is resistant to outliers.

### 69. Gasoline 2014.

a)

Stem	Leaf
31	1
31	5
32	1233
32	6678
33	
33	9
34	23
34	556

32|1 = \$3.21/gallon

- b) The distribution of gas prices is bimodal, with two clusters, one centered around \$3.25 per gallon, and another centered around \$3.45 per gallon. The lowest and highest prices were \$3.11 and \$3.46 per gallon.
- c) There is a gap in the distribution of gasoline prices. There were no stations that charged between \$3.28 and \$3.39.

**70. The great one.**

a)

Stem	Leaf
8	00000122
7	8899
7	0344
6	
6	4
5	
5	
4	58
4	

$$4|5 = 45 \text{ games}$$

- b) The distribution of games played by Gretzky is unimodal and skewed to the left.
- c) Typically, Gretzky played about 80 games per season. The numbers of games played are tightly clustered in the upper 70s and low 80s.
- d) Two seasons are low outliers, when Gretzky played fewer than 50 games. He might have been injured.

**71. States.**

- a) There are 50 entries in the stemplot, so the median must be between the 25th and 26th population values. Counting in the ordered stemplot gives median = 4.5 million people. The middle of the lower 50% of the list (25 state populations) is the 13<sup>th</sup> population, or 2 million people. The middle of the upper half of the list (25 state populations) is the 13th population from the top, or 7 million people. The IQR = Q3 – Q1 = 7 – 2 = 5 million people.
- b) The distribution of population for the 50 U.S. States is unimodal and skewed heavily to the right. The median population is 4.5 million people, with 50% of states having populations between 2 and 7 million people. There are two outliers, a state with 37 million people, and a state with 25 million people. The next highest population is only 19 million.

**72. Wayne Gretzky.**

- a) Since the distribution of number of games played is highly skewed to the left, the median is the better choice for reporting the center of the distribution. The median is resistant to the skewed shape of the distribution, so it is a better choice for most summaries.
- b) The median is 79 games played.
- c) The mean should be lower. There are two seasons when Gretzky played an unusually low number of games. Those will pull the mean down.

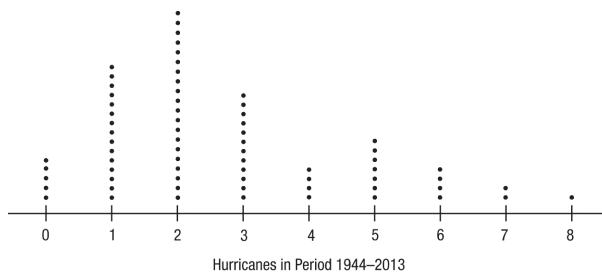
**73. A-Rod 2016.**

The distribution of the number of homeruns hit by Alex Rodriguez during the 1994 – 2016 seasons is reasonably symmetric, with the exception of a second mode around 10 homeruns. A typical number of homeruns per season was in the high 30s to low 40s. With the exception of 5 seasons in which A-Rod hit 0, 0, 5, 7, and 9 homeruns, his total number of homeruns per season was between 16 and the maximum of 57.

## 18 Part I Exploring and Understanding Data

### 74. Major hurricanes 2013.

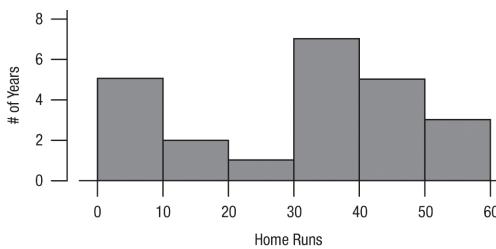
a)



- b) The distribution of the yearly number of hurricanes appears to be unimodal and slightly skewed to the right, with no outliers. The mode is near 2. There is possibly a second mode near 5.

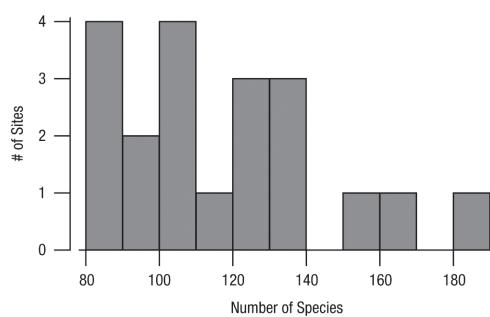
### 75. A-Rod again 2016.

- a) This is not a histogram. The horizontal axis should contain the number of home runs per year, split into bins of a convenient width. The vertical axis should show the frequency; that is, the number of years in which A-Rod hit a number of home runs within the interval of each bin. The display shown is a bar chart/time plot hybrid that simply displays the data table visually. It is of no use in describing the shape, center, spread, or unusual features of the distribution of home runs hit per year by A-Rod.
- b) The histogram is shown below.



### 76. Return of the birds 2013.

- a) This is not a histogram. The horizontal axis should split the number of counts from each site into bins. The vertical axis should show the number of sites in each bin.
- b)



### 77. Acid rain.

- a) The distribution of the pH readings of water samples in Allegheny County, Pennsylvania is skewed to the right and possibly bimodal. A roughly uniform cluster is centered around a pH of 4.4. This cluster ranges from pH of 4.1 to 4.9. Another smaller, tightly packed cluster is centered around a pH of 5.6. Two readings in the middle seem to belong to neither cluster.

77. (continued)

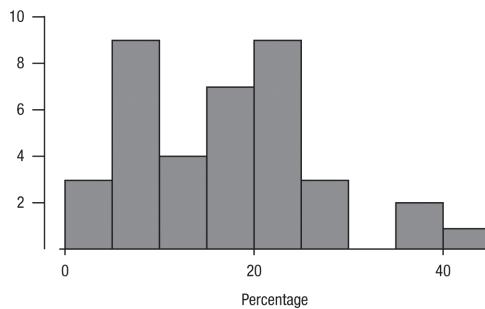
Stem	Leaf
57	8
56	27
55	1
54	
53	
52	9
51	
50	8
49	
48	2
47	3
46	034
45	267
44	015
43	0199
42	669
41	22

$$41|2 = 4.12 \text{ pH}$$

- b) The cluster of high outliers contains many dates that were holidays in 1973. Traffic patterns would probably be different then, which might account for the difference.

#### 78. Marijuana 2015.

The distribution of the percentages of 16-year-olds who had used marijuana is unimodal (or possibly bimodal) with center around 18%, but with France, Monaco, and the Czech Republic reporting values near 40%.



#### 79. Final grades.

The width of the bars is much too wide to be of much use. The distribution of grades is skewed to the left, but not much more information can be gathered.

#### 80. Final grades revisited.

- a) The histogram bars are too narrow to show the shape of the distribution.  
 b) The distribution of final grades is skewed to the left, with a mode near 170, and several outliers below 100. The distribution is fairly tightly clustered except for outliers.

#### 81. Zip codes.

Even though zip codes are numbers, they are not quantitative in nature. Zip codes are categories. A histogram is not an appropriate display for categorical data. The histogram the Holes R Us staff member displayed doesn't take into account that some 5-digit numbers do not correspond to zip codes or that zip codes falling into the same classes may not even represent similar cities or towns. The employee could design a better display by constructing a bar chart that groups together zip codes representing areas with similar demographics and geographic locations.

#### 82. ZIP codes revisited.

They can't tell very much because ZIP codes are categorical. However, there is some information in the first digit of ZIP codes. It indicates a general east (0–1) to west (8–9) direction. So, the distribution shows that a large portion of the sales occur in the West with another in the 32000 area. A bar chart of the first digits might be a better display of this.

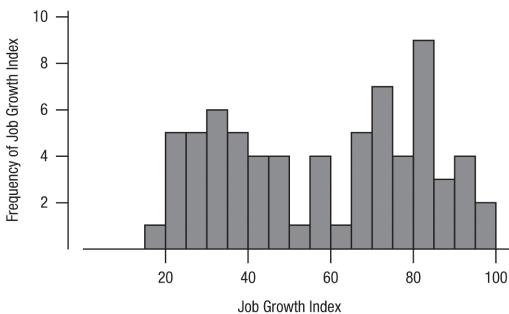
## 20 Part I Exploring and Understanding Data

### 83. Math scores 2013.

- a) Median: 285  
Mean: 284.36
- IQR: 9  
Standard deviation: 6.84
- b) Since the distribution of Math scores is skewed to the left, it is probably better to report the median and IQR.
- c) The distribution of average math achievement scores for eighth graders in the United States is skewed to the left, and roughly unimodal. The distribution is centered at 285. Scores range from 269 to 301, with the middle 50% of the scores falling between 280 and 289.

### 84. Boomtowns 2015.

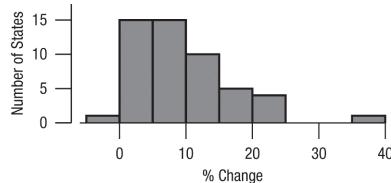
a)



- b) The mean growth index is 58.26% and the median growth index is 61.7%. With the bimodal distribution, it is difficult to predict the effect on the mean.
- c) It might be more informative to report one mode at about 33% and the other at about 80%.
- d) The IQR is 45.1 and the standard deviation is 23.85.
- e) Neither measure is ideal. It might be more informative to discuss a measure of spread for each mode.
- f) The mean and median would be 49.23 percentage points lower. The standard deviation and IQR would not change.
- g) The median and IQR won't change very much. The middle value and the two quartiles will shift at most one data value. The mean and standard deviation will decrease.
- h) Answers will vary but should include a discussion of the distribution and what it tells us about growth indices.

### 85. Population growth 2010.

The histogram shows that the distribution of *Percent Change* is unimodal and skewed to the right. The states vary from a minimum of  $-0.6\%$  (Michigan) to  $35.1\%$  (Nevada) growth in the decade. The median was  $7.8\%$  and half of the states had growth between  $4.3\%$  and  $14.1\%$ . Not surprisingly, the top three states in terms of growth were all in the West: Nevada ( $35.1\%$ ), Arizona ( $24.6\%$ ), and Utah ( $23.8\%$ ).



### 86. Student survey.

Answers will vary. The best answers will be about students and their responses, not about numbers and displays. Write what you learn about students, not about what your display looks like. Do not give the value of a summary statistic without a discussion of what it might mean.