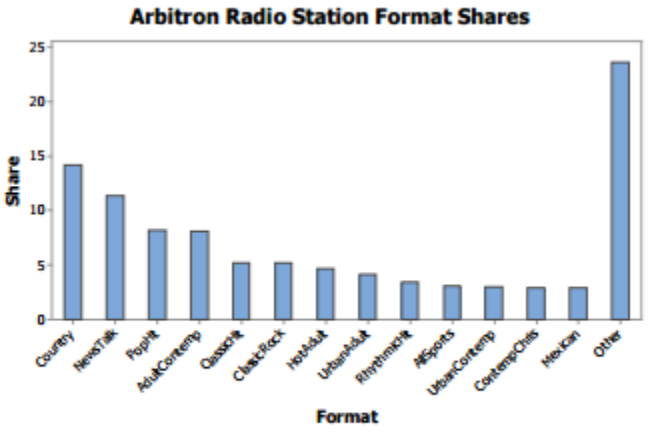
Chapter 1 – Picturing Distributions with Graphs

1.1 (a) The individuals are the car makes and models. **(b)** For each individual, the variables recorded are Vehicle class (categorical), Transmission type (categorical), Number of cylinders (usually treated as quantitative), City mpg (quantitative), Highway mpg (quantitative), and Annual fuel cost (dollars, quantitative).

1.3 (a) The given shares sum to 76.4%; 100% – 76.4% = 23.6% of the radio audience listens to stations with other formats. **(b)** The bar graph is shown at right. **(c)** A pie chart based only on the data presented would be inappropriate, because the areas of the pie wedges would be relative to the total of the categories presented (76.4%). If you include a wedge for "Other format" that accounts for 23.6% of the total, a pie chart would be reasonable.



1.11 Here is a stemplot for health expenditure per capita (in PPP). Data are rounded to units of hundreds. For example, Argentina's "1434" becomes 14. Stems are thousands and are split, as prescribed. This distribution is right-skewed, with a single high outlier (United States). There seem to be two clusters of countries. The center of this distribution is around 20 ($2000 spent per capita), ignoring the outlier. The distribution varies from 0|1 (about $100 spent per capita) to 8|6 (about $8600 spent per capita).

```
0 | 1144
0 | 6679999
1 | 02344
1 | 7
2 | 2
2 | 9
3 | 0123
3 | 79
4 | 114
4 | 556
5 | 1
5 | 67
6 |
6 |
7 |
7 |
8 |
8 | 6
```

1.13 (a) the students

1.14 (c) Either a pie chart or a bar graph would be appropriate, because you would have data for all the students (a complete "whole").

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) 58–61%

1.18 (b) 2, 3, 4, 5, 6, 7, 8, 9

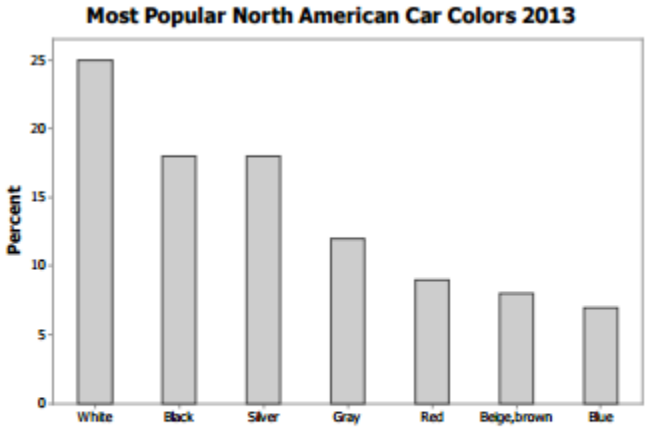
1.19 (b) 80%. There are 50 observations, so the center would be between the 25th and 26th observations; both of these are 80%.

1.20 (a)

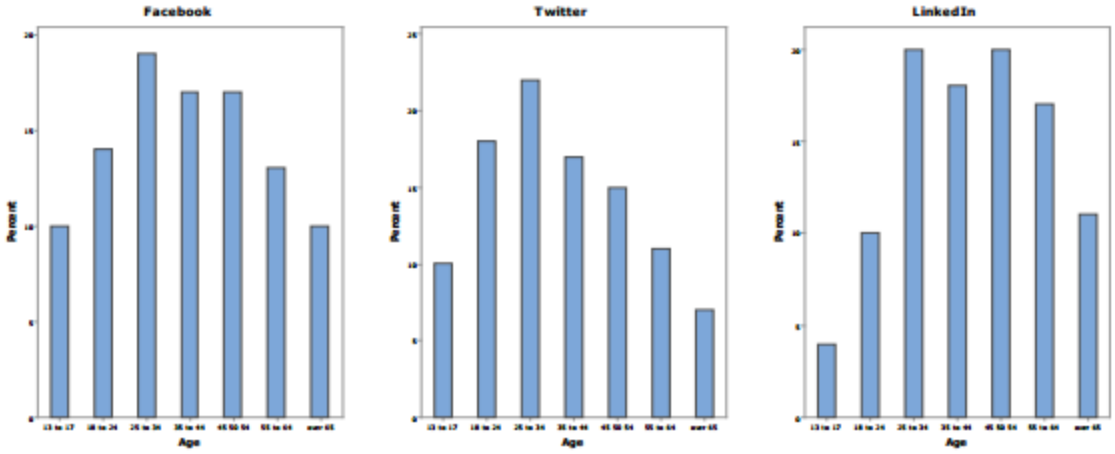
1.21 (b) 92%. The stems are rounded to whole percentages; you cannot make finer judgments.

1.22 (c) skewed to the right

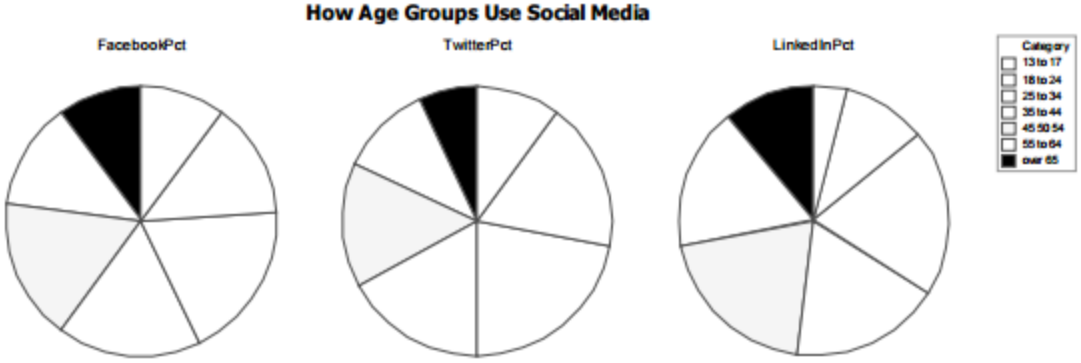
1.25 "Other colors" should *acco*unt for 3%. A bar graph would be an appropriate display. If you included the "other" category, a pie chart could be made.



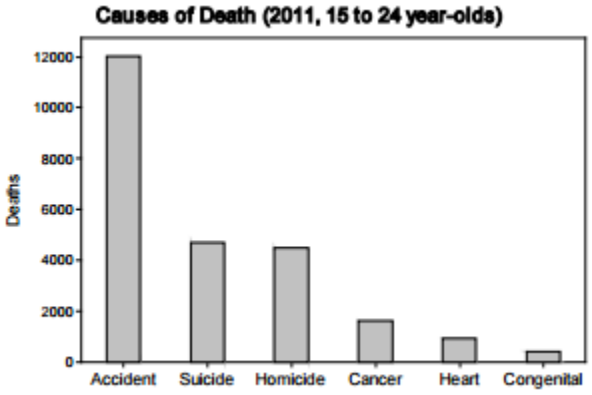
1.26 (a) Bar graphs for the age distributions are given below.



(b) Twitter and Facebook are more popular than LinkedIn (oriented toward careers) among younger users, but are least popular among older age groups. To see the similarities and differences, it is important to order the bars by age. **(c)** Pie charts follow and are appropriate, because the percentages in each distribution should (and do) sum to 100%. Many people feel that it is easier to compare distributions using bar graphs. Bar graphs invite comparison by heights, whereas pie charts require comparing areas. Comparing areas is difficult for many people.



1.27 (a) A bar graph is given. **(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.



1.33 (1.) "Are you male or female?" is Histogram (c). Two outcomes are 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?" is Histogram (b), because there are more right-handed people than left handed people, and the difference is likely larger than the sex difference in (1). **(3.)** "Heights" is Histogram (d). Height distribution is likely to be symmetric. **(4.)** "Time spent studying" is 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 a great deal of time studying.

1.37 The shape of the distribution is roughly symmetric (it might be called left-skewed if we ignore the high outlier); with this scaling, 245 seems to be a high outlier. The center is about 171 (the 12th observation). The data range is from about 94 to about 245.

9	46
10	29
11	
12	2
13	
14	5
15	8
16	57
17	011399
18	22
19	2
20	1233
21	
22	
23	
24	5

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