Chapter 2 Solutions

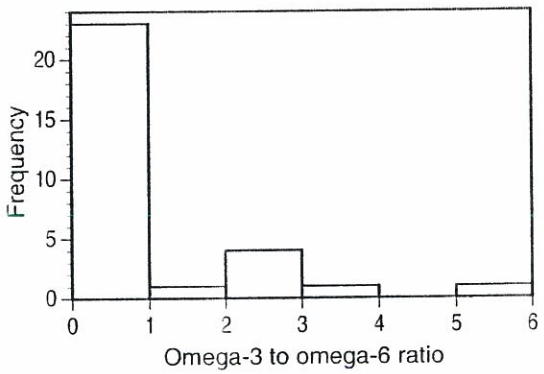
2.1. The mean is $\bar{x} = 30{,}841$ pounds. Only 6 of the 20 pieces of wood had breaking strengths below the mean. The distribution is skewed to the left, which makes the mean smaller than the "middle" of the set of numbers (the median).

2.2. With all countries included, the mean is $\bar{x} \doteq \$2007.18$ per person. Without the United States, the mean drops about \$100, to $\bar{x}^* \doteq \$1907.08$ per person.

2.3. The mean is 31.25 minutes, while the median is 22.5 minutes. This is what we expect for a right-skewed distribution like this one.

2.4. The median is \$218,900, and the mean is \$265,800. The distribution of housing prices will be right-skewed, so the mean will be higher.

2.5. The mean ratio is $\bar{x} \doteq 0.7607$, while the median is $M = 0.075$. The histogram (copied from the solution to Exercise 1.34) shows a sharp right skew, which accounts for this difference.



2.6. **(a)** and **(b)** The five-number summaries (in units of pounds) are

	Min	Q_1	M	Q_3	Max
Defensive line	255	296.5	300	300	310
Offensive line	305	305	313.5	324	366

Defensive line		Offensive line
25	5	25
26		26
27		27
28		28
29	588	29
30	0000	30 | 5559
31	0	31 | 25
		32 | 04
		33
		34 | 0
		35
		36 | 6

Note that extra (unused) stems were added to the beginning of the O-line stemplot, making it easier to compare the two distributions. A back-to-back stemplot (see Exercise 1.38) would also be useful for such a comparison. **(c)** The lightest defensive lineman is certainly a low outlier. Among offensive lineman, some students might view the heaviest as an outlier, or perhaps the two heaviest. Even if we ignore the outlier(s), offensive linemen are generally heavier than defensive linemen.

2.7. **(a)** The stock fund varied between about -3.5% and 3.0%. **(b)** The median return for the stock fund was slightly positive—about 0.1%—while the median real estate fund return appears to be close to 0%. **(c)** The stock fund is much more variable—it has higher positive returns, but also lower negative returns.

2.8. No (barely): the *IQR* is $Q_3 - Q_1 = 30 - 10 = 20$ minutes, so we would consider any numbers *greater than* $Q_3 + 1.5 \times IQR = 30 + 30 = 60$ minutes to be outliers.

2.9. Yes: the *IQR* is $Q_3 - Q_1 = 12.6\% - 3.8\% = 8.8\%$, so we would consider any numbers greater than $Q_3 + 1.5 \times IQR = 12.6\% + 13.2\% = 25.8\%$ to be outliers.

2.10. **(a)** The mean is

$$\bar{x} = \frac{3175 + 2526 + 1763 + 1090}{4} = \frac{8554}{4} = 2138.5 \text{ CFU/m}^3.$$

(b) The details of the computation are shown on the right. The variance is

$$s^2 = \frac{2{,}464{,}841}{3} = 821{,}613.\overline{6},$$

and the standard deviation is $s = \sqrt{s^2} \doteq 906.43$ CFU/m^3.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
3175	1036.5	1,074,332.25
2526	387.5	150,156.25
1763	−375.5	141,000.25
1090	−1048.5	1,099,352.25
8554	0	2,464,841

2.11. The means and standard deviations are basically the same: for set A, $\bar{x}_A \doteq 7.501$ and $s_A \doteq 2.032$, while for set B, $\bar{x}_B \doteq 7.501$ and $s_B \doteq 2.031$. Set A is left-skewed, while set B has a high outlier.

Set A		Set B	
3	1	5	257
4	7	6	58
5		7	079
6	1	8	48
7	2	9	
8	1177	10	
9	112	11	
		12	5

2.12. **(a)** Not appropriate: the distribution of percents of foreign-born residents is clearly skewed to the right. (Furthermore, the histogram of this same data set in Figure 1.6, page 16, suggests that there may be a high outlier.) **(b)** $\bar{x}$ and s are fine: the Iowa Test score distribution is quite symmetric and has no outliers. **(c)** Not appropriate: the wood breaking-strength distribution is strongly skewed to the left.

2.13. STATE: How does logging affect tree count?

PLAN: We need to compare the distributions, including appropriate measures of center and spread.

SOLVE: Stemplots are shown below. Based on these, $\bar{x}$ and s are reasonable choices; the means and standard deviations (in units of trees) are given in the table (below, right).

CONCLUDE: The means and the stemplots appear to suggest that logging reduces the number of trees per plot and that recovery is slow (the 1-year-after and 8-years-after means and stemplots are similar).

Never logged		1 year earlier		8 years earlier	
0		0	2	0	4
0		0	9	0	
1		1	2244	1	22
1	699	1	57789	1	5889
2	0124	2	0	2	22
2	7789	2		2	
3	3	3		3	

Group	$\bar{x}$	s
1	23.7500	5.06548
2	14.0833	4.98102
3	15.7778	5.76146

2.14. STATE: Do diplomats from developed countries and those from developing countries differ in the number of unpaid parking tickets?

PLAN: We need to compare the distributions, including appropriate measures of center and spread.

SOLVE: Stemplots are shown below; note that for developed countries, the stems are hundreds digits (split five ways), while in the second stemplot, the stems are the tens digits. Because these distributions are sharply skewed, five-number summaries are the most appropriate choice; they are given in the table (below, right). The maximum number of tickets for developed-country diplomats—a mind-boggling 246.2 for Kuwait—is an outlier.

CONCLUDE: Apart from the outlier, diplomats from developed countries generally had fewer parking tickets than those from developing countries.

Developed countries

```
0 | 00000000000000000000000001111
0 | 33
0 |
0 |
0 |
1 |
1 |
1 |
1 |
1 |
2 |
2 |
2 | 4
```

	Min	Q_1	M	Q_3	Max
Developed	0	0	0.7	7.45	246.2
Developing	0	3.2	9.5	22.80	139.6

Developing countries

```
 0 | 00000000000000111112222223333333344444455556666677888899999999
 1 | 0000111223335566678889
 2 | 014559
 3 | 3446778
 4 | 359
 5 | 289
 6 | 009
 7 | 9
 8 | 14
 9 |
10 |
11 | 079
12 | 4
13 | 9
```

2.15. (c) The mean is $\bar{x} = 43.3$.

2.16. (b) The median is 42.5.

2.17. (a) The five-number summary is Min $= 0$, $Q_1 = 0$, $M = 42.5$, $Q_3 = 76$, Max $= 97$.

2.18. (c) The mean is pulled in the direction of the skew.

2.19. (b) Half the observations lie between the quartiles.

2.20. (c) A boxplot is a picture of the five-number summary.

2.21. (b) The standard deviation is about 37.24.

2.22. (a) Standard deviations can be any nonnegative number.

2.23. (b) s is measured in the same units as the data.

2.24. (a) The median is resistant to outliers.

2.25. The median is \$46,453 and the mean is \$58,886: income distributions will be skewed to the right, so the mean will be larger.

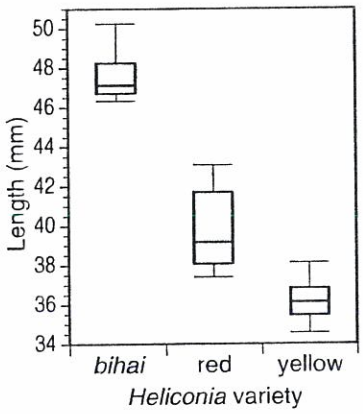
2.26. These distributions are sharply right-skewed, because many—probably most—of those with retirement savings have not saved very much, but a few have saved hundreds of thousands, or even millions.

2.27. The median is at position $\frac{785+1}{2} = 393$, Q_1 is at position $\frac{392+1}{2} = 196.5$ (the average of the 196th and 197th values), and Q_3 is at position $393 + 196.5 = 589.5$ (the average of the 589th and 590th values).

2.28. **(a)** The five-number summary (all quantities in units of pounds) is Min = 23,040, Q_1 = 30,315, M = 31,975, Q_3 = 32,710, Max = 33,650. **(b)** Note the distances between the numbers in the five-number summary: in order, the gaps are 7275, 1660, 735, and 940 pounds. That the first two gaps are larger gives some indication of the left skew.

2.29. The five-number summaries (all in millimeters) are

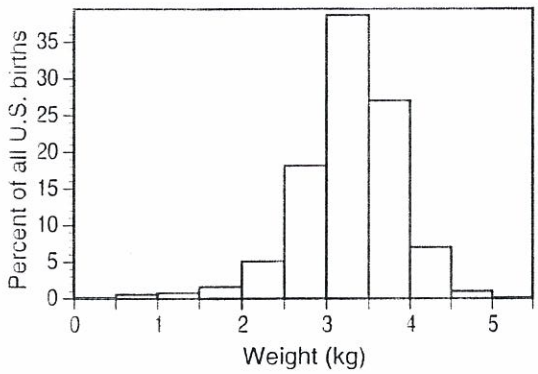
	Min	Q_1	M	Q_3	Max
bihai	46.34	46.71	47.12	48.245	50.26
red	37.40	38.07	39.16	41.69	43.09
yellow	34.57	35.45	36.11	36.82	38.13



Although we lose the detail of the individual measurements visible in the stemplots, we can draw essentially the same conclusions: *H. bihai* is clearly the tallest variety—the shortest *bihai* was more than 3 mm taller than the tallest red. Red is generally taller than yellow, with a few exceptions. Another noteworthy fact: the red variety is more variable than either of the other varieties.

2.30. $M = 2$, $Q_1 = 1$, and $Q_3 = 4$ servings: we can use the frequencies shown in the histogram to reconstruct the (sorted) data list; it begins with 15 zeros, then 11 ones, etc. The median is halfway between the 37th and 38th numbers in this list; because the 27th through 41st numbers in the list are all "2," that is the median. The first quartile is the 19th number in the list, and Q_3 is the 56th number.

2.31. **(a)** The total number of births in a year will vary greatly from one country to another; it would be difficult to compare counts for a small country with those of a large country. **(b)** There were 4,134,370 total births recorded in the table; divide each count by this number to compute the percents. For example, the first weight class accounts for $\frac{6599}{4{,}134{,}370} \doteq 0.16\%$. **(c)** The positions and weight classes are given in the table below.



Measurement	Position	Weight class
Median	$\frac{4{,}134{,}370 + 1}{2} = 2{,}067{,}185.5$	3,000 to 3,499 grams
Q_1	$\frac{2{,}067{,}185 + 1}{2} = 1{,}033{,}593$	2,500 to 2,999 grams
Q_3	$2{,}067{,}185 + 1{,}033{,}593 = 3{,}100{,}778$	3,500 to 3,999 grams

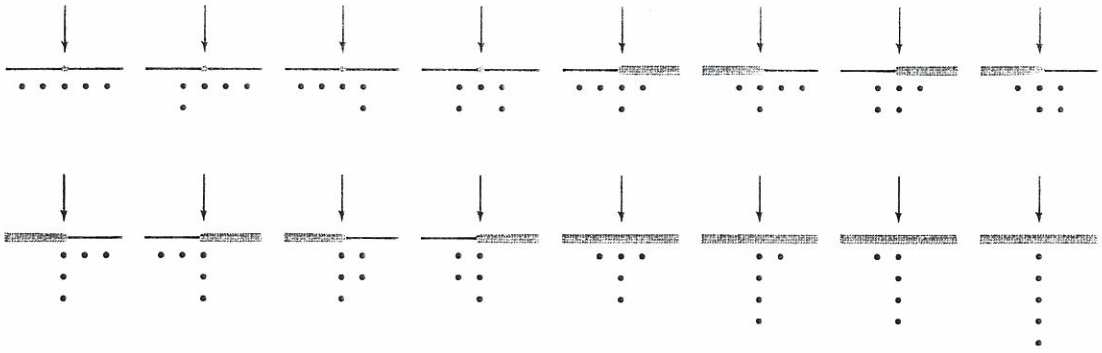
2.32. **(a)** $\bar{x}$ and s are appropriate for symmetric distributions with no outliers. **(b)** The table on the right shows the effect of removing these outliers; both $\bar{x}$ and s decrease.

	Women		Men	
	$\bar{x}$	s	$\bar{x}$	s
Before	165.2	56.5	117.2	74.2
After	158.4	43.7	110.9	66.9

2.33. **(a)** The mean (green arrow) moves along with the moving point (in fact, it moves in the same direction as the moving point, at one-third the speed). At the same time, as long as the moving point remains to the right of the other two, the median (red arrow) points to the middle point (the rightmost nonmoving point). **(b)** The mean follows the moving point as before. When the moving point passes the rightmost fixed point, the median slides along with it until the moving point passes the leftmost fixed point, then the median stays there.

2.34. **(a)** There are several different answers, depending on the configuration of the first five points. *Most students* will likely assume that the first five points should be distinct (no repeats), in which case the sixth point *must* be placed at the median. This is because the median of 5 (sorted) points is the third, while the median of 6 points is the average of the third and fourth. If these are to be the same, the third and fourth points of the set of 6 must both equal the third point of the set of 5.

The diagram on the next page illustrates all of the possibilities; in each case, the arrow shows the location of the median of the initial five points, and the shaded region (or dot) on the line indicates where the sixth point can be placed without changing the median. Notice that there are four cases where the median does not change regardless of the location of the sixth point. (The points need not be equally spaced; these diagrams were drawn that way for convenience.)



(b) Regardless of the configuration of the first 5 points, if the sixth point is added so as to leave the median unchanged, then in that (sorted) set of 6, the third and fourth points must be equal. One of these 2 points will be the middle (fourth) point of the (sorted) set of 7, no matter where the seventh point is placed.

Note: *If you have a student who illustrates all possible cases above, then it is likely that the student (1) obtained a copy of this solutions manual, (2) should consider a career in writing solutions manuals, (3) has too much time on his or her hands, or (4) both 2 and 3 (and perhaps 1) are true.*

2.35. (a) A stemplot is shown; a histogram would also be appropriate. The expected right skew is clearly evident; the split stems emphasize the skewness by showing the gaps. The main peak occurs from 50 to 150 days—the guinea pigs that lived more than 500 days seem to be outliers. **(b)** Because of the skew, choose the five-number summary:

43 82.5 102.5 151.5 598

(all measured in days). The difference between Q_3 and the maximum is relatively much larger than the other differences between successive numbers. This indicates a large spread among the high observations—that is, it shows that the data are skewed to the right.

```
0 | 44
0 | 55555667778888888888899999999
1 | 00000000000111222233344 4
1 | 56777899
2 | 1144
2 |
3 | 2
3 | 8
4 | 0
4 |
5 | 12
5 | 9
```

2.36. Students observations will vary. As in the United States, weekend births are less common. Additionally, Monday stands out as slightly lower than the rest of the weekdays. The means in Exercise 1.5 also suggested that this might be the case, but the additional detail visible in the boxplots gives stronger evidence that this is true. (For example, Monday's third quartile is below the median count for the other four weekdays.)

2.37. The mean is 8.4%—much lower than the true national value of 12.5%. The largest states in population have high percents of foreign-born residents (for example, California has 27.2% and Texas, 15.9%). When we average the 51 states, smaller states—some of which have lower percents of foreign-born residents—are overrepresented (given too much weight).

A simplified example illustrates what is happening here: suppose that a two-year college has 1000 students, of which 600 are first-year students and 400 are sophomores. If 60% of the first-year students and 50% of the sophomores are women, note that the percent of

women at the college is *not* 55%, the "straight average" of 50% and 60%. Instead, it is 56%, the "weighted average," because there are 360 first-year women and 200 sophomore women.

Note: *A good supplemental exercise for stronger students is to ask how to find the* correct *national average from this data. To do this, find a list of state populations (e.g., from* `http://www.census.gov/popest/states/`*). Ideally, one should use population figures from the same year as the data in Table 1.1, but any recent year will produce fairly good results. Next, compute the* number *of foreign-born residents in each state by taking the appropriate percent of each state population. Add up these numbers to find the total foreign-born population in the U.S., then divide by the total population.*

2.38. Households with no credit cards, as well as those which pay off the balance each month, have no credit card debt (see note below). If we list the credit card debt figures for all American households, more than half of the numbers in that list equal zero, so the median is zero.

Note: *One might question whether someone who routinely pays off the balance on his or her credit card really has "no credit card debt." For more detail about this, see the Survey of Consumer Finance, conducted by the Federal Reserve Board. For the purposes of that report, credit card debt "excludes purchases made after the most recent bill was paid."*

At the time this solution was written, the 2004 report was the most recent report available; there we learn that 25.1% of households had no credit cards, and another 31.5% had cards but did not carry a balance. Therefore, about 56.6% of American households had no credit card debt in 2004.

2.39. **(a)** One possible answer is 1, 1, 1, 1. **(b)** 0, 0, 10, 10. **(c)** For (a), any set of four identical numbers will have $s = 0$. For (b), the answer is unique; here is a rough description of why. We want to maximize the "spread-out-ness" of the numbers (which is what standard deviation measures), so 0 and 10 seem to be reasonable choices based on that idea. We also want to make each individual squared deviation—$(x_1 - \bar{x})^2$, $(x_2 - \bar{x})^2$, $(x_3 - \bar{x})^2$, and $(x_4 - \bar{x})^2$—as large as possible. If we choose 0, 10, 10, 10—or 10, 0, 0, 0—we make the first squared deviation 7.5^2, but the other three are only 2.5^2. Our best choice is two at each extreme, which makes all four squared deviations equal to 5^2.

2.40. Answers will vary. Typical calculators will carry only about 12 to 15 digits; for example, a TI-83 fails (gives $s = 0$) for 13-digit numbers (the TI-83+ does somewhat better). Excel (at least the version I checked) gives $s = 0$ for nine-digit numbers. The (old) version of Minitab used to prepare these answers fails at 100,000,001 (nine digits).

2.41. Because the mean is to be 7, the five numbers must add to 35. Also, the third number (in order from smallest to largest) must be 10 because that is the median. Beyond that, there is some freedom in how the numbers are chosen.

Note: *It is likely that most students will interpret "positive numbers" as meaning positive integers only, which leads to eight possible solutions, shown below.*

1	1	10	10	13
1	3	10	10	11
1	1	10	11	12
1	4	10	10	10
1	2	10	10	12
2	2	10	10	11
1	2	10	11	11
2	3	10	10	10

2.42. The simplest approach is to take (at least) 6 numbers; call them (in increasing order) a, b, c, d, e, f. For this set, $Q_3 = e$; we can cause the mean to be larger than e simply by choosing f to be *much* larger than e. For example, if all numbers are nonnegative, $f > 5e$ would accomplish the goal because then $\bar{x} = (a + b + c + d + e + f)/6 > (e + f)/6 > (e + 5e)/6 = e$.

2.43. PLAN: We need to display the salary distribution, including appropriate measures of center and spread.

SOLVE: A stemplot is shown; a histogram would also be appropriate. Because the distribution is clearly skewed to the right, we should report the five-number summary rather than $\bar{x}$ and s:

\$380,000 \$424,500 \$2,800,000 \$8,625,000 \$17,016,381

While they are poor choices for this distribution, some students might compute the mean and standard deviation: $\bar{x} \doteq \$5{,}066{,}389$ and $s \doteq \$5{,}234{,}351$. Using the $1.5 \times IQR$ criterion, none of the salaries are outliers. (The question of outliers is also asked in Exercise 2.51.)

0	0000000001
0	2223
0	5
0	666
0	89
1	1
1	33
1	4
1	7

CONCLUDE: Student comments will vary. Some observations: player salaries range from \$380,000 to over \$17 million; nine players make less than \$1 million; the total payroll is over \$126 million, with the top five salaries accounting for over half that total.

2.44. STATE: How have returns on stocks behaved over the years?

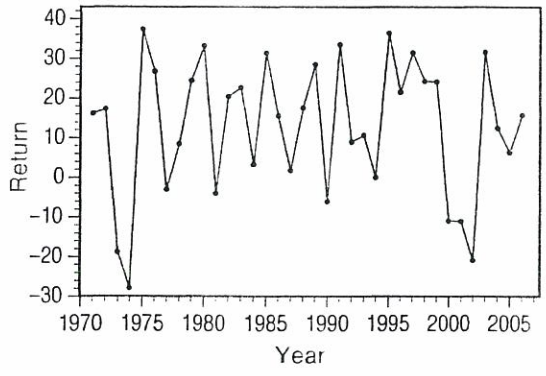
PLAN: We should examine the distribution through graphs and numerical summaries. Because this is a variable that changes over time, we should also look at a time plot.

SOLVE: A stemplot and time plot are shown below. Because the stemplot appears to be somewhat skewed to the left, the five-number summary is preferred, but some students may compute the mean and standard deviation:

$\bar{x}$	s	Min	Q_1	M	Q_3	Max
12.79%	17.28%	−27.87%	0.905%	15.98%	25.585%	37.38%

CONCLUDE: The time plot shows no particular pattern. From the stemplot and the summary statistics, we see that returns have typically been positive (in 28 of the 36 years listed), but the wide fluctuations are an indication of the risk involved for short-term investing.

−2	7
−2	0
−1	8
−1	00
−0	6
−0	32
0	013
0	689
1	02
1	55677
2	012444
2	68
3	11133
3	67



2.45. STATE: What effect do lavender and lemon odors have on customer spending?
PLAN: We will compare the three distributions through graphs and numerical summaries.
SOLVE: Side-by-side stemplots are shown below; it would also be appropriate to produce histograms or boxplots. All three stemplots show clustering—presumably because of the pricing of the items on the menu; for example, perhaps a medium pizza costs €15.90. The clustering makes it hard to comment on shape, but the lavender distribution is skewed to the right. For that reason, the five-number summary is preferred, but some students may compute the mean and standard deviation:

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
No odor	€17.513	€2.359	€12.9	€15.9	€17.2	€18.5	€25.5
Lemon	€18.157	€2.218	€15.9	€15.9	€18.5	€18.5	€25.9
Lavender	€21.123	€2.345	€18.5	€18.5	€21.9	€22.3	€25.9

CONCLUDE: There was little difference in spending between the control and lemon-scented evenings, but spending was noticeably higher with the lavender odor.

No odor		Lemon		Lavender	
12	9	12		12	
13		13		13	
14		14		14	
15	99999999999999	15	999999999	15	
16		16		16	
17		17		17	
18	555555555555	18	5555555555555555	18	55555555555
19		19		19	
20	5	20		20	7
21	9	21	59	21	5599999999
22		22		22	3558
23		23		23	
24		24		24	99
25	5	25	9	25	59

2.46. STATE: How do lean and obese people differ in time spent in activity and in time spent lying down?

PLAN: We will compare each pair of distributions using graphs and numerical summaries.

SOLVE: On the right are two back-to-back stemplots; histograms or boxplots could also be used. None of the stemplots show any particular skewness, so either means and standard deviations or five-number summaries would be suitable. All values in the table are in units of minutes.

Time active

Lean		Obese
	2	66
1	3	4
7	3	56
	4	1112
	4	6
410	5	
875	5	
0	6	
7	6	

Time lying down

Lean		Obese
9	3	
	4	1
	4	
5	4	44
6	4	6
8	4	
10	5	001
33	5	23
5	5	
6	5	6

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Lean/Active	525.751	107.121	319.212	504.700	549.522	584.644	677.188
Obese/Active	373.269	67.498	260.244	347.375	388.885	416.531	464.756
Lean/Lying down	501.646	52.045	396.962	467.700	510.290	537.362	567.006
Obese/Lying down	491.743	46.593	412.919	448.856	507.456	521.044	563.300

CONCLUDE: In both the stemplots and the numerical summaries, we observe that lean subjects spent more active time than the obese subjects. There was little difference in time spent lying down.

2.47. STATE: How does increasing compression affect soil penetrability?

PLAN: We need to compare the distributions, including appropriate measures of center and spread.

SOLVE: Shown are three stemplots; it would also be appropriate to produce histograms or boxplots (five-number summaries are given below).

Here are numerical summaries; students may give all or just some of these in response to this question. The slight skew evident in the "Intermediate" stemplot makes the five-number summary preferable, but note that the mean and median for that group are nearly identical.

Compressed	Intermediate	Loose
2 | 67777	2 |	2 |
2 | 8888899999	2 | 99	2 |
3 | 00011	3 | 0111111	3 |
	3 | 2333	3 |
	3 | 4445	3 |
	3 | 6	3 |
	3 | 8	3 | 9999
	4 |	4 | 0011111
	4 | 2	4 | 22233
		4 | 44
		4 |
		4 | 89

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Compressed	2.9075	0.1390	2.68	2.795	2.880	2.99	3.18
Intermediate	3.3360	0.3190	2.92	3.130	3.310	3.45	4.26
Loose	4.2315	0.2713	3.94	4.015	4.175	4.32	4.91

CONCLUDE: Both the graphs and the numerical summaries suggest that soil penetrability is greatest for loose soil and least for compressed soil.

2.48. STATE: Is bone mineral loss greater among the breast-feeding women?
PLAN: We need to compare the distributions, including appropriate measures of center and spread.
SOLVE: Shown are two stemplots; it would also be appropriate to produce two histograms, a back-to-back stemplot (see the solution to Exercise 1.38), or two boxplots (five-number summaries are given below). Note that for negative stems, the leaves are in "reverse" order, so that they increase from left to right like the leaves on the positive stems. Students who create stemplots by hand might not consider this issue.

BF women		Other women
−8	3	−8
−7	80	−7
−6	88552	−6
−5	97633221	−5
−4	9977430	−4
−3	86310	−3
−2	755322110	−2 2
−1	800	−1 65
−0	83	−0 64442111
0	234	0 0379
1	7	1 0127
2	2	2 249

Here are numerical summaries; students may give all or just some of these in response to this question.

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
BF women	−3.5872%	2.5056%	−8.3%	−5.3%	−3.8%	−2.1%	2.2%
Other women	0.3091%	1.2983%	−2.2%	−0.4%	−0.05%	1.1%	2.9%

CONCLUDE: Both the graphs and the numerical summaries suggest that breast-feeding women lose calcium.

2.49. **(a)** The five-number summary is Min = 6.8%, Q_1 = 12.1%, M = 12.8%, Q_3 = 13.4%, Max = 16.8%. **(b)** The IQR is $Q_3 - Q_1$ = 13.4% − 12.1% = 1.3%, so we would consider to be outliers any numbers below 12.1% − 1.95% = 10.15% or above 13.4% + 1.95% = 15.35%. Five states are flagged as low outliers, and one as a high outlier. (Those states are Alaska, Utah, Georgia, Texas, and Colorado on the low end, and Florida on the high end.)

2.50. See also the solution to Exercise 1.36. **(a)** The five-number summary (in units of metric tons per person) is Min = 0.1, Q_1 = 0.95, M = 3.3, Q_3 = 7.4, Max = 19.6. The evidence for the skew is in the large gaps between the higher numbers; that is, the differences $Q_3 - M$ = 4.1 and Max − Q_3 = 12.2 are large compared to Q_1 − Min = 0.85 and $M - Q_1$ = 2.35. **(b)** The IQR is $Q_3 - Q_1$ = 6.45, so outliers would be less than −8.725 or greater than 17.075. According to this rule, only the United States and Australia qualify as outliers, but it seems reasonable to include Canada as well.

0	0000000000001111111
0	222333333
0	4455
0	6667777
0	8999
1	0
1	3
1	
1	7
1	89

2.51. See also the solution to Exercise 2.43. We find Q_1 = \$424,500 and Q_3 = \$8,625,000, so the interquartile range is IQR = \$8,200,500. Outliers are those salaries above \$20,925,750; there are no such salaries.

2.52. See also the solution to Exercise 2.44. We find Q_1 = 0.905% and Q_3 = 25.585%, so IQR = 24.68% and 1.5 × IQR = 37.02%. None of the returns would be considered outliers by this rule, because all of them fall in the range $Q_1 - 1.5 \times IQR$ = −36.115% to $Q_3 + 1.5 \times IQR$ = 62.605%.