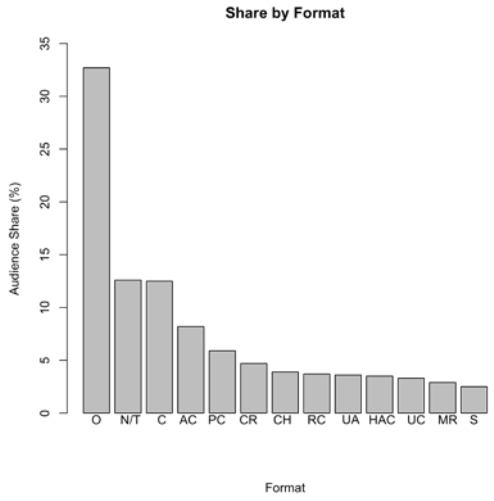
Chapter 1 Solutions

1.1: (a) The individuals are the car makes and models. (b) For each individual, the variables recorded are Vehicle type (categorical), Transmission type (categorical), Number of cylinders (usually treated as quantitative), City mpg (quantitative), Highway mpg (quantitative), and Carbon footprint (tons, quantitative).

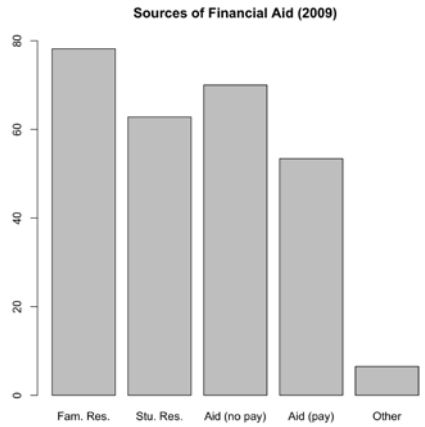
1.2: Answers will vary. Some possible categorical variables: Whether or not student plays on a sport team or club; Sex; Whether or not the student smokes; Attitude about exercise, etc. Some possible quantitative variables: Weight (kilograms or pounds), Height (centimeters or inches); Resting heart rate (beats per minute); Body mass index (kg/m^2 or lb/ft^2).

1.3: (a) These shares sum to 67.3%. Hence, 100% – 67.3% = 32.7% of the radio audience listens to stations with other formats. (b)



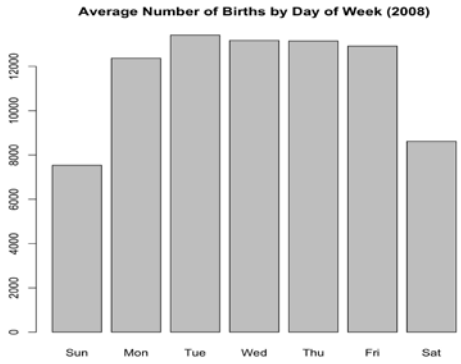
(c) A pie chart would be inappropriate based only on the data presented because the areas of the pie wedges would be relative to the total of the categories presented (67.3%). If you include a wedge for "other" that accounts for 32.7% of the total, a pie chart would be reasonable.

1.4: (a) Individuals fall into more than one of the categories. (b) A bar graph follows:



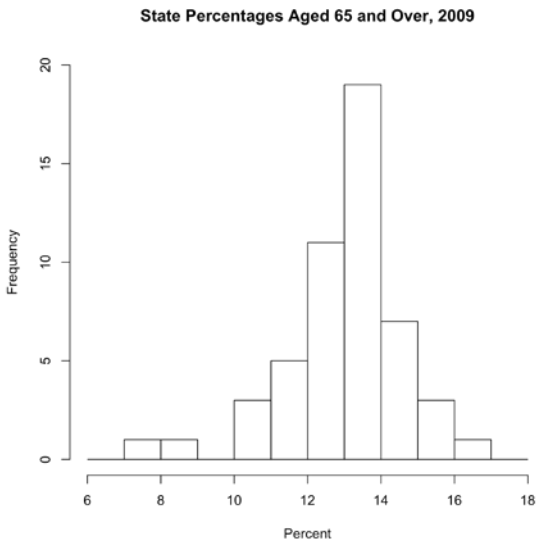
1.5:

A pie chart would make it more difficult to distinguish between the weekend days and the weekdays. Some births are scheduled (induced labor, for example), and probably most are scheduled for weekdays.



1.6:

Make this histogram by hand, as the instructions suggest:



1.7: Use the applet to answer these questions.

1.8: The distribution is roughly symmetric, although one could argue that it is very slightly left-skewed. The center is around 13%. The statewide percentages range from about 7% to 17%. One state has only about 7% of its residents aged 65 or older (Alaska, 7.0%), while another has almost 17% of its residents aged 65 or older (Florida, 16.9%).

1.9: (a) The District of Columbia is the center of Federal government and hence has many, many young professionals, many of whom may not be married. (b) The 26th ordered value falls between 26 and 28. The values in this distribution fall between 20 and 54, but virtually all are between 20 and 34. Again, the District of Columbia is an outlier.

1.10: A stemplot for the state percentages of residents aged 65 years and over:

7	0
8	8
9	
10	0139
11	378888
12	1114456689999
13	012222334556688
14	0011367
15	035
16	9

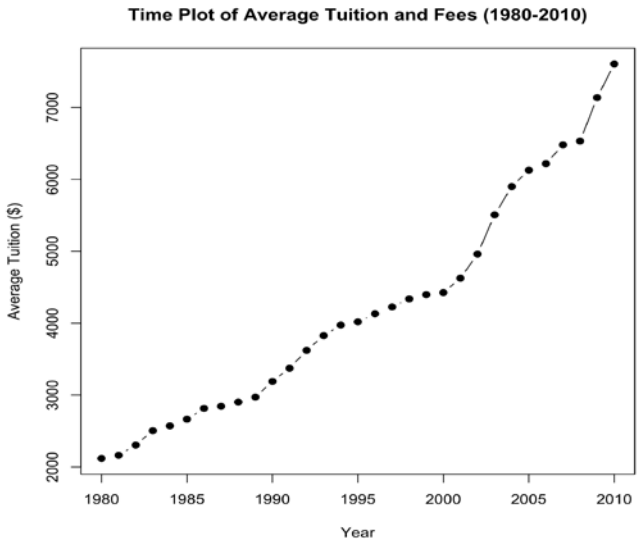
The midpoint is 13.0%. The spread is 7.0% to 16.9%.

1.11: Here is a stemplot for health expenditure per capita (PPP). Data are rounded to units of hundreds. For example, Argentina's "1332" becomes 13. Stems are thousands, and are split, as prescribed.

0 1 1 2 3
0 7 7 7 8 8 8 8 8
1 0 3
1 7
2 3
2 7 7 7 7 8
3 0 3 3 4 4
3 5 5 6 7 8 9
4 4
4 8
5
5
6
6
7 3

This distribution is somewhat right-skewed, with a single high outlier (United States). There are two clusters of countries. The center of this distribution is around 25 ($2500 spent per capita), ignoring the outlier. The distribution's spread is from 1 ($100 spent per capita) to 73 ($7300 spent per capita).

1.12: (a) A time plot of Average tuition follows:



(b) Tuition has steadily climbed during the 30-year period, with sharpest absolute increases in the last 10 years. (c) It would be better to use percent increases, rather than dollar increases. A 10% increase in tuition in 1980 should correspond to a 10% increase in tuition in 2005, but the absolute dollar increases in these cases are very different.

1.13: (a) the students.

1.14: (c) either a pie chart or a bar graph.

1.15: (b) Square footage and average monthly gas bill are both quantitative variables.

1.16: (b) Zip code is a categorical variable. Zip codes are equivalent to town (or zone) names or identifications, and you can't do arithmetic meaningfully with them.

1.17: (b) 20% to 22%.

1.18: (a) 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

1.19: (c) 30.9 minutes.

1.20: (b) roughly symmetric.

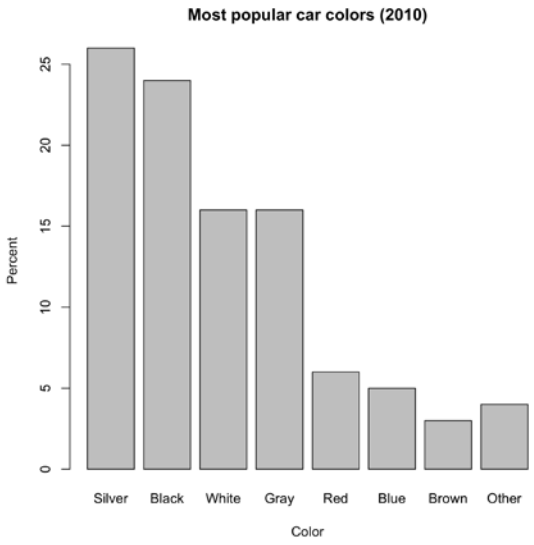
1.21: (b) close to 23.4 minutes. Take the 26th ordered value.

1.22: (c) skewed to the right.

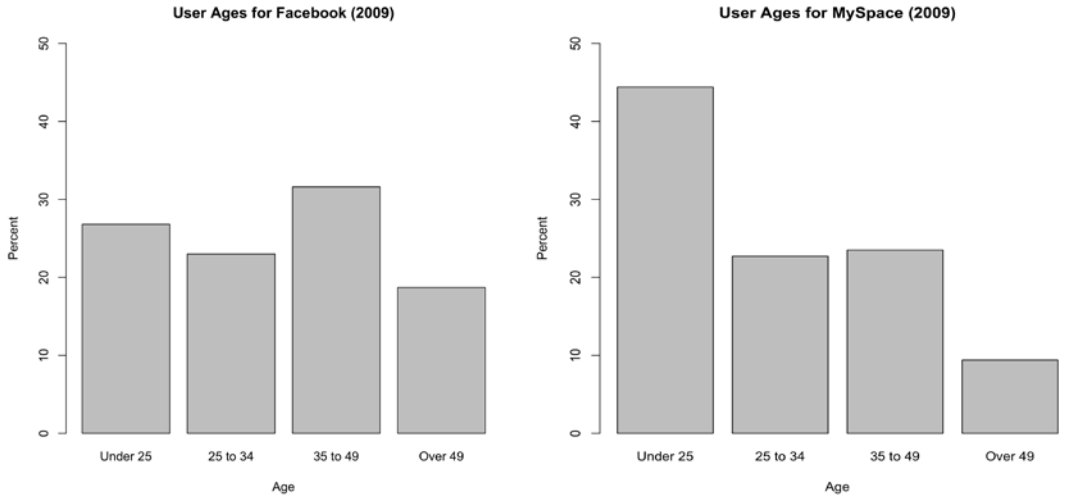
1.23: (a) Individuals are students who have finished medical school. (b) 6, including "Name." "Age" and "USMLE" are quantitative. The others are categorical.

1.24: The categorical variables are (a) Type of wood, (b) Type of water repellent, (d) Paint color. The quantitative variables are (c) Paint thickness and (e) Weathering time.

1.25: "Other colors" should account for 4%. A bar graph would be an appropriate display:

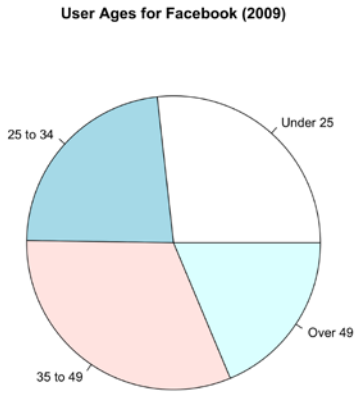


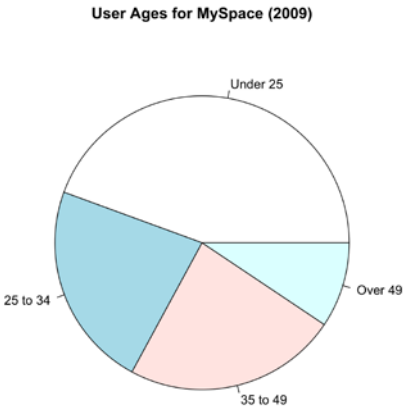
1.26: (a) Bar graphs for the age distribution of Facebook and MySpace users follow.



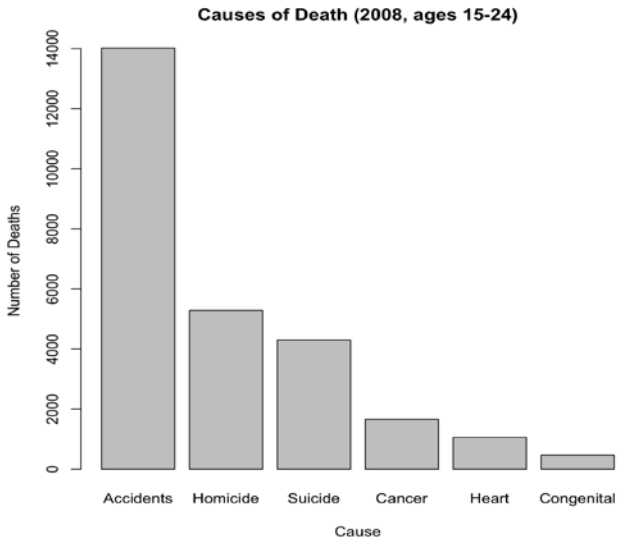
(b) Comparing these distributions, notice that Facebook users are almost evenly distributed across the age categories, while MySpace users are more heavily concentrated among the youngest demographic. Keeping the age groups arranged in order is important here.

(c) Pie charts follow, and are appropriate since the percentages in each distribution sum to 100%, though note that the percentages for Facebook categories sum to 100.1% due to rounding. Many feel that it is easier to compare distributions using bar graphs. Bar graphs invite comparison by heights, while pie charts require comparing areas. Comparing areas is difficult for many people.



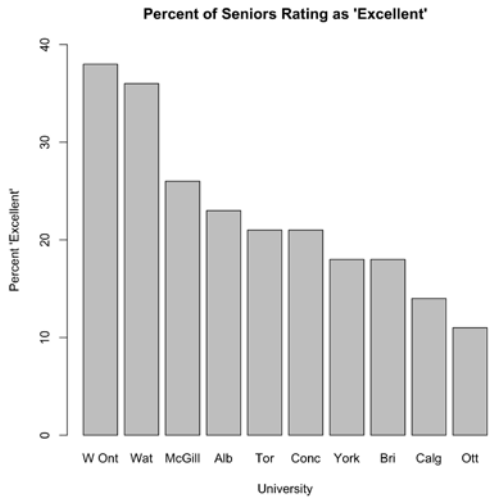


1.27: (a) A bar graph follows. (b) To make a pie chart, you would need to know the total number of deaths in this age group, or (equivalently) the number of deaths due to "other" causes.



1.28: Perhaps 60-65% of the Hispanic population in the United States is Mexican, while perhaps 10% is Puerto Rican.

1.29: (a) A bar graph is provided below. (b) A pie chat would be inappropriate, because these percentages aren't "shares." That is, the percentages don't sum to 100%.



1.30: This distribution is right-skewed, with center around 2 servings, and spread from 0 to 8 servings. There are no outliers. About 12% (9 out of 74) consumed 6 or more servings, and about 35% (26 out of 74) ate fewer than 2 servings (which means 0 or 1 serving).

1.31: (a) Ignoring the four lower outliers, the distribution is roughly symmetric, centered at a score of about 110, and having spread in scores of 86 to 136. (b) 64 of the 78 scores are more than 100. This is 82.1%.

1.32: (a) The distribution is slightly left-skewed. (b) The center is somewhere between 0% and 2.5%. (c) The smallest value is somewhere between –10% and –12.5%, and the largest value is between 12.5% and 15%. (d) There are about 130 negative returns, although your estimate could differ. This corresponds to about 42%.

1.33:

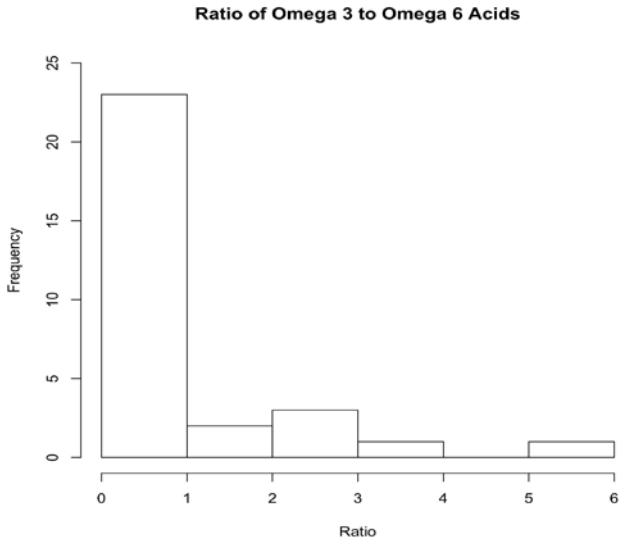
1. Are you male or female → Histogram (c). There are two outcomes possible, and the difference in frequencies is likely to be smaller than the right-handed/left-handed difference in (2).

2. Are you right-handed or left-handed → Histogram (b), since there are more right-handed people than left handed people, and the difference is likely larger than the sex difference in (1).

3. Heights → Histogram (d). Height distribution is likely to be symmetric.

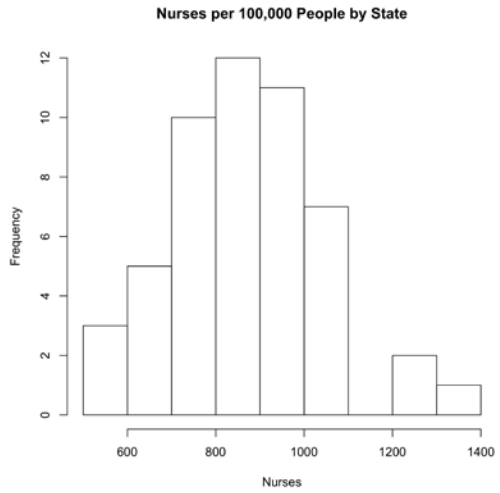
4. Time spent studying → Histogram (a). The variable takes on more than one value, and time spent studying may well be a right-skewed distribution, with most students spending less time studying, and some students spending more time studying.

1.34: (a) A histogram is provided below. (b) This is an extremely right-skewed distribution. Ratios greater than 1 correspond to an acid with more omega-3 than omega-6. Hence, this would be 7 of the 30 acids, or 23.3%. Most foods' oils aren't this healthy. (c) Of the 7 healthier foods, 5 are types of fish. Furthermore, all of the fish in the list have ratios higher than 1. Clearly, fish provide a healthier ratio of omega-3 to omega-6 acids.



1.35:

(a) States vary in population, so you would expect more nurses in California than in New Hampshire, for example. Nurses per 100,000 provides a better measure of how many nurses are available to serve a state's population. (b) A histogram is provided below. The District of Columbia, South Dakota, and Massachusetts are the three states different from the others. Perhaps they could be considered outliers. It's difficult to know why these states would have more nurses than other states.



1.36: (a) Since the countries have varying populations, it is easier to compare them by emissions per person than by total emissions. (b) Round data to the nearest tenth place, and use whole numbers as stems. The United States and Canada are the extreme outliers in this data set – a larger economy tends to coincide with more fuel consumption. The distribution is right-skewed, with center near 4 and a range of 0 to about 19.

0	001113333689
1	344589
2	3
3	
4	011479
5	
6	089
7	7
8	2378
9	68
10	58
11	
12	
13	
14	
15	
16	9
17	
18	9

1.37: Here is a stemplot for the pups data, using split stems at the tens place. This is a right-skewed distribution, with center around 25 pups and spread of 17 pups to 56 pups. There were several extremely good years for pups, resulting in more than 45 births.

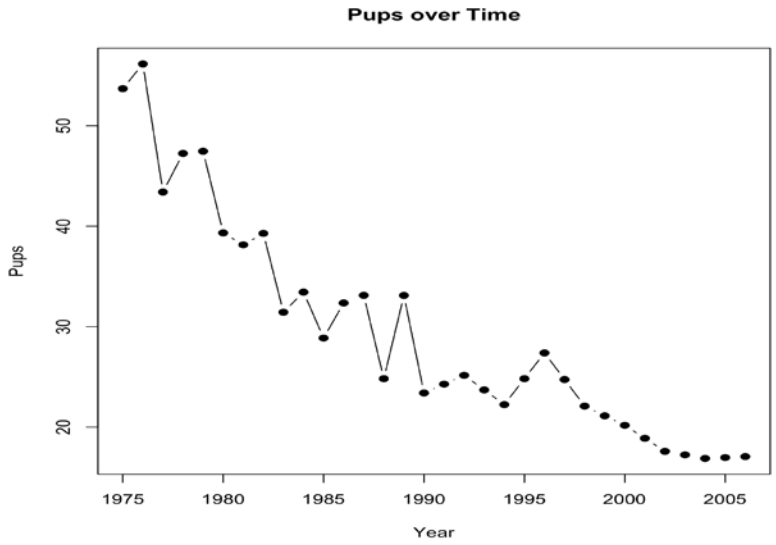
1 777789
2 0122344
2 555579
3 12333
3 899
4 3
4 77
5 4
5 6

1.38: (a) It is natural for people to round to common multiples like 5 and 10. In fact, multiples of 5 are the other common occurrence.

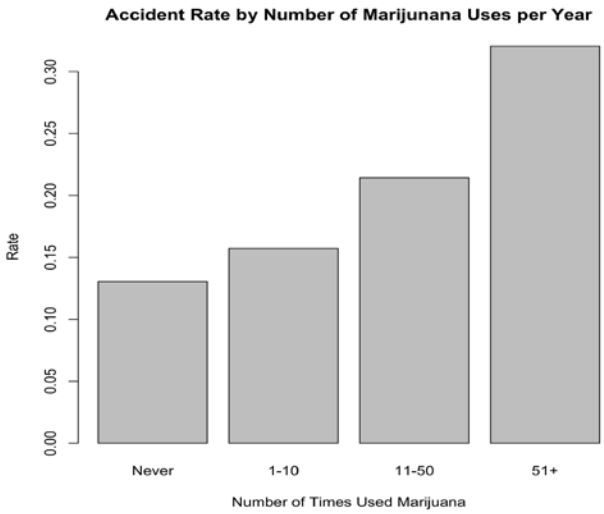
Women		Men
	0	033334
986	0	66679999
22222221	1	222222
88888888755555	1	5558
4440	2	44300
7	2	
	3	0
6	3	

Both distributions are somewhat right-skewed. The women's distribution is shifted a bit to the right of the men, suggesting that they tend to study a bit more. One woman claimed to study 360 minutes (6 hours) per night.

1.39: A time plot of seal pups. The decline in population is not described by the stemplot made in Exercise 1.37.



1.40: Rates are appropriate (rather than number of accidents) because the group sizes are different. If marijuana did not increase with the rate of accidents, then you would still have more accidents (by count) in the largest groups.



Accident rates go up with the number of times drivers use marijuana. The data are "observational," and no cause-and-effect conclusion is reasonable.

1.41: Coins with earlier (lower) dates are older, and rarer. Hence, there are more coins with larger dates (newer coins) than with smaller dates (older coins).

1.42: (a) Here are stem and split-stem plots. In both cases, stems denote the hundreds place. The slight right skew in the distribution is more apparent in the original version, but can be seen in both versions of the stem plot.

6 0 3 5 5 7
7 0 1 2 4 4 8 8 9 9 9
8 1 1 3 6 6 7
9 0 6

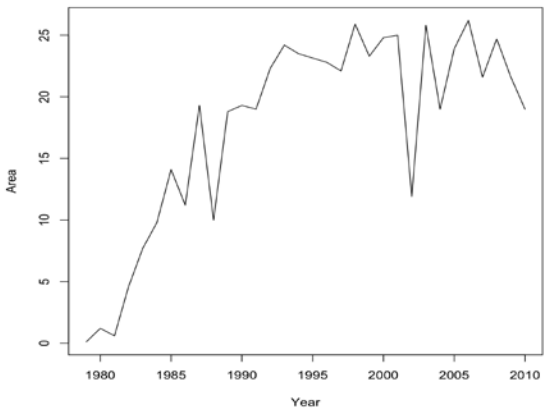
6 0 3
6 5 5 7
7 0 1 2 4 4
7 8 8 9 9 9
8 1 1 3
8 6 6 7
9 0
9 6

(b) The center of this distribution is around 780 millimeters. The distribution is somewhat right-skewed. There are no outliers. (c) It seems that El Niño strength is associated with volume of monsoon rains. No cause-and-effect relationship can be established, however, since there may be another factor driving both strong El Niño weather patterns and reduced monsoon rain volume.

1.43: (a) Graph (a) appears to show the greatest increase, even though both plots describe the same data. Vertical scaling can impact one's perception of the data. (b.) In both graphs, tuition starts around $2000 and rises to $7700. Again, both plots describe the same data.

1.44: (a) It seems as though winter quarters are typically associated with lower housing starts. (b) and (c) Over the long run, housing starts have risen, except for the most recent years, which correspond with the 2007–10 economic "crisis."

1.45: (a) A time plot of ozone hole size (area) is provided below. There is a trend, as well as year-to-year variability. The hole has grown a lot over the period studied, but may have leveled out in recent years.



(b) A stemplot of ozone hole size (area) is provided below. The midpoint is 19.3 millions of km^2. A stemplot fails to capture the relationship between size of hole and year.

0	0004
0	79
1	0114
1	899999
2	11222333444
2	5556

1.46: Use the Applet to investigate this problem.

1.47 and 1.48 are Web-based exercises.