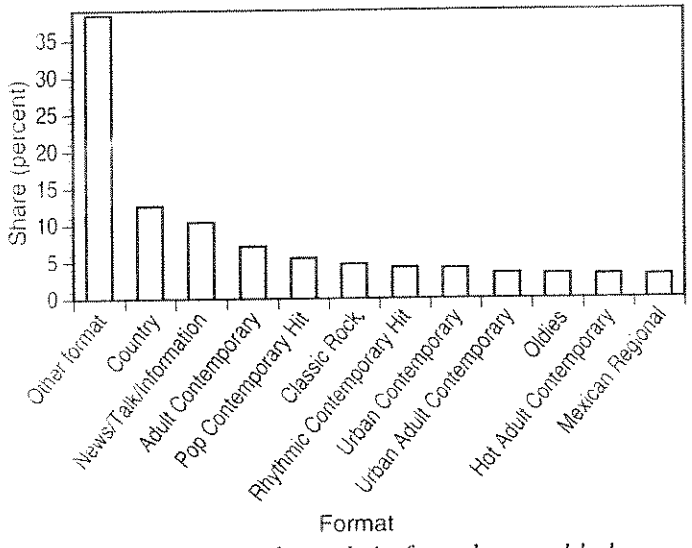
Chapter 1 Solutions

1.1. **(a)** The individuals are vehicles (or "cars"). **(b)** The variables are make/model (categorical), vehicle type (categorical), transmission type (categorical), number of cylinders (quantitative), city MPG (quantitative), and highway MPG (quantitative).

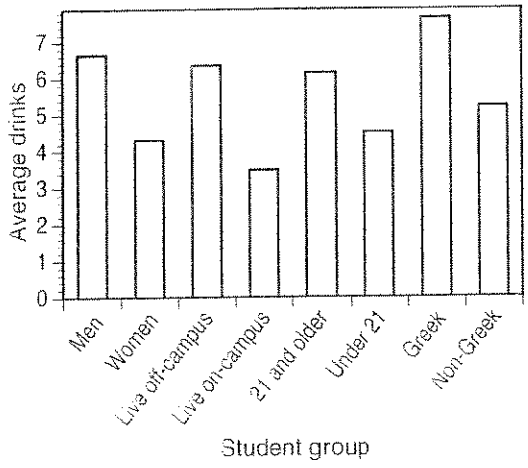
1.2. Possible categorical variables: year in school, gender, major. Possible quantitative variables: age (years), time watching TV (hours), time in class (hours), time sleeping (hours), time studying (hours—or perhaps minutes).

1.3. **(a)** The given percents add up to 61.6%, so the remaining 38.4% of the audience listens to other formats. **(b)** The bar graph is shown on the right. **(c)** A pie chart could be used (provided "Other format" is included), because these percents represent nonoverlapping parts of a single whole (the radio listening audience).



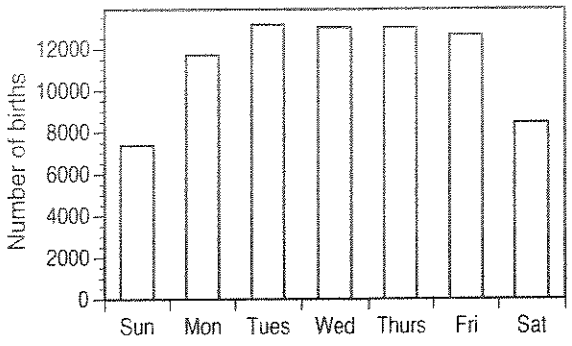
Note: *In the bar graph, the "Other format" bar has been placed at the beginning, so that the bars are decreasing in height from left to right. Some students might instead place it at the end; in fact, they could choose to rearrange the bars in any order they wish (e.g., alphabetically).*

1.4. **(a)** A pie chart is not appropriate because these numbers do not represent nonoverlapping portions of a single group. In fact, they are not even *percents*—they are average counts of drinks consumed. **(b)** The bar graph is shown on the right.

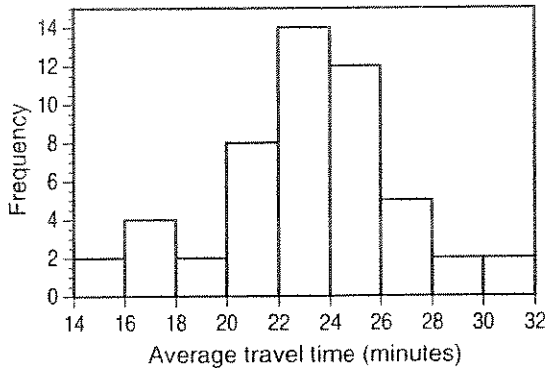


1.5. A pie graph could also be made, but the relative heights of the bars are easier to compare than the relative sizes of the "slices" of the pie.

The most likely explanation for the lower weekend numbers is that, when a birth is "planned" (either by inducement or cesarean section), it is usually scheduled for a weekday—perhaps more due to the preferences of the physician or midwife.



1.6. With the intervals 14.0–15.9, 16.0–17.9, etc., the histogram should look like the one on the right. When asked to make intervals that are 2 minutes wide, some students might fail to read the rest of the instructions and use 14–16, 16–18, etc., which causes confusion about where to place the three states (Minnesota, Mississippi, and Oklahoma) that fall on an interval boundary. If a student's histogram looks different from this one, that may be the reason.



1.7. **(a)** The applet creates a histogram with 23 classes. **(b)** It is possible to get to one class ranging from 1.20 to 27.30 (not a very useful histogram). **(c)** The most classes the applet will allow is 46; the largest class has 5 observations. **(d)** Choices will vary; anything from about 10 to 30 classes is reasonable.

1.8. The distribution is roughly symmetric. Based on the histogram, the center is near 23 minutes, and the spread is from 14 to 31.9 minutes. If we look at the actual data, we find that the center (median) is 23.4 minutes, and the times range from 15.5 to 30.9 minutes. See also the solution to Exercise 1.6.

1.9. **(a)** Women in the District of Columbia are more likely to be career-oriented. (In general, we might expect that women in large cities are less likely to be married. The states consist of some large cities mixed with smaller towns and rural areas, but D.C. is essentially one big city.) **(b)** The midpoint is between 24% and 26%: the first three bars represent (respectively) 5, 7, and 14 states, so the 26th state is included in the third bar. The spread is from 20% to 34% (ignoring the outlier).

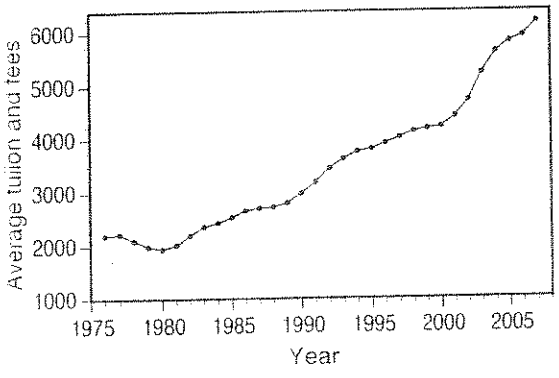
1.10. See also the solution to Exercise 1.8. In the stemplot, the 26th observation is underlined. The midpoint is 23.4 minutes, and the times range from 15.5 to 30.9 minutes.

```
15 | 59
16 |
17 | 6779
18 | 25
19 |
20 | 017889
21 | 28
22 | 01333499
23 | 445669
24 | 01266
25 | 0012569
26 | 689
27 | 39
28 |
29 | 12
30 | 69
```

1.11. The distribution is somewhat right-skewed, although with the United States (a high outlier) removed, the remaining data are relatively symmetric. The center (median) is \$1954 per person; aside from the U.S., health care spending ranges from \$419 to \$3809 per person.

```
0 | 4
0 | 66777888
1 | 11233
1 | 78999
2 | 012334
2 | 578899
3 | 0001
3 | 88
4 |
4 |
5 |
5 | 7
```

1.12. **(a)** The time plot is shown on the right. **(b)** After adjusting for inflation, tuition and fees have increased, apart from a slight dip in the late 1970s. **(c)** There are no outliers in the plot. As noted in (b), charges decreased at the beginning of this time period. From about 2000 to 2004, costs increased more rapidly (about \$200 to \$500 per year) before returning to the previous rate of about \$100 to \$200 per year.



1.13. (a) The individuals are the students.

1.14. (c) Either a pie chart or a bar graph would be appropriate.

1.15. (c) Sex is a categorical variable, and college debt is a quantitative variable.

1.16. (b) People with the same zip code live in the same area, but one cannot compute (for example) an "average zip code."

1.17. (b) The tick marks are 4 units apart, the first bar begins at 20%, and there are 2 bars between each tick.

1.18. (a) With a range of 0 to 97, we should make a stemplot with the tens digit as the stem and the ones digit as the leaf, using all stems from 0 to 9.

1.19. (a) The lowest percent is 6.8% (a stem of 6, with leaf 0.8).

1.20. (b) "Skewed to the left" means that distribution trails off toward lower numbers. (Note that in a stemplot, the lower numbers are at the *top* of the list.)

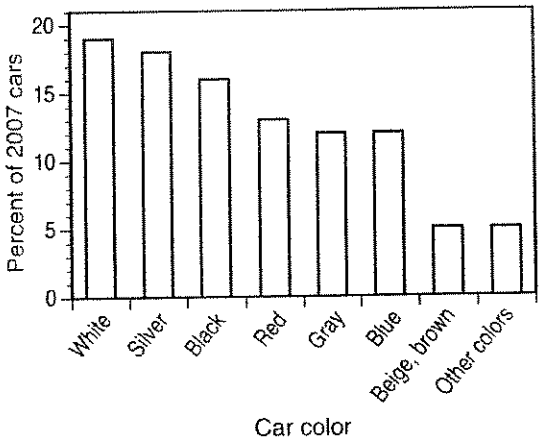
1.21. (a) There are 51 numbers represented in the stemplot, so the center (median) is the 26th number. Counting from the start of the list, the 26th leaf is the third "8" on the stem "12." (Of course, one can also count up from the end of the list.)

1.22. (c) These housing prices are (fairly sharply) right-skewed.

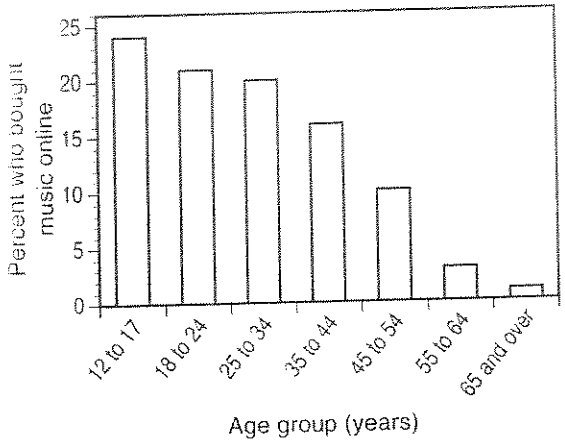
1.23. **(a)** The individuals are medical students. **(b)** The data set has five variables. Of those, three (medical school, sex, and specialty) are categorical, and the other two (age and USMLE) are quantitative.

1.24. **(a)** Type of wood is categorical. **(b)** Type of water repellent is categorical. **(c)** Paint thickness is quantitative. **(d)** Paint color is categorical. **(e)** Weathering time is quantitative.

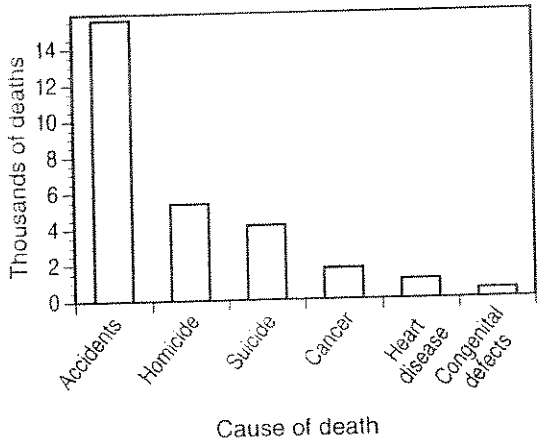
1.25. **(a)** The given percents add up to 95%, so 5% must be some other color. **(b)** The bar graph shown includes the "Other colors" category, although some students might leave it off. With the "Other" category, a pie chart could be used because these percents show parts of a whole (if we assume, as we did in part (a), that a car can be only one color).



1.26. **(a)** The given percents represent fractions of different age groups, rather than parts of a single whole. **(b)** The bar graph is shown on the right.

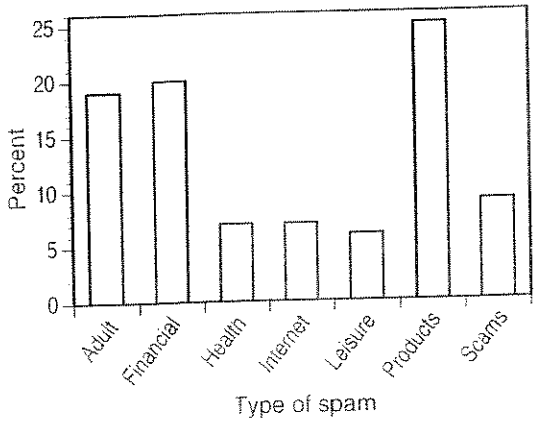


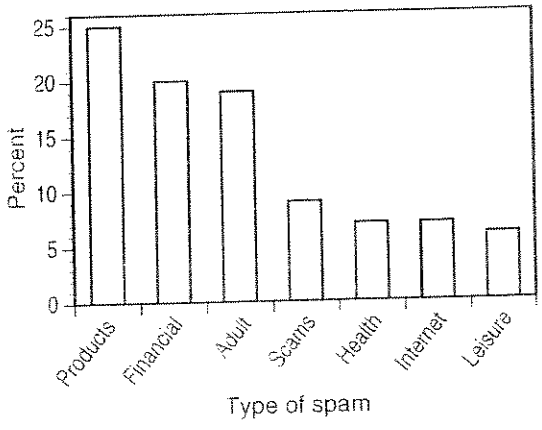
1.27. **(a)** The bar graph is shown on the right. **(b)** In order to make a pie chart, we would need to know the total number of deaths in this age group (so that we could compute the number of deaths due to other causes).



1.28. Estimates will vary, but should be close to the actual reported numbers (which can be found at the Census Bureau Web site): 64% Mexican, 9% Puerto Rican.

1.29. The two bar graphs are shown below.





1.30. The distribution is skewed to the right, spread from 0 to 8 servings, and the center is about 2 or 3 servings of fruit. (The median number of servings is 2, but student judgments of the "center" may vary from the median.) About 35% (26 out of 74) ate 0 or 1 ("fewer than two") servings of fruit.

1.31. (a) The distribution is fairly symmetric—perhaps slightly left-skewed—with center near 110. Apart from the four scores in the 70s, IQs range from the mid 80s to the high 130s. **(b)** $\frac{62}{78} \doteq 79.49\%$ scored above 100. (It is easier to count the 16 students who scored 100 or less, then subtract from 78.)

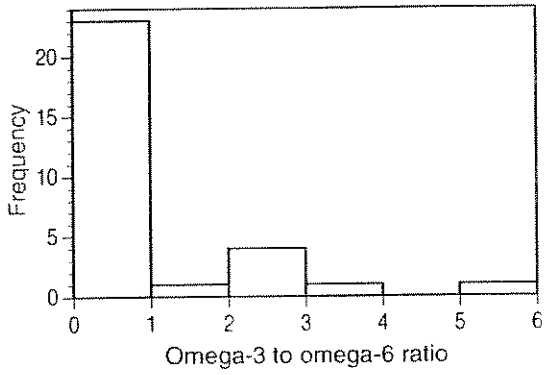
1.32. The text mentions the "extreme low outlier," so students might infer that there is only one outlier. Other students might also consider the second lowest return to be an outlier (which is a reasonable opinion). **(a)** The distribution is slightly skewed to the left, although not strikingly so if one ignores the low outlier(s). **(b)** The center is between 0% and 2.5%; student estimates will vary. **(c)** The highest return was between 10% and 12.5%. Ignoring only one low outlier, the lowest return was between −17.5% and −15%. If we ignore two low outliers, the lowest return was between −12.5% and −10%. **(d)** About 37% of these months (102 out of 273) had negative returns. Student estimates of the count may vary, but the percentage should be roughly 35–40% in every case.

Note: *By examining the raw data, we find that the median return was 1.2%, the largest return was 12.5%, and the three lowest returns were −23%, −15.99%, and −10.63%. In addition to being in the data files accompanying the text, these numbers are in the first column of the file "Fama/French Benchmark Factors(Monthly)" found at*

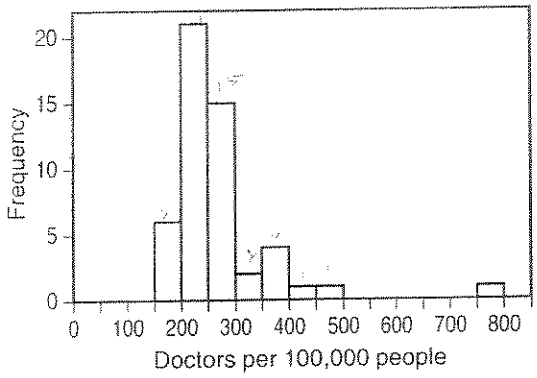
`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

1.33. **(a)** is variable 4. Minutes spent studying would likely be skewed to the right (many study for a short time, a few study longer). **(b)** is variable 2, and **(c)** is variable 1—unless this was a particularly unusual class! We would expect that male/female counts should be somewhat close, while right-handed students should outnumber lefties substantially. (Roughly 10% to 15% of the population as a whole is left-handed.) **(d)** is variable 3. One would expect a fair amount of variation in student heights, but no particular skewness to such a distribution.

1.34. **(a)** The histogram is shown on the right. **(b)** The distribution is sharply skewed to the right. Only 7 of the 30 food oils have more omega-3 than omega-6, so most oils do not have the desired ratio. **(c)** The fish oils are 5 of the 7 oils that have a ratio over 1.



1.35. **(a)** In a state with many people, more doctors are needed to serve the larger population. For example, having 1000 doctors in Rhode Island would be very different from having 1000 doctors in California. **(b)** The distribution is clearly skewed to the right, with the District of Columbia a high outlier. The states all have numbers between 169 and 450; D.C. is different from the states in that it includes very little area that would be considered "rural," where we would expect the density of doctors would drop off considerably. (Observe that the states with large cities tend to have high numbers; D.C. is an extreme case, because it consists mainly of a large city.)



1.36. **(a)** Totals emissions would almost certainly be higher for very large countries; for example, we would expect that even with great attempts to control emissions, China (with over 1 billion people) would have higher total emissions than the smallest countries in the data set. **(b)** We see a strong right skew with a peak from 0 to 0.2 metric tons per person, and a smaller peak from 0.6 to 0.8. The three highest countries (the United States, Canada, and Australia) appear to be outliers; apart from those countries, the distribution is spread from 0 to 14 metric tons per person.

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

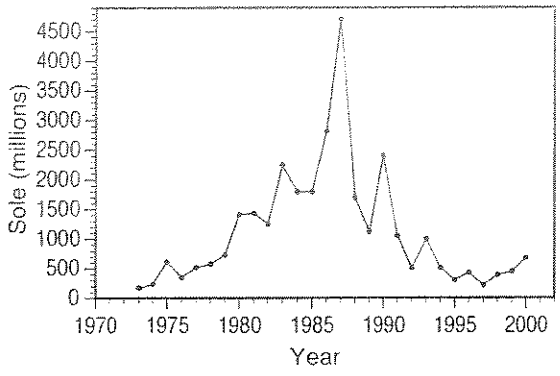
1.37. Shown are two versions of this stemplot. For the first, we have (as the text suggests) rounded to the nearest 100; for the second, we have trimmed the numbers (dropped the last two digits). The distribution is clearly skewed to the right, with a high outlier (4700 million sole, from 1987). The center is around 700 million, and the spread is from 173 million to 4700 million (2809 million, if we omit the outlier).

0	22233444	0	12233344
0	5556677	0	55556679
1	001344	1	01244
1	788	1	777
2	24	2	24
2	8	2	8
3		3	
3		3	
4		4	
4	7	4	7

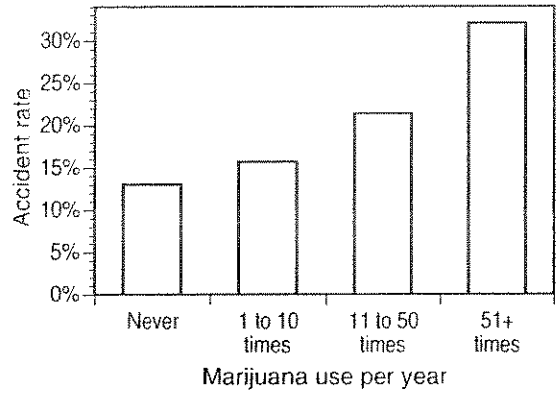
1.38. **(a)** Not only are most responses multiples of 10; many are multiples of 30 and 60. Most people will round their answers when asked to give an estimate like this; in fact, the most striking answers are ones such as 115, 170, or 230. The students who claimed 360 minutes (6 hours) and 300 minutes (5 hours) may have been exaggerating. (Some students might also "consider suspicious" the student who claimed to study 0 minutes per night. As a teacher, I can easily believe that such students exist.) **(b)** The stemplots suggest that women (claim to) study more than men. The approximate centers are 175 minutes for women and 120 minutes for men.

Women		Men
	0	033334
96	0	66679999
22222221	1	2222222
888888888875555	1	558
4440	2	00344
	2	
	3	0
6	3	

1.39. The time plot shows that the number of recruits peaked in the mid-1980s, and in recent years has fallen back to levels similar to those in the 1970s.



1.40. **(a)** If the four groups were roughly equal in size, then it would be valid to compare accident counts. However, because those who never used marijuana (or at least denied using it) accounted for about half of the group, we would expect the number of accidents for that group to be larger—and it is the largest of the four. By computing accidents per driver (e.g., $\frac{59}{452} \doteq 13\%$), we can compare the relative risks for the four unequal-sized groups. **(b)** Increasing marijuana usage is associated with increasing accident rates. (Students will not necessarily use the technical language of association or correlation, but should somehow acknowledge that when one number is high, the other is, too.)



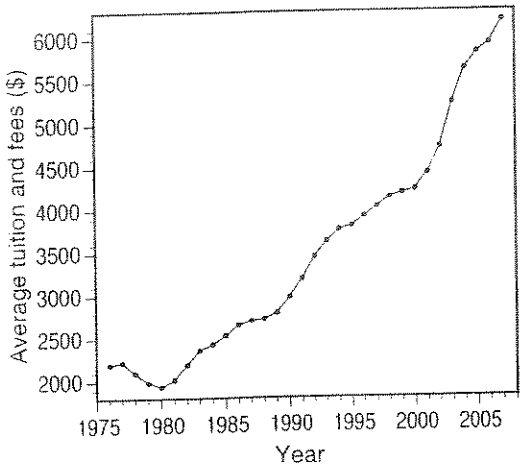
1.41. Sketches will vary. The distribution of coin years would be left-skewed because newer coins are more common than older coins.

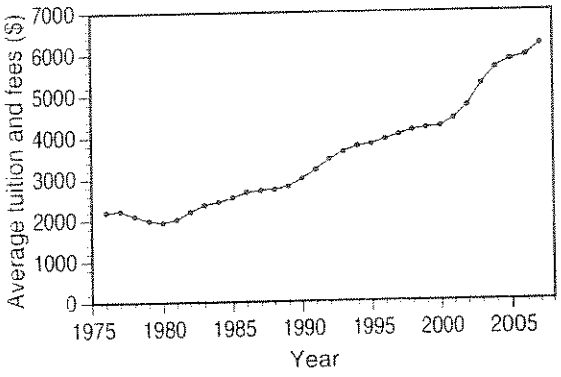
1.42. **(a)** The two stemplots are shown on the right. Student preferences might vary, but the split stems of the first stemplot show more detail. **(b)** The distribution is relatively symmetric, with center near 780 mm (the median is 784 mm), and spread from 604 to 957 mm. **(c)** Monsoon rainfall was below average in 18 of the 23 El Niño years, and only exceeded 900 mm in one of those years.

6	03
6	557
7	01244
7	88999
8	113
8	667
9	0
9	6

6	03557
7	0124488999
8	113667
9	06

1.43. Here are examples that do the trick. With most software that can create such plots, it is very easy to achieve the desired effects by using the mouse to resize the plot.





1.44. **(a)** Housing starts are highest in the spring, and lowest in the winter. **(b)** With the exception of the final part of the graph, there is an overall increasing trend. **(c)** The downturn is clearly visible; the increases of the first 15 years shown in the graph are essentially erased in the last year and a half. In particular, the spring peak in 2007 barely rose above the January 2006 low.

1.45. **(a)** One possible histogram is shown below on the left; the exact appearance of the histogram will vary with the choice of interval width. The middle count is 13 people bitten per year. **(b)** The time plot shows a lot of fluctuation, but suggests that bite counts have been generally higher in recent years. From 1986 to 2007, the annual count of people bitten has exceeded the midpoint 15 times (and the count was *equal to* the midpoint in an additional four years). By contrast, the bite count exceeded the median only once (barely) in the first 14 years.