Chapter 2
Organizing and Summarizing Data

Section 2.1

1. Raw data are the data as originally collected, before they have been organized or coded.

2. Number (or count); proportion (or percent)

3. The relative frequencies should add to 1, although rounding may cause the answers to vary slightly.

4. A bar graph is used to illustrate qualitative data. It is a chart in which rectangles are used to illustrate the frequency or relative frequency with which a category appears. A Pareto chart is a bar chart with bars drawn in order of decreasing frequency or relative frequency.

5. **(a)** The largest segment in the pie chart is for "Washing your hands" so the most commonly used approach to beat the flu bug is washing your hands. 61% of respondents selected this as their primary method for beating the flu.

(b) The smallest segment in the pie chart is for "Drinking Orange Juice" so the least used method is drinking orange juice. 2% of respondents selected this as their primary method for beating the flu.

(c) 25% of respondents felt that flu shots were the best way to beat the flu.

6. **(a)** $\frac{128,000}{1,350,000} \approx 0.0948$; approximately 9.48% of cosmetic surgeries in 2009 were for tummy tucks.

(b) $\frac{138,000}{1,350,000} \approx 0.102$; approximately 10.2% of cosmetic surgeries in 2009 were for nose reshaping.

(c) The graph accounts for 312,000 + 284,000 + 150,000 + 138,000 + 128,000 = 1,012,000 surgeries. Thus, 1,350,000 − 1,012,000 = 338,000 surgeries are not accounted for in the graph.

7. **(a)** The highest bar corresponds to the position OF (outfield), so OF is the position with the most MVPs.

(b) The bar for first base (1B) reaches the line for 15. Thus, there were 15 MVPs who played first base.

(c) The bar for outfield (OF) is 30 on the vertical axis. The bar for first base (1B) reaches 15. Since 30 – 15=15, there were 15 more MVPs who played outfield than first base.

(d) Each of the three outfield positions should be reported as MVPs, rather than treating the three positions as one position.

8. **(a)** 29,936,000 whites were living in poverty.

(b) 12745/(29936 + 11041 + 12745 + 1974) = 0.229 = 22.9%
In 2013, about 22.9% of the impoverished in the United States were Hispanic.

(c) This graph should use relative frequencies, rather than frequencies. The graph does not account for the different population size of each ethnic group. Without knowing the population sizes, we cannot determine whether a group is disproportionally impoverished.

9. **(a)** 69% of the respondents believe divorce is morally acceptable.

(b) 23% believe divorce is morally wrong. So, 240 million * 0.23 = 55.2 million adult Americans believe divorce is morally wrong.

(c) This statement is inferential, since it is a generalization based on the observed data.

10. **(a)** 5% of identity theft was loan fraud.

(b) 26% of the identity fraud cases in a recent year involved credit card fraud. So, 10 million * 0.26 = 2.6 million cases of credit card fraud occurred in a recent year.

11. **(a)** The proportion of 18–34 year old respondents who are more likely to buy when made in America is 0.42. For 34–44 year olds, the proportion is 0.61.

(b) The 55+ age group has the greatest proportion of respondents who are more likely to buy when made in America.

(c) The 18–34 age group has a majority of respondents who are less likely to buy when made in America.

(d) As age increases, so does the likelihood that a respondent will be more likely to buy a product that is made in America.

12. **(a)** The proportion of males who would like to be richer is 0.46. The proportion of females who would like to be richer is 0.41.

(b) The attribute that females desire more than males is to be thinner.

(c) The attribute that males prefer over females two-to-one is to be younger.

(d) Equal proportions of males and females desire to be smarter.

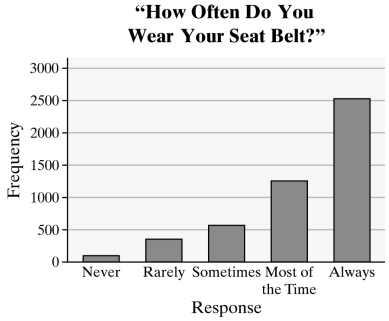
13. **(a)** Total students surveyed = 125 + 324 + 552 + 1257 + 2518 = 4776
Relative frequency of "Never"
= 125 / 4776 ≈ 0.0262, and so on.

Response	Relative Frequency
Never	0.0262
Rarely	0.0678
Sometimes	0.1156
Most of the time	0.2632
Always	0.5272

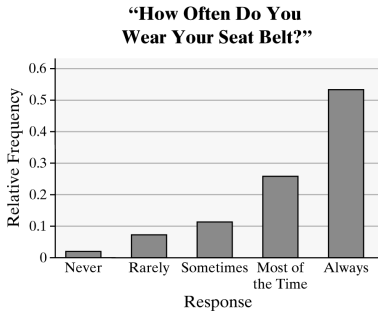
(b) 52.72%

(c) 0.0262 + 0.0678 = 0.0940 or 9.40%

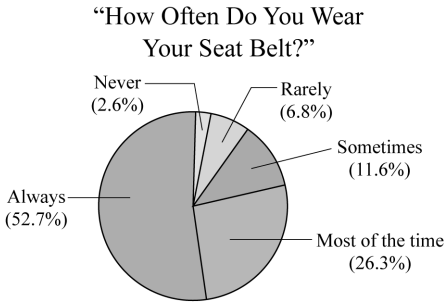
(d)



(e)



(f)



(g) The statement is inferential since it is inferring something about the entire population based on the results of a sample survey.

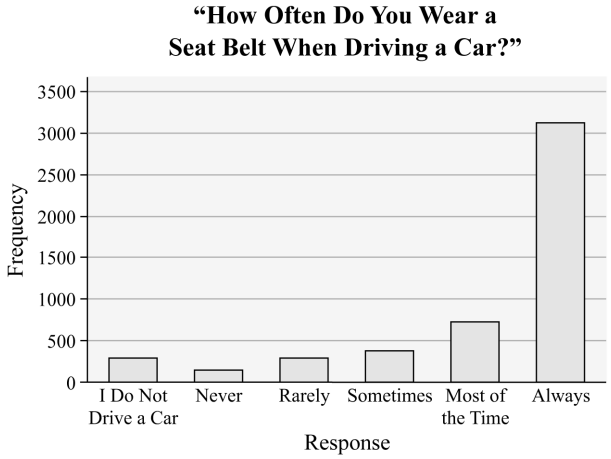
14. **(a)** Total students surveyed = 249 + 118 + 249 + 345 + 716 + 3093 = 4770
Relative frequency of " I do not drive"
$= \frac{249}{4770} \approx 0.0522$, and so on.

Response	Relative Frequency
I do not drive	0.0522
Never	0.0247
Rarely	0.0522
Sometimes	0.0723
Most of the time	0.1501
Always	0.6484

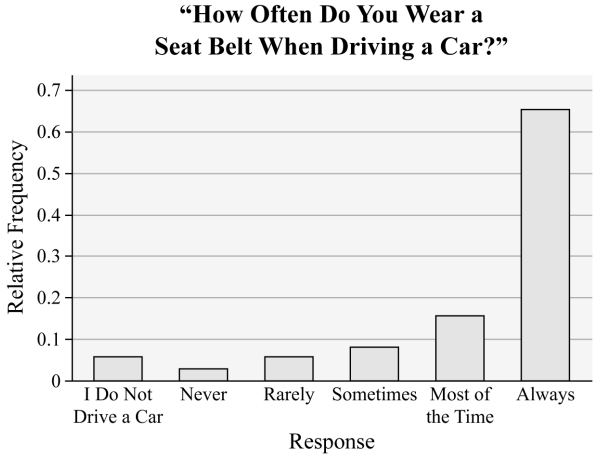
(b) 64.84%

(c) 0.0247 + 0.0522 = 0.0769 or 7.7%

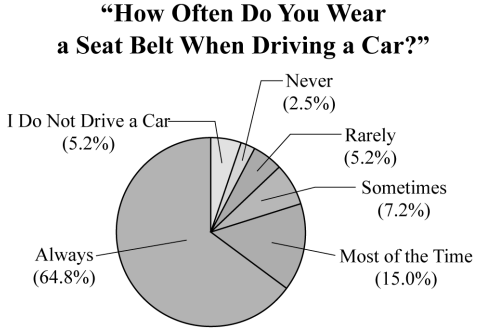
(d)



(e)



(f)



(g) Total students = 118 + 249 + 345 + 716 + 3093 = 4521

Relative frequency of "Never"

$$= \frac{118}{4521} \approx 0.0261, \text{ and so on.}$$

Response	Relative Frequency
Never	0.0261
Rarely	0.0551
Sometimes	0.0763
Most of the time	0.1584
Always	0.6841

The relative frequencies of all categories are very similar except that students are more likely to wear their seatbelt 'Always' when driving.

(h) The statement is descriptive because it is describing the particular sample.

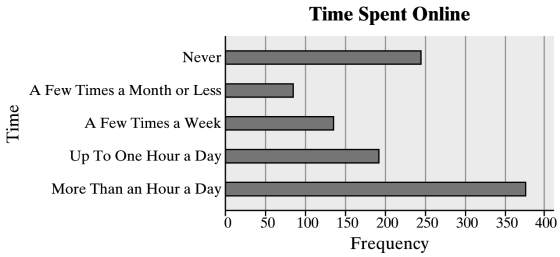
15. (a) Total adults surveyed = 377 + 192 + 132 + 81 + 243 = 1025

Relative frequency of "More than 1 hour a day" = $377/1025 \approx 0.3678$, and so on.

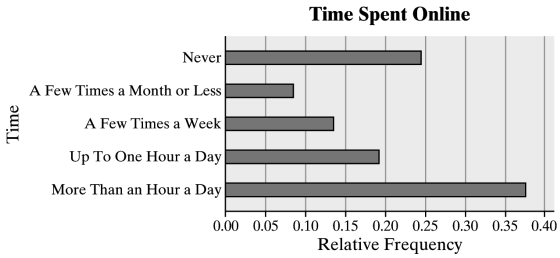
Response	Relative Frequency
More than 1 hr a day	0.3678
Up to 1 hr a day	0.1873
A few times a week	0.1288
A few times a month or less	0.0790
Never	0.2371

(b) 0.2371 (about 24%)

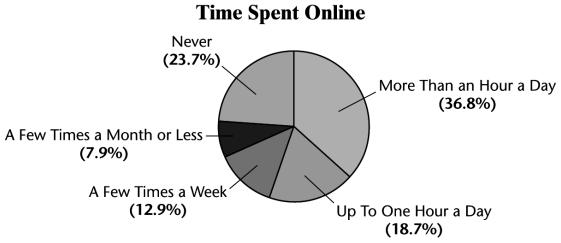
(c)



(d)



(e)



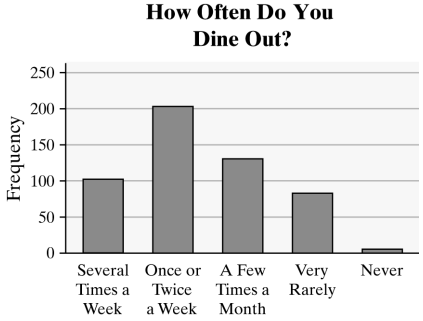
(f) The statement provides an estimate, but no level of confidence is given.

16. (a) Total adults surveyed = 103 + 204 + 130 + 79 + 5 = 521
Relative frequency of "Several times a week" = $\frac{103}{521} \approx 0.197$, and so on.

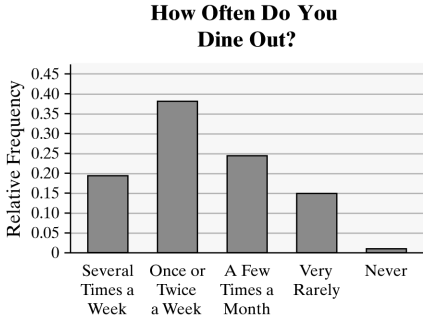
Response	Relative Frequency
Several times a week	0.197
Once or twice a week	0.392
A few times a month	0.250
Vary rarely	0.152
Never	0.010

(b) The proportion surveyed who dine out once or twice a week is 204/(103 + 204 + 130 + 79 + 5) = 0.392

(c)



(d)



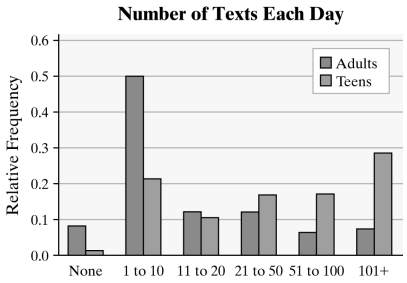
17. (a) Total adults = 1936
Relative frequency for "none" is: 173/1936 = 0.09, and so on.

Number of Texts	Rel. Freq. (Adults)
None	0.089
1 to 10	0.505
11 to 20	0.129
21 to 50	0.129
51 to 100	0.069
101+	0.079

(b) Total teens = 627
Relative frequency for "none" is: 13/627 = 0.021, and so on.

Number of Texts	Rel. Freq. (Teens)
None	0.021
1 to 10	0.220
11 to 20	0.110
21 to 50	0.180
51 to 100	0.180
101+	0.289

(c)



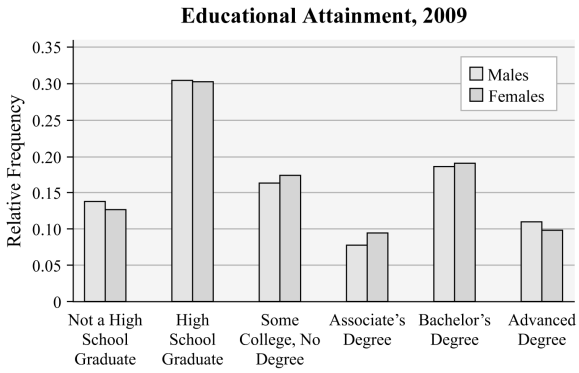
(d) Answers will vary. Adults are much more likely to send fewer texts per day, while teens are much more likely to do more texting.

18. (a), (b)
Total males = 99.4 million
Relative frequency for "Not HS graduate" is 12.3/99.4 = 0.124, and so on.

Total females = 107.6 million
Relative frequency for "Not HS graduate" is 12.2/107.6 = 0.113, and so on.

Educational Attainment	Males	Females
Not a HS graduate	0.124	0.113
High school graduate	0.302	0.295
Some college, no degree	0.166	0.170
Associate's degree	0.089	0.108
Bachelor's degree	0.200	0.202
Advanced degree	0.120	0.112

(c)



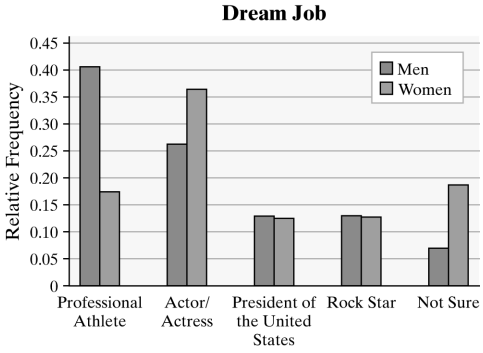
(d) Answers will vary. It appears that females are slightly more likely to start, but not finish college. Males appear to be slightly more likely to attain an advanced degree.

19. (a) Total males = 99; Relative frequency for "Professional Athlete" is 40/99 = 0.404, and so on.

Total number of females = 100; Relative frequency for "Professional Athlete" is 18/100 = 0.18, and so on.

Dream Job	Men	Women
Professional Athlete	0.404	0.180
Actor/Actress	0.263	0.370
President of the United States	0.131	0.130
Rock Star	0.131	0.130
Not Sure	0.071	0.190

(b)

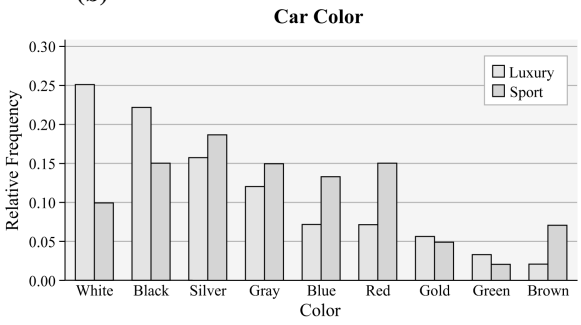


(c) Answers will vary. Males are much more likely to want to be a professional athlete. Women are more likely to aspire to a career in acting than men. Men's desire to become athletes may be influenced by the prominence of male sporting figures in popular culture. Women may aspire to careers in acting due to the perceived glamour of famous female actresses.

20. (a) Relative frequency for "White" luxury cars = $\frac{25}{100} = 0.25,$ and so on.
Relative frequency for "White" sport cars = $\frac{10}{100} = 0.10,$ and so on.

	Relative Frequencies	
Color	Luxury Cars	Sport Cars
White	0.25	0.10
Black	0.22	0.15
Silver	0.16	0.18
Gray	0.12	0.15
Blue	0.07	0.13
Red	0.07	0.15
Gold	0.06	0.05
Green	0.03	0.02
Brown	0.02	0.07

(b)

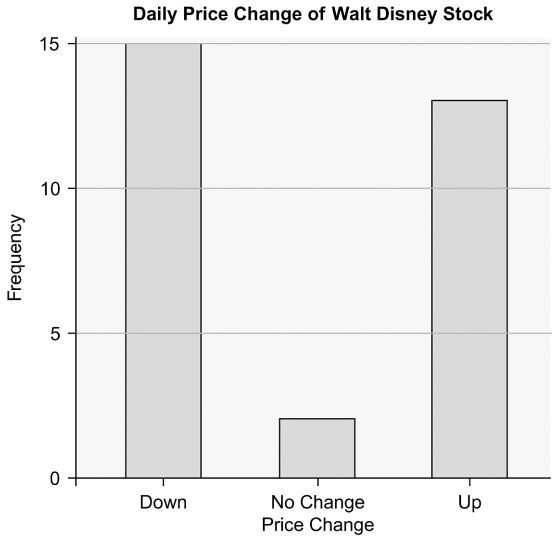


(c) Answers will vary. White is the most popular color for luxury cars, while silver is the most popular for sports cars. People who drive luxury cars may enjoy the clean look of a white vehicle. People who drive sports cars may prefer the flashier look of silver.

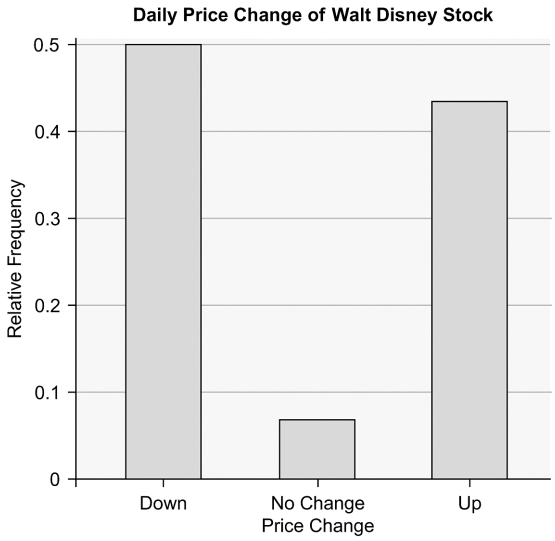
21. **(a), (b)**
Total number of Trading Days = 30; relative frequency for Down is 15/30 = 0.5, and so on.

Winner	Freq.	Rel. Freq.
Down	15	0.500
No Change	2	0.067
Up	13	0.433

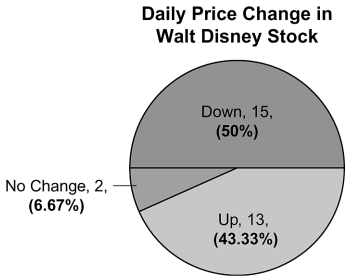
(c)



(d)



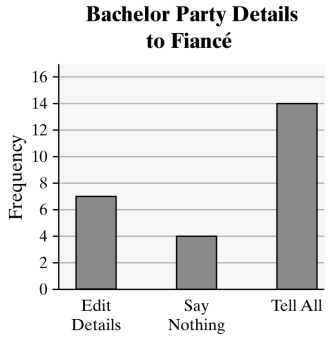
(e)



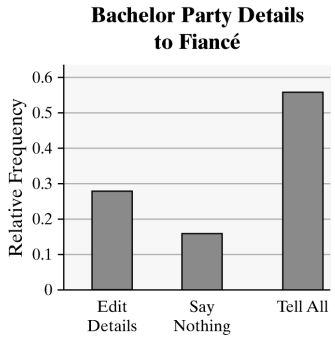
22. **(a), (b)** Total number of responses = 25; relative frequency for "edit details" is 7/25 = 0.28.

Response	Freq.	Rel. Freq.
Edit details	7	0.28
Say nothing	4	0.16
Tell all	14	0.56

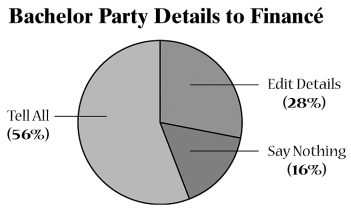
(c)



(d)



(e)

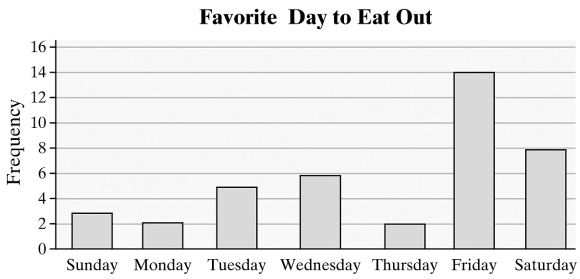


23. **(a), (b)**
Total number of responses = 40; relative frequency for "Sunday" is 3/40 = 0.075.

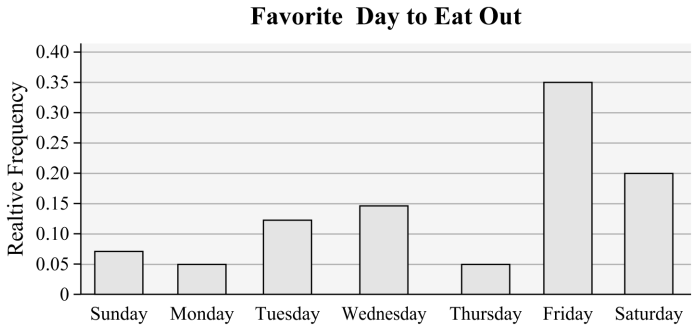
Response	Freq.	Rel. Freq.
Sunday	3	0.075
Monday	2	0.050
Tuesday	5	0.125
Wednesday	6	0.150
Thursday	2	0.050
Friday	14	0.350
Saturday	8	0.200

(c) Answers will vary. If you own a restaurant, you will probably want to advertize on the days when people will be most likely to order takeout: Friday. You might consider avoiding placing an ad on Monday and Thursday, since the readers are least likely to choose to order takeout on these days.

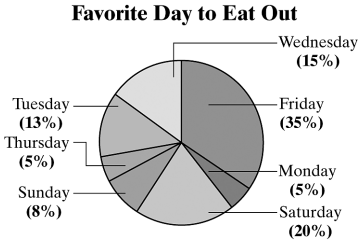
(d)



(e)



(f)



24. **(a), (b)**
Total number of patients = 50
Relative frequency for "Type A"
$= \frac{18}{50} = 0.36$, and so on.

Blood Type	Freq.	Rel. Freq.
A	18	0.36
AB	4	0.08
B	6	0.12
O	22	0.44

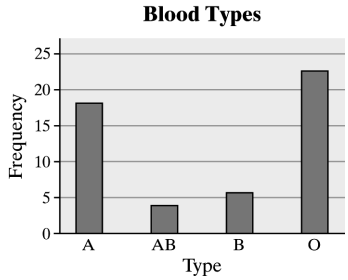
(c) Type O is the most common.

(d) Type AB is the least common.

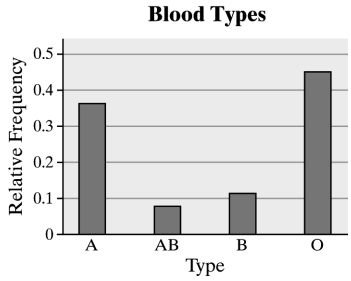
(e) We estimate that 44% of the population has type O blood. This is considered inferential statistics because a conclusion about the population is being drawn based on sample data.

(f) Answers will vary; in 2008 the Red Cross reported that 45% of the population had type O blood (either + or –). Results will differ because of sampling variability.

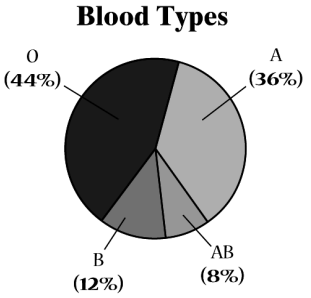
(g)



(h)



(i)



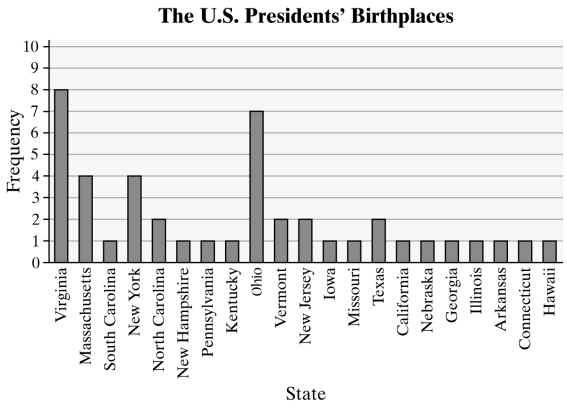
25. (a)

State	AR	CA	CT	GA	HI	IL
Freq.	1	1	1	1	1	1

State	IA	KY	MA	MO	NE
Freq.	1	1	4	1	1

State	NH	NJ	NY	NC	OH
Freq.	1	2	4	2	7

State	PA	SC	TX	VT	VA
Freq.	1	1	2	2	8

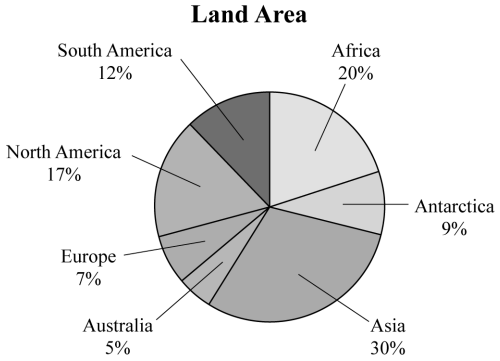


(b) More presidents were born in Virginia than in any other state.

(c) Answers will vary. The data do not take the year of statehood into account. For example, Virginia has been a state for roughly 62 years more than California. The population of the United States was more concentrated in the east in the early years, so it was more likely that the president would be from that part of the country.

26. (a) It would make sense to draw a pie chart for land area since the 7 continents contain all the land area on Earth.
Total land area is 11,608,000 + 5,100,000 + … + 9,449,000 + 6,879,000 = 57,217,000 square miles.
The relative frequency (percentage) for Africa is $\frac{11,608,000}{57,217,000} = 0.2029$.

Continent	Land Area (mi^2)	Rel. Freq.
Africa	11,608,000	0.2029
Antarctica	5,100,000	0.0891
Asia	17,212,000	0.3008
Australia	3,132,000	0.0547
Europe	3,837,000	0.0671
North America	9,449,000	0.1651
South America	6,879,000	0.1202



(b) It would not make sense to draw a pie chart for the highest elevation because there is no whole to which to compare the parts.

27. Answers will vary.

28. Answers will vary.

29. (a) The researcher wants to determine if online homework improves student learning over traditional pencil-and-paper homework.

(b) This study is an experiment because the researcher is actively imposing treatments (the homework style) on subjects.

(c) Answers will vary. Some examples are same teacher, same semester, and same course.

(d) Assigning different homework methods to entire classes could confound the results because there may be differences between the classes. The instructor may give more instruction to one class than the other. The instructor is not blinded, so he or she may treat one group differently from the other.

(e) *Number of students*: quantitative, discrete
Average age: quantitative, continuous
Average exam score: quantitative, continuous
Type of homework: qualitative
College experience: qualitative

(f) Letter grade is a qualitative variable at the ordinal level of measurement.
Answers will vary. It is possible that ordering the data from A to F is better because it might give more "weight" to the higher grade and the researcher wants to show that a higher percent of students passed using the online homework.

(g) The graph being displayed is a side-by-side relative frequency bar graph.

(h) Yes; the "whole" is the set of students who received a grade for the course for each homework method.

(i) The table shows that the two groups with no prior college experience had roughly the same average exam grade. From the bar graph, we see that the students using online homework had a lower percent for As, but had a higher percent who passed with a C or better.

30. Relative frequencies should be used when the size of two samples or populations differ.

31. Answers will vary. If the goal is to illustrate the levels of importance, then arranging the bars in a bar chart in decreasing order makes sense. Sometimes it is useful to arrange the categorical data in a bar chart in alphabetical order. A pie chart does not readily allow for arranging the data in order.

32. A bar chart is preferred when trying to compare two specific values. Pie charts are helpful for comparing parts of a whole. A pie chart cannot be drawn if the data do not include all possible values of the qualitative variable.

33. No, the percentages do not sum to 100%.

Section 2.2

1. classes

2. lower; upper

3. class width

4. Skewed left means that the left tail is longer than the right tail.

5. True

6. False

7. False. The distribution shape shown is skewed right.

8. False. The distribution shape is bell-shaped.

9. (a) The value with the highest frequency is 8.

(b) The value with the lowest frequency is 2.

(c) The value of 7 was observed 15 times.

(d) The value of 5 was observed 11 times and the value of 4 was observed 7 times. Therefore, the value of 5 was observed 4 more times than the value of 4 (e.g. $11 - 7 = 4$).

(e) $\frac{15}{100} = 0.15$ or 15% of the time a 7 was observed.

(f) The distribution is approximately bell-shaped.

10. (a) The most frequent number of cars sold in a week was 4 cars.

(b) There were 9 weeks in which 2 cars sold.

(c) Total frequency = 4 + 2 + 9 + 8 + 12 + 8 + 5 + 2 + 1 + 1 = 52 (as required)
Percentage of time two cars are sold
$= \frac{9}{52} \cdot 100 = 17.3\%$

(d) Slightly skewed to the right

11. **(a)** Total frequency = 2 + 3 + 13 + 42 + 58 + 40 + 31 + 8 +2 + 1 = 200

(b) 10 (e.g. 70 – 60 = 10)

(c)

IQ Score (class)	Frequency
60–69	2
70–79	3
80–89	13
90–99	42
100–109	58
110–119	40
120–129	31
130–139	8
140–149	2
150–159	1

(d) The class "100 – 109" has the highest frequency.

(e) The class "150 – 159" has the lowest frequency.

(f) $\frac{8+2+1}{200} = 0.055 = 5.5\%$

(g) No, there were no IQs above 159.

12. **(a)** The class width is 200 (e.g. 200 – 0 = 200).

(b) 0–199, 200–399, 400–599, 600–799, 800–999, 1000–1199, 1200–1399

(c) The highest frequency is in class 0–199.

(d) The distribution is skewed right.

(e) Answers will vary. The statement is incorrect because they are comparing counts from populations of different size. To make a fair comparison, the reporter should use rates of fatalities such as the number of fatalities per 1000 residents.

13. **(a)** Likely skewed right. Most household incomes will be to the left (perhaps in the $50,000 to $150,000 range), with fewer higher incomes to the right (in the millions).

(b) Likely bell-shaped. Most scores will occur near the middle range, with scores tapering off equally in both directions.

(c) Likely skewed right. Most households will have, say, 1 to 4 occupants, with fewer households having a higher number of occupants.

(d) Likely skewed left. Most Alzheimer's patients will fall in older-aged categories, with fewer patients being younger.

14. **(a)** Likely skewed right. More individuals would consume fewer alcoholic drinks per week, while less individuals would consume more alcoholic drinks per week.

(b) Likely uniform. There will be approximately an equal number of students in each age category.

(c) Likely skewed left. Most hearing-aid patients will fall in older-aged categories, with fewer patients being younger.

(d) Likely bell-shaped. Most heights will occur, say, in the 66- to 70-inch range, with heights tapering off equally in both directions.

15. **(a)** Total number of households = $16+18+12+3+1=50$

Relative frequency of 0 children = 16/50 = 0.32, and so on.

Number of Children Under Five	Relative Frequency
0	0.32
1	0.36
2	0.24
3	0.06
4	0.02

(b) $\frac{12}{50} = 0.24$ or 24% of households have two children under the age of 5.

(c) $\frac{18+12}{50} = \frac{30}{50} = 0.6$ or 60% of households have one or two children under the age of 5.

16. **(a)** Total number of free throws = $16+11+9+7+2+3+0+1+0+1=50$. Relative frequency of 1 throw until a miss = 16/50 = 0.32, and so on.

Number of Free Throws Until a Miss	Relative Frequency
1	0.32
2	0.22
3	0.18
4	0.14
5	0.04
6	0.06
7	0.00
8	0.02
9	0.00
10	0.02

(b) $\frac{7}{50}=0.14$; 14% of the time she first missed on the fourth try.

(c) $\frac{1}{50}=0.02$; 2% of the time she first missed on the tenth try.

(d) "At least 5" means that the basketball player misses on the 6^{th} shot or 7^{th} shot or 8^{th}, etc. $\frac{3+0+1+0+1}{50}=\frac{5}{50}=0.10$ or 10% of the time.

17. From the legend, 1|0 represents 10, so the original data set is 10, 11, 14, 21, 24, 24, 27, 29, 33, 35, 35, 35, 37, 37, 38, 40, 40, 41, 42, 46, 46, 48, 49, 49, 53, 53, 55, 58, 61, 62.

18. From the legend, 24|0 represents 240, so the original data set is 240, 244, 247, 252, 252, 253, 259, 259, 263, 264, 265, 268, 268, 269, 270, 271, 271, 273, 276, 276, 282, 283, 288.

19. From the legend, 1|2 represents 1.2, so the original data set is 1.2, 1.4, 1.6, 2.1, 2.4, 2.7, 2.7, 2.9, 3.3, 3.3, 3.3, 3.5, 3.7, 3.7, 3.8, 4.0, 4.1, 4.1, 4.3, 4.6, 4.6, 4.8, 4.8, 4.9, 5.3, 5.4, 5.5, 5.8, 6.2, 6.4.

20. From the legend, 12|3 represents 12.3, so the original data set is 12.3, 12.7, 12.9, 12.9, 13.0, 13.4, 13.5, 13.7, 13.8, 13.9, 13.9, 14.2, 14.4, 14.4, 14.7, 14.7, 14.8, 14.9, 15.1, 15.2, 15.2, 15.5, 15.6, 16.0, 16.3.

21. **(a)** There are six classes.

(b) Lower class limits: 10, 14, 18, 22, 26, 30
Upper class limits: 13.9, 17.9, 21.9, 25.9, 29.9, 33.9

(c) The class width can be found by subtracting consecutive lower class limits. For example, 14 – 10 = 4. Therefore, the class width is 4 (players).

22. **(a)** There are eight classes.

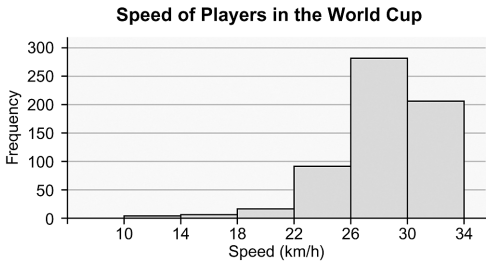
(b) Lower class limits: 0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0
Upper class limits: 0.9, 1.9, 2.9, 3.9, 4.9, 5.9, 6.9, 7.9

(c) The class width can be found by subtracting consecutive lower class limits. For example, $2.0-1.0=1.0$. Therefore, the class width is 1.0.

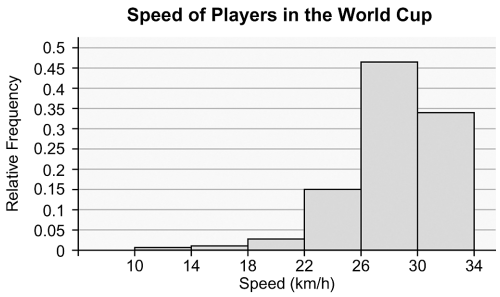
23. **(a)** Total frequency = 4 + 7 + 17 + 91+ 282 + 206 = 607
Relative frequency for 10–13.9 is 4/607 = 0.0066, and so on.

Speed (Km/hr)	Relative Frequency
10–13.9	0.0066
14–17.9	0.0115
18–21.9	0.0280
22–25.9	0.1499
26–29.9	0.4646
30–33.9	0.3394

(b)



(c)

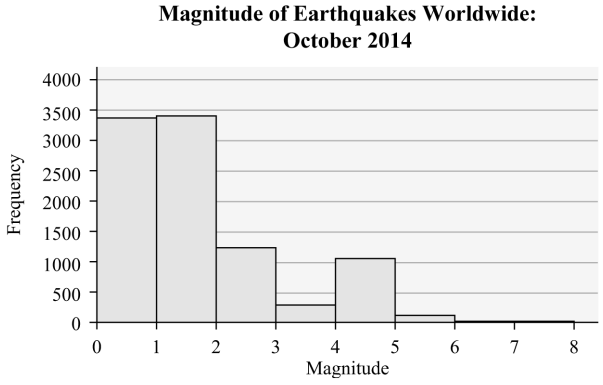


The percentage of players who had a top speed between 30 and 33.9 km/h is 33.94%. The percent of players who had a top speed less than 13.9 km/h is 0.66%.

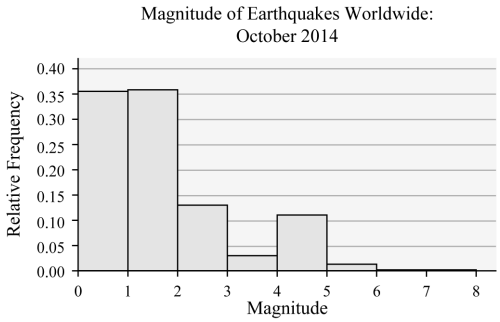
24. **(a)** Total frequency = 3371 + 3400 + 1237 + 286 + 1045 + 121 + 7 + 2 = 9469
Relative frequency for 0–0.9 is 3371/9469 = 0.3561, and so on.

Magnitude	Relative Frequency
0–0.9	0.3560
1.0–1.9	0.3591
2.0–2.9	0.1306
3.0–3.9	0.0302
4.0–4.9	0.1104
5.0–5.9	0.0128
6.0–6.9	0.0007
7.0–7.9	0.0002

(b)



(c)



The percentage of earthquakes that registered between 4.0 and 4.9 km/h is 11.04%. The percent of earthquakes that registered 4.9 or less is 98.63%.

25. **(a)** The data are discrete. The possible values for the number of color televisions in a household are countable.

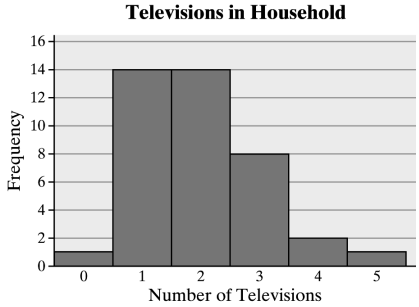
(b), (c) The relative frequency for 0 color televisions is 1/40 = 0.025, and so on.

Number of Color TVs	Frequency	Relative Frequency
0	1	0.025
1	14	0.350
2	14	0.350
3	8	0.200
4	2	0.050
5	1	0.025

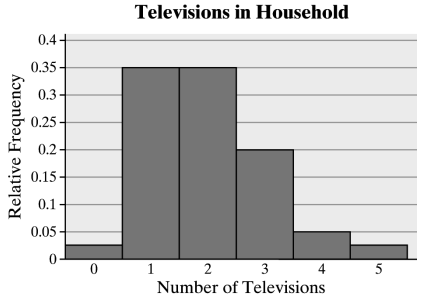
(d) The relative frequency is 0.2, so 20% of the households surveyed had 3 color televisions.

(e) 0.05 + 0.025 = 0.075
7.5% of the households in the survey had 4 or more color televisions.

(f)



(g)



(h) The distribution is skewed right.

26. **(a)** The data are discrete. The possible values for the number of customers waiting for a table are countable.

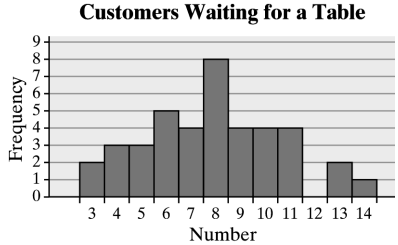
(b), (c) Relative frequency of 3 customers waiting = 2/40 = 0.05, and so on.

Number of Customers	Freq.	Rel. Freq.
3	2	0.050
4	3	0.075
5	3	0.075
6	5	0.125
7	4	0.100
8	8	0.200
9	4	0.100
10	4	0.100
11	4	0.100
12	0	0.000
13	2	0.050
14	1	0.025

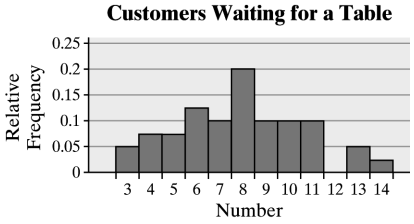
(d) $10.0 + 10.0 + 0.0 + 5.0 + 2.5 = 27.5\%$ of the Saturdays had 10 or more customers waiting for a table at 6 p.m.

(e) $5.0 + 7.5 + 7.5 = 20.0\%$ of the Saturdays had 5 or fewer customers waiting for a table at 6 p.m.

(f)



(g)

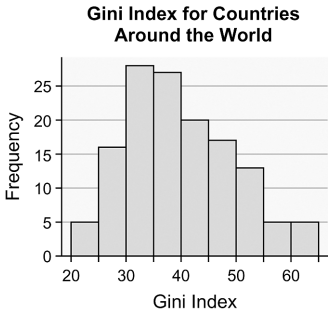


(h) The distribution is more or less symmetric.

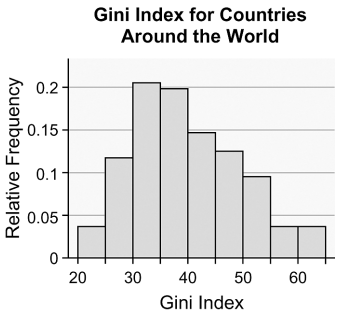
27. **(a), (b)** Relative frequency of a Gini Index of 20–24.9 = 5/136 = 0.037, and so on.

Gini Index	Freq.	Rel. Freq.
20–24.9	5	0.037
25–29.9	16	0.118
30–34.9	28	0.206
35–39.9	27	0.199
40–44.9	20	0.147
45–49.9	17	0.125
50–54.9	13	0.096
55–59.9	5	0.037
60–64.9	5	0.037

(c)



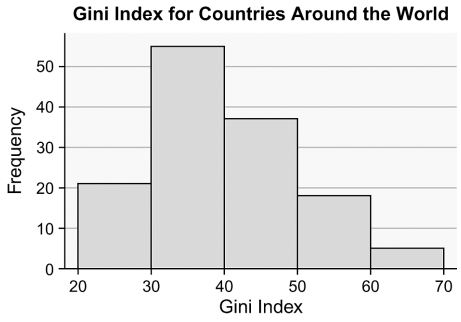
(d)

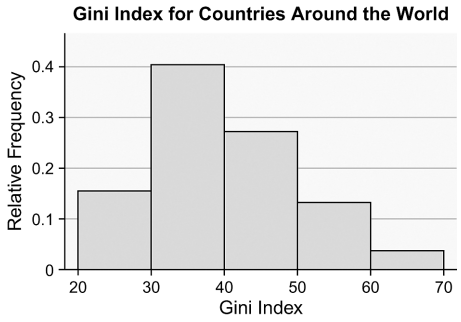


(e) The shape of the distribution is skewed right.

(f) Relative frequency of a Gini Index of 20–29.9 = 21/136 = 0.154, and so on.

Gini Index	Freq.	Rel. Freq.
20–29.9	21	0.154
30–39.9	55	0.404
40–49.9	37	0.272
50–59.9	18	0.132
60–69.9	5	0.037





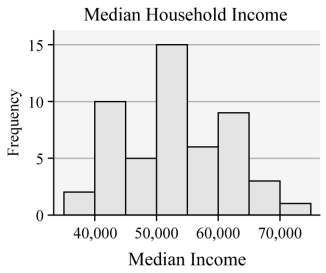
The shape of the distribution is skewed right.

(g) Answers will vary. The graph with a class width of 5 provides more detail, so it seems to be a superior graph.

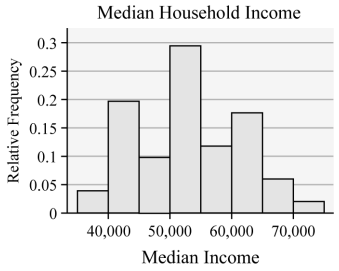
28. (a), (b) Relative frequency for the median income 35,000–39,999 is 2/51 = 0.0392, and so on.

Income	Freq.	Rel. Freq.
35,000–39,999	2	0.0392
40,000–44,999	10	0.1961
45,000–49,999	5	0.0980
50,000–54,999	15	0.2941
55,000–59,999	6	0.1176
60,000–64,999	9	0.1765
65,000–69,999	3	0.0588
70,000–74,999	1	0.0196

(c)



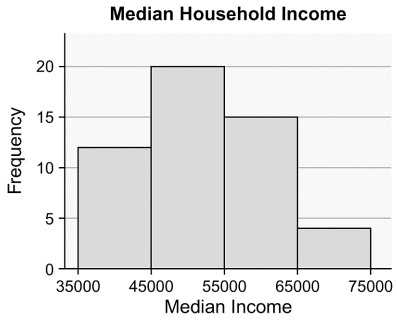
(d)

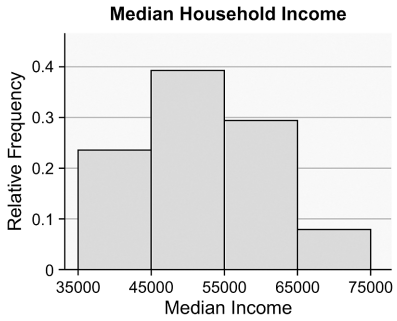


(e) The shape of the distribution is approximately symmetric.

(f) Relative frequency for the median income 35,000–44,999 is 12/51 = 0.2353, and so on.

Income	Freq.	Rel. Freq.
35,000–44,999	12	0.2353
45,000–54,999	20	0.3922
55,000–64,999	15	0.2941
65,000–74,999	4	0.0784





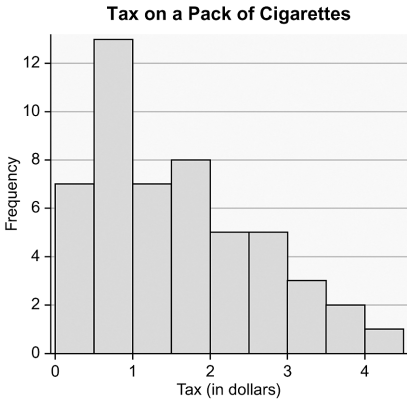
Distribution appears to be approximately symmetric, however, one could argue it is slightly skewed to the right (since the tail on the right is longer than the tail on the left).

(g) Answers will vary, but the graph with a class width of $5000 seems to show more details about the data so it seems better.

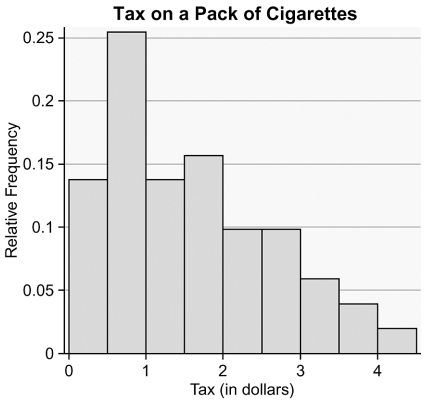
29. (a), (b)
Total number of data points = 51
Relative frequency of 0–0.499 is 7/51 = 0.1373, and so on.

Cigarette Tax	Frequency	Relative Frequency
0.00–0.499	7	0.1373
0.50–0.999	13	0.2549
1.00–1.499	7	0.1373
1.50–1.999	8	0.1569
2.00–2.499	5	0.0980
2.50–2.999	5	0.0980
3.00–3.499	3	0.0588
3.50–3.999	2	0.0392
4.00–4.499	1	0.0196

(c)



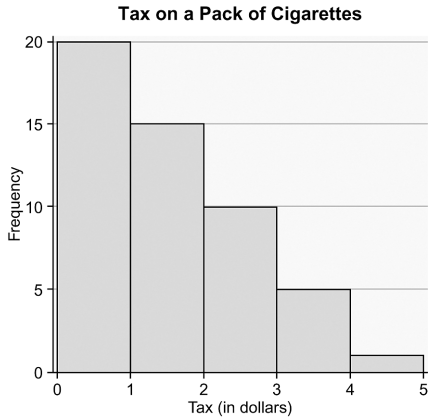
(d)

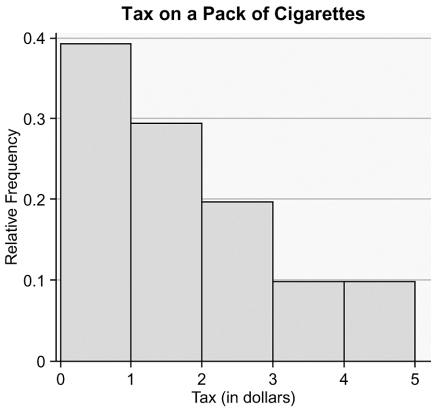


(e) The distribution appears to be right skewed.

(f) Relative frequency of 0–0.999 is: 20/51 = 0.3922, and so on.

Cigarette Tax	Frequency	Relative Frequency
0.00–0.999	20	0.3922
1.00–1.999	15	0.2941
2.00–2.999	10	0.1961
3.00–3.999	5	0.0980
4.00–4.999	1	0.0196





The distribution is right skewed.

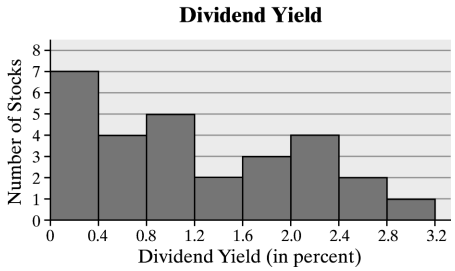
(g) Answers will vary. The first distribution gives a more detailed pattern and does a nice job summarizing the data.

30. (a), (b)

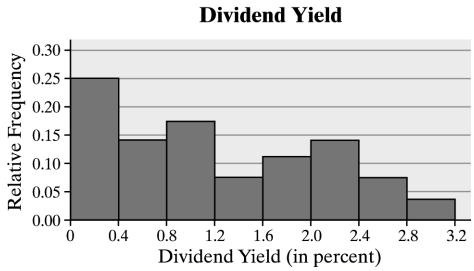
Relative frequency for 0.00–0.39 = 7/28 = 0.2500, and so on.

Dividend	Freq.	Rel. Freq.
0.00 – 0.39	7	0.2500
0.40 – 0.79	4	0.1429
0.80 – 1.19	5	0.1786
1.20 – 1.59	2	0.0714
1.60 – 1.99	3	0.1071
2.00 – 2.39	4	0.1429
2.40 – 2.79	2	0.0714
2.80 – 3.19	1	0.0357

(c)



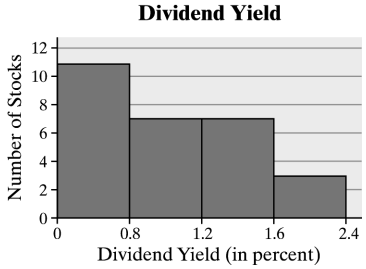
(d)

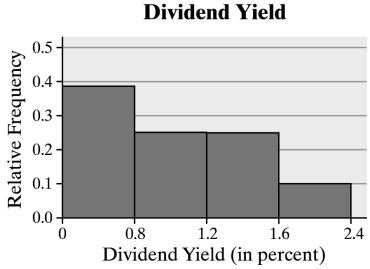


(e) The distribution is skewed right.

(f) Relative frequency for 0.00–0.79 = 11/28 = 0.3929, and so on.

Dividend	Freq.	Rel. Freq.
0.00 – 0.79	11	0.3929
0.80 – 1.59	7	0.2500
1.60 – 2.39	7	0.2500
2.40 – 3.19	3	0.1071





The distribution is skewed right.

(g) Answers will vary. Both distributions indicate the data are skewed right. The first graph is preferred because it gives more detailed information. The second graph is a little too compressed to get a complete view of what is happening with the data.

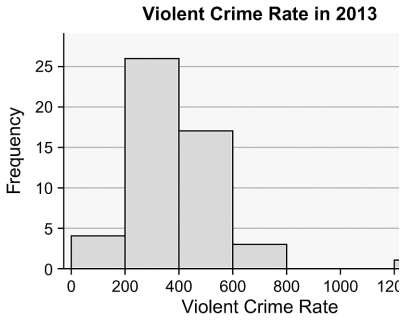
31. Answers will vary. One possibility follows.

(a) Choose a lower class limit of first class of 0 with a class width of 200.

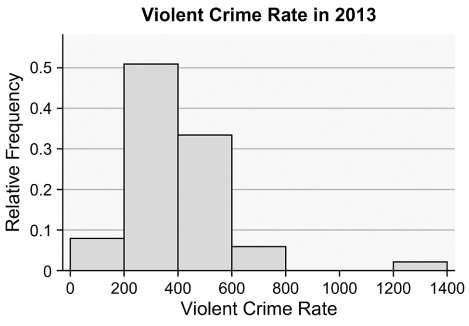
(b), (c)
Relative frequency for 0–199 is 4/51 = 0.0784, and so on.

Violent Crime Rate	Frequency	Relative Frequency
0–199.9	4	0.0784
200–399.9	26	0.5098
400–599.9	17	0.3333
600–799.9	3	0.0588
800–999.9	0	0.0000
1000–1199.9	0	0.0000
1200–1399.9	1	0.0196

(d)



(e)



(f) The distribution is skewed right.

32. Answers will vary. One possibility follows.

(a) We can determine a class width by subtracting the smallest value from the largest, dividing by the desired number of classes, then rounding up. For example,

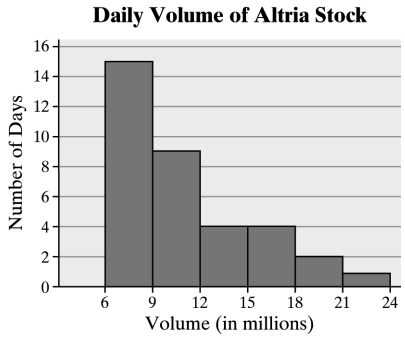
$$\frac{23.59 - 6.37}{6} = 2.87 \rightarrow 3$$

Our first lower class limit should be a nice number below the smallest data value. In this case, 6 is a good first lower limit since it is the nearest whole number below the smallest data value. Thus, we will have a class width of 3, and the first class will have a lower limit of 6.

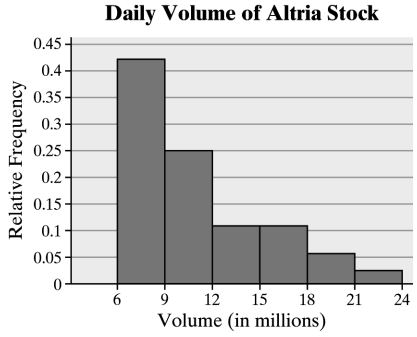
(b), (c)
Relative frequency for 6–8.99 = 15/35 = 0.4286, and so on.

Volume	Freq.	Rel. Freq.
6 – 8.99	15	0.4286
9 – 11.99	9	0.2571
12 – 14.99	4	0.1143
15 – 17.99	4	0.1143
18 – 20.99	2	0.0571
21 – 23.99	1	0.0286

(d)



(e)



(f) The distribution is skewed right.

33. **(a)** **President Ages at Inauguration**

```
4 | 23
4 | 6677899
5 | 0011112244444
5 | 555566677778
6 | 0111244
6 | 589
```

Legend: 4 | 2 represents 42 years.

(b) The distribution appears to be roughly symmetric and bell-shaped.

34. **(a)** Divorce Rates in 2011

```
2 | 4
2 | 677899999
3 | 12223344
3 | 56677888999
4 | 00123344
4 | 5889
5 | 223
5 | 6
```

Legend: 2|4 represents 2.4 per 1000 population

(b) The distribution appears to be roughly symmetric and bell-shaped. One could argue that the distribution is slightly skewed right.

35. **(a)** **Fat in McDonald's Breakfast**

```
0 | 39
1 | 1266
2 | 1224577
3 | 0012267
4 | 6
5 | 159
```

Legend: 5 | 1 represents 51 grams of fat.

(b) The distribution appears to be roughly symmetric and bell-shaped.

36. **(a)** **Gasoline Mileages**

```
2 | 233
2 | 55567889999
3 | 000001111111112222233333333333444444444444
3 | 55555555566666666666778
4 | 0223
```

Legend: 2 | 2 represents 22 miles per gallon.

(b) The distribution appears to be symmetric and bell-shaped.

37. **(a)** Five Year Rate of Return Rounded to the nearest tenth:

10.9	14.2	12.4	13.6	13.0
10.5	10.3	13.1	15.7	14.9
14.1	12.8	13.3	9.9	15.6
12.3	13.9	13.4	19.4	13.4
12.2	14.8	11.9	10.1	13.6
14.6	14.8	13.5	13.9	13.2
14.0	15.2	8.3	9.0	8.7
14.9	16.0	13.7	13.9	12.8

(b) Five Year Rate of Return

```
 8 | 37
 9 | 09
10 | 1359
11 | 9
12 | 23488
13 | 0123445667999
14 | 01268899
15 | 267
16 | 0
17 |
18 |
19 | 4
```

Legend: 8|3 represents 8.3%

(c) The distribution is bell-shaped.

38. **(a)** Home appreciation values rounded to the nearest whole number:

69	149	94	118	87
113	130	65	113	109
350	122	104	94	101
185	150	225	117	107
136	135	113	87	113
115	197	71	96	91
85	105	210	109	125
220	136	127	110	87
75	105	104	97	133
207	93	80	145	67
80				

Home Appreciation

6	579
7	15
8	005777
9	134467
10	14455799
11	03333578
12	257
13	03566
14	59
15	0
16	
17	
18	5
19	7
20	7
21	0
22	05
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	0

Legend: 6|5 represents 65

(b) The shape of the distribution is skewed to the right.

(c) Answers will vary. However, a histogram is probably a better choice because of the wide range of possible values.

39. **(a)** Violent crime rates rounded to the nearest tens:

450	1240	350	260	410	560	320
600	490	220	450	350	320	280
430	380	500	270	240	640	200
470	240	120	260	300	410	
420	210	480	610	470	210	
310	410	410	190	250	140	
280	350	450	290	350	190	
550	260	230	560	250	300	

(b) Violent Crime Rates by State, 2013

1	2499
2	0112344556667889
3	0012255558
4	1111235557789
5	0566
6	014
7	
8	
9	
10	
11	
12	4

Legend: 1|2 represents 120 violent crimes per 100,000 population

(c) Violent Crime Rates by State, 2013

1	24
1	99
2	0112344
2	556667889
3	00122
3	55558
4	111123
4	5557789
5	0
5	566
6	014
6	
7	
7	
8	
8	
9	
9	
10	
10	
11	
11	
12	4

Legend: 1|2 represents 120 violent crimes per 100,000 population

(d) Answers will vary. The first display is decent. It clearly shows that the distribution is skewed right and has an outlier. The second display is not as good as the first. Splitting the stems did not reveal any additional information and has made the display more cluttered and cumbersome.

40. **(a)** **Ages of Academy Award Winners**

Best Actor		Best Actress
9	2	125668999
998877766220	3	012233333455689
8765554332200	4	11245599
5432100	5	
200	6	112
6	7	4
	8	0

Legend: 6|7|4 represents 76 years old for Best Actor and 74 years old for Best Actress.

(b) Answers will vary. It appears that Academy Award winners for best actor tend to be older on the whole than winners for best actress.

41. (a)

Home Run Distances

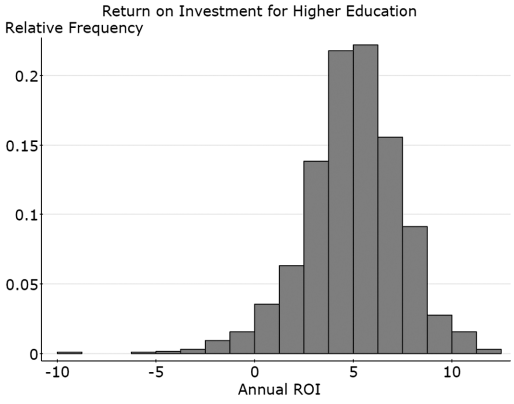
McGwire		Bonds
	32	0 0
	33	
1 0	34	7
0 0	35	0
9 0 0 0	36	0 0 0 1 5
7 0 0 0 0	37	0 0 5 5 5 5
8 5 5 0 0 0 0 0	38	0 0 0 0 0 5
8 0 0 0 0	39	0 0 1 4 6
9 0 0	40	0 0 0 0 4 5
0 0 0 0 0	41	0 0 0 0 0 0 0 0 0 0 0 1 5 5 6 7 7
5 3 0 0 0 0 0	42	0 0 0 0 0 0 0 0 9
0 0 0 0 0 0 0	43	0 0 0 0 0 5 5 6
0 0 0	44	0 0 0 0 2
8 2 0 0 0 0	45	0 4
1 0 0	46	
8 0 0 0	47	
0	48	8
	49	
0	50	
0 0	51	
7	52	
	53	
	54	
0	55	

Legend: 0 | 34 | 7 represents 340 feet for McGwire and 347 feet for Bonds.

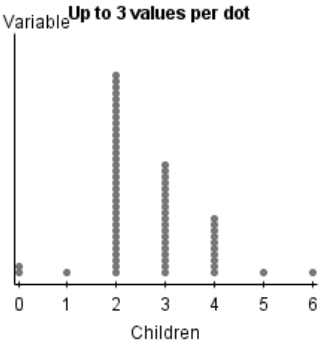
(b) Answers will vary. For both players, the distances of home runs mainly fall from 360 to 450 feet. McGwire has quite a few extremely long distances.

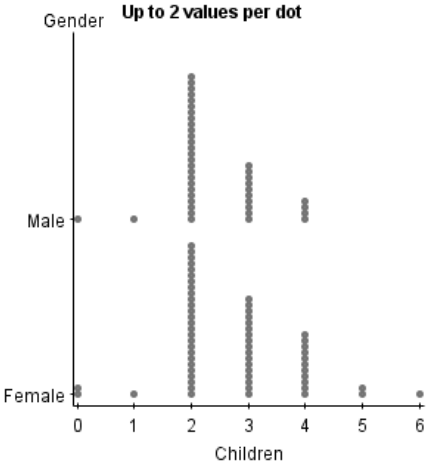
42. Answers will vary.

43. Answers will vary. It is disconcerting that some schools have a negative ROI.



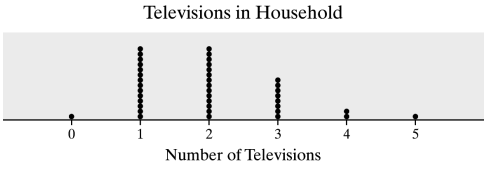
44.



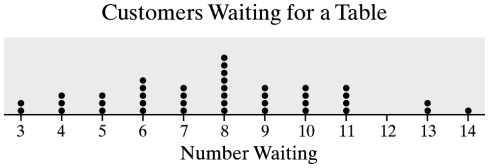


There are several similarities in the distribution of the ideal number of children, as reported by males and females. However, females seem more likely to deem larger families as ideal.
A histogram would better serve us in comparing the preferences between males and females.

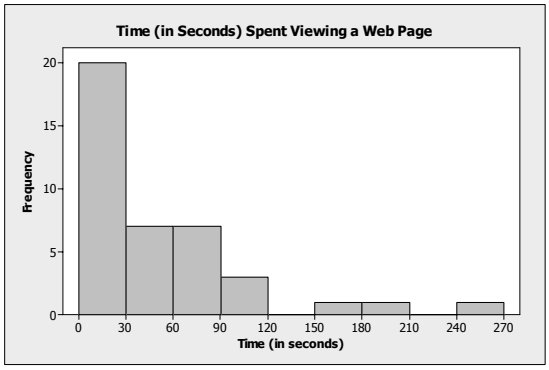
45.



46.



47. Because the data are quantitative, either a stem-and-leaf plot or a histogram would be appropriate. There were 20 people who spent less than 30 seconds, 7 people spent at least 30 seconds but less than 60 seconds, etc. One possible histogram is:



The data appear to be skewed right with a gap and one potential outlier. It seems as if the majority of surfers spent less than one minute viewing the page, while a few surfers spent several minutes viewing the page.

48. Age: histogram, stem-and-leaf plot, or dot plot; Income: histogram or stem-and-leaf plot; Marital status: bar graph or pie chart; Number of vehicles: histogram, stem-and-leaf plot, or dot plot

49. Classes should not overlap to avoid any confusion as to which class an observation belongs to.

50. Histograms are useful for large data sets or data sets with a large amount of spread. Stem-and-leaf plots are nice because the raw data can easily be retrieved. A disadvantage of stem-and-leaf plots is that sometimes the data must be rounded, truncated, or adjusted in some way that requires extra work. Furthermore, if these steps are taken, the original data is lost and a primary advantage of stem-and-leaf plots is lost.

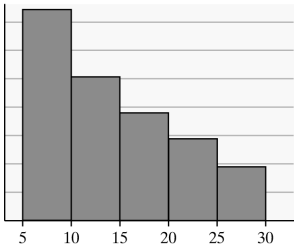
51. There is no such thing as the correct choice for a class width, however some choices are better than others. For example, if the class width is too small, the histogram will show many gaps between the bars. If the class width is too large, the histogram may not provide enough detail.

52. Relative frequencies should be used when comparing two data sets with different sample sizes.

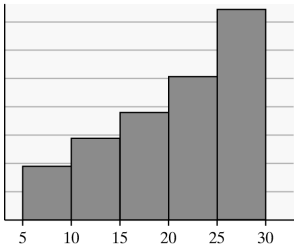
53. Answers will vary. The exercise illustrates the fact that there is no such thing as the "correct" histogram. However, some histograms are better than others and class width can affect the shape of a graph.

54. Answers will vary. Sample histograms are given below.

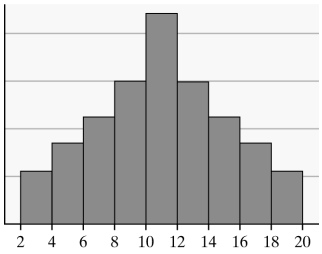
Skewed Right



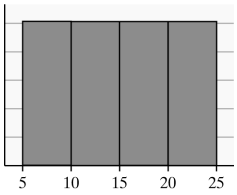
Skewed Left



Bell-Shaped



Uniform



A histogram is skewed left if it has a long tail on the left side. A histogram is skewed right if it has a long tail on the right side. A histogram is symmetric if the left and right sides of the graph are roughly mirror images of each other.

Section 2.3

1. An ogive is a line graph of cumulative frequencies or cumulative relative frequencies against upper class limits.

2. Time series are data measuring the value of a variable at different points in time.

3. True

4. False; when plotting a frequency polygon, we plot the frequency (or number) for each class above the class midpoint and connect the points with straight line segments.

5. **(a)** 0.5; The class width is the difference between successive class midpoints. (e.g. 1.25 – 0.75 = 0.5).
There are 6 classes represented in the graph. The 8 plotted points are for the 6 class midpoints plus two additional points to connect the graph to the horizontal axis on the ends.

(b) The midpoint of the first class is 1.25. The lower limit of the first class is 1.00 and the upper limit is 1.49.

(c) The midpoint of the last class is 3.75. The lower limit of the last class is 3.50 and the upper limit is 3.99.

(d) The lower and upper limits of the class with 25 students are 2.50 and 2.99, respectively.

(e) The lower and upper limits of the class with the fewest students are 1.00 and 1.49, respectively.

6. **(a)** 4; The class width is the difference between successive class upper limits (e.g. 2–(–2) =4).
There are 6 classes represented in the graph. The 8 plotted points are for the 6 class midpoints plus two additional points to connect the graph to the horizontal axis on the ends.

(b) The midpoint of the first class is 2. The lower limit of the first class is 0 and the upper limit is 3.9.

(c) The midpoint of the last class is 22. The lower limit of the last class is 20 and the upper limit is 23.9.

(d) The most popular number of hours spent exercising each week is 0 to 3.9 hours, which is the group with the highest frequency.

(e) The least popular number of hours spent exercising each week is 16 to 19.9 hours, which is the group with the lowest frequency.

(f) The lower and upper limits of the class with 55 students are 4 and 7.9, respectively.

7. **(a)** 5; The class width is the difference between successive class midpoints. (e.g. 24.9 – 19.9 = 5).

(b) From the graph, it appears that approximately 10% of all four-year universities have a graduation rate below 34.9%.

(c) From the graph, it appears that approximately 60% of all four-year universities have a graduation rate below 59.9%.

(d) From the graph, it appears that approximately 5% of all four-year universities have a graduation rate above 90%.

8. **(a)** 2; The class width is the difference between successive class upper limits (e.g. 3.9–1.9).

(b) From the graph, it appears that approximately 92% of all tornadoes had a length less than 9.9 miles.

(c) From the graph, it appears that approximately 80% of all tornadoes had a length less than 5.9 miles.

(d) From the graph, it appears that more than 50% of all tornadoes exceed a length of 2 miles.

9. **(a)** From the graph, it appears the unemployment rate in 2011 was about 9%.

(b) The highest unemployment rate was about 9.8%. This occurred in 2010.

(c) The highest inflation rate was about 4.3%. This occurred in 2008.

(d) The unemployment rate and inflation rate were closest in 2001. The unemployment rate and inflation rate were furthest in 2009.

(e) The misery index for 1999 was approximately 4.2 + 1.8 = 6. The misery index for 2014 was approximately 6.5 + 1.5 = 8. According to the misery index, the year 2014 was more "miserable" than the year 1999.

(f) Since 2010, the misery index has been declining due to the decreases in unemployment each year.

10. (a) To the nearest year, the average age of a man who first married in 1980 was 25.

(b) To the nearest year, the average age of a woman who first married in 1960 was 21.

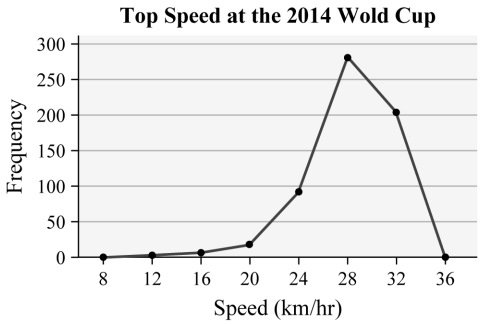
(c) The largest difference in the average age of men and women at which they first married occurred in 1950. The approximate age difference was 24 – 20.5 = 3.5 years.

(d) The least amount of difference in the average age of men and women at which they first married occurred in 2000. The approximate age difference was 26.5 – 25.2 = 1.3 years.

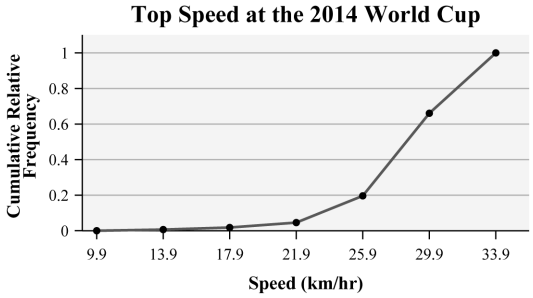
11. (a), (b)

Speed (km/h)	Cumulative Frequency	Cumulative Relative Frequency
10–13.9	4	0.0066
14–17.9	11	0.0181
18–21.9	28	0.0461
22–25.9	119	0.1960
26–29.9	401	0.6606
30–33.9	607	1.0000

(c)



(d)

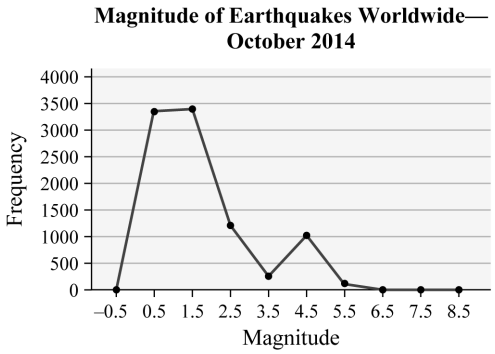


12. (a), (b)

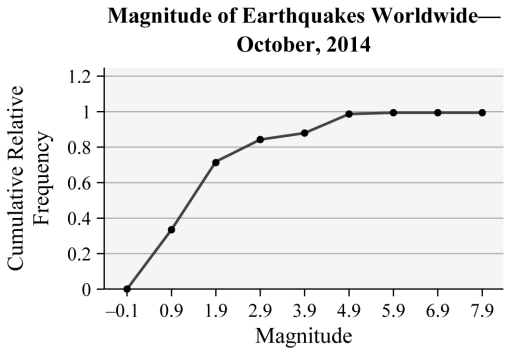
Total number of earthquakes is: 9469
Second class cumulative frequency is 3371 + 3400 = 6671, and so on.
Second class cumulative relative frequency = 6771/9469 = 0.7151, and so on.

Earthquake Magnitude	Cumulative Frequency	Cumulative Relative Frequency
0.0–0.9	3371	0.3560
1.0–1.9	6771	0.7151
2.0–2.9	8008	0.8457
3.0–3.9	8294	0.8759
4.0–4.9	9339	0.9863
5.0–5.9	9460	0.9990
6.0–6.9	9467	0.9998
7.0–7.9	9469	1.0000

(c)



(d)

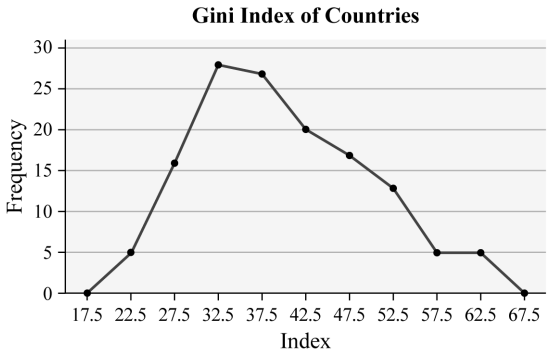


13. (a), (b)

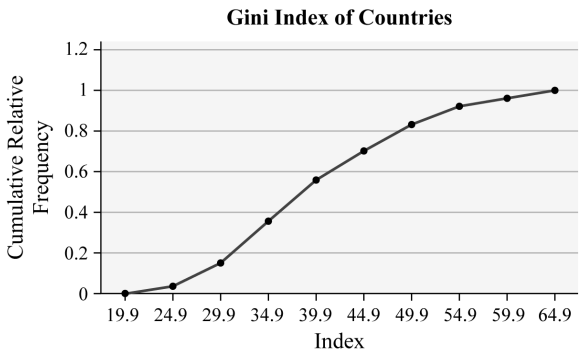
Cumulative frequency of a Gini Index of 25–29.9 = 5 + 16 = 21, and so on. The relative cumulative frequency of 25–29.9 is 21/136 = –.1544, and so on.

Gini Index	Cumulative Frequency	Rel. Cum. Freq.
20–24.9	5	0.0368
25–29.9	21	0.1544
30–34.9	49	0.3603
35–39.9	76	0.5588
40–44.9	96	0.7059
45–49.9	113	0.8309
50–54.9	126	0.9265
55–59.9	131	0.9632
60–64.9	136	1.0000

(c)



(d)

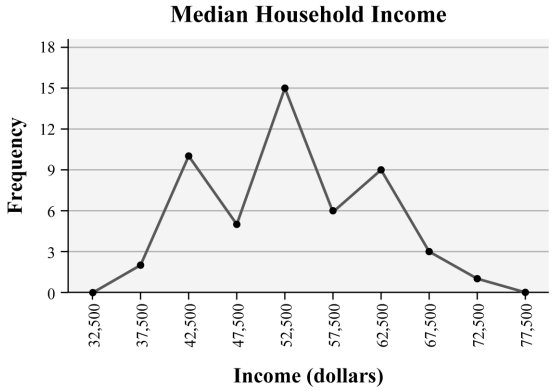


14. (a), (b)

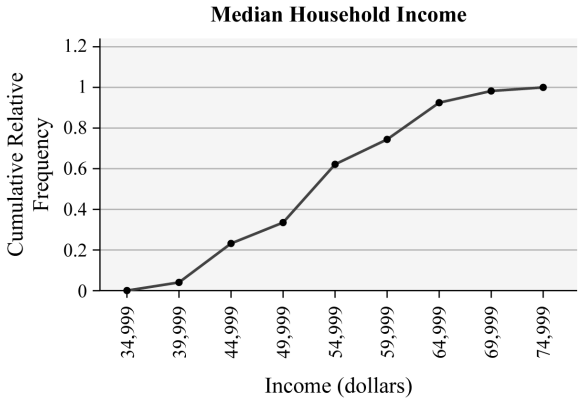
Second class cumulative frequency is 2 + 10 = 12, and so on.
Second class cumulative relative frequency =12/51 = 0.2353, and so on.

Income	Cum. Freq.	Rel. Cum. Freq.
35,000–39,999	2	0.0392
40,000–44,999	12	0.2353
45,000–49,999	17	0.3333
50,000–54,999	32	0.6275
55,000–59,999	38	0.7451
60,000–64,999	47	0.9216
65,000–69,999	50	0.9804
70,000–74,999	51	1.0000

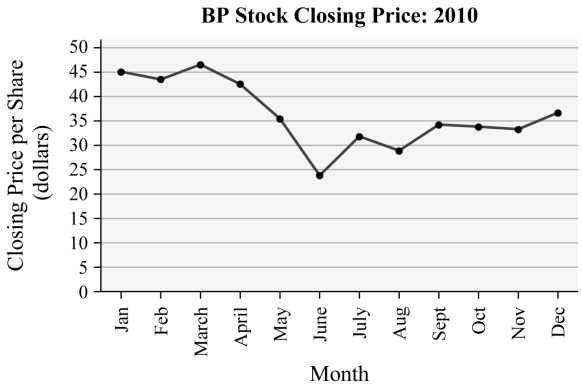
(c)



(d)



15. (a)



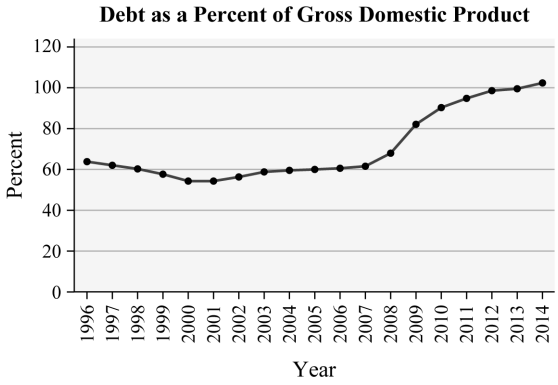
(b) The value of the BP stock at the end of May 2010 was 35.72 and was only 24.02 at the end of June 2010. The percentage change in the BP stock price from May to June 2010 was (24.02–35.72)/35.72 = –0.328, which is a decrease of 32.8%.

16. (a)

(b) The closing price of Twitter stock in November 2013 was 41.57 and was 63.65 in December 2013. The percentage change from November to December 2013 was (63.65–41.57)/41.57 = 0.531, which is an increase of 53.1%.

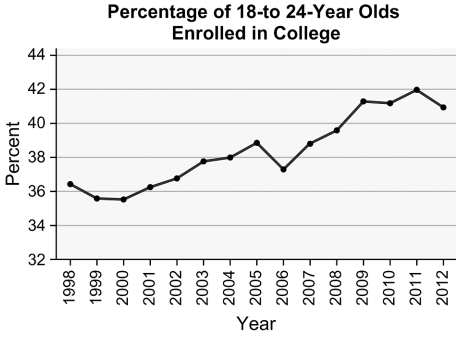
The closing price of Twitter stock in November 2013 was 41.57 and was 41.47 in October 2014. The percentage change from November 2013 to October 2014 was (41.47–41.57)/41.57 = –0.002, which is a decrease of 0.2%. The closing price of Twitter stock increased 53.1% between November and December 2013, so an investor would have been wise to sell in December 2013.

17. (a)



Answers will vary. The time-series plot shows that debt as a percent of gross domestic product remained relatively stable around 60% from 1996 to 2007 and then began to increase steadily from 2008 to 2014.

18.

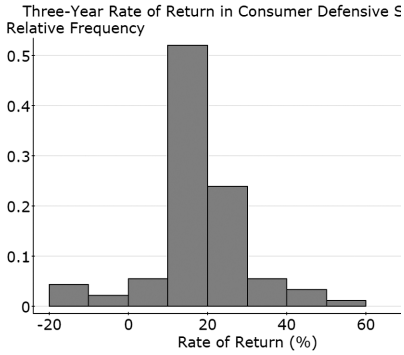


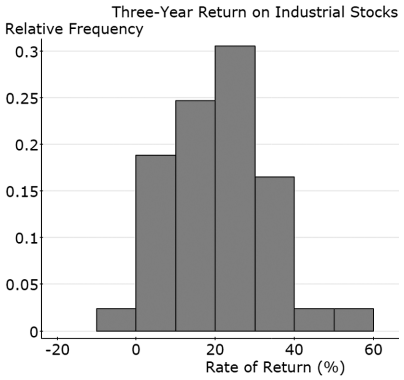
Answers will vary. The time-series plot shows that the percentage of high school graduates enrolling in college seems to have increased over the given time period amid a variety of fluctuations with a slight downturn in 2012.

19. (a)

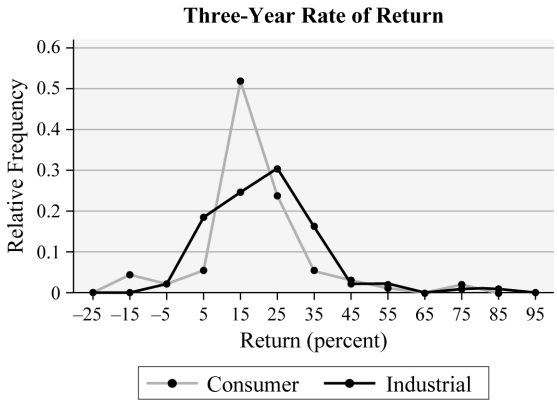
Return (Percent)	Consumer Defense	Industrial
–20 to –10.01	0.0435	0.0000
–10 to –0.01	0.0217	0.0235
0 to 9.99	0.0543	0.1882
10 to 19.99	0.5217	0.2471
20 to 29.99	0.2391	0.3059
30 to 39.99	0.0543	0.1647
40 to 49.99	0.0326	0.0235
50 to 59.99	0.0109	0.0235
60 to 69.99	0.0000	0.0000
70 to 79.99	0.0217	0.0118
80 to 89.99	0.0000	0.0118

(b)

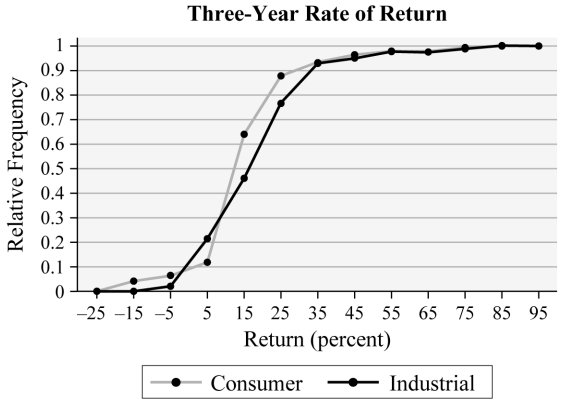




(c)



(d)



(e) Answers will vary. However, industrial stocks had a higher proportion of high-return stocks and had fewer stocks with negative returns.

20. Answers will vary. Reports should address the fact that the number of people going to the beach and participating in underwater activities (e.g. scuba diving, snorkeling) has also increased, so an increase in shark attacks is not unexpected. A better comparison would be the rate of attacks per 100,000 beach visitors. The number of fatalities could decrease due to better safety equipment (e.g. bite resistant suits) and better medical care.

21. Answers will vary.

22. Answers will vary. The cumulative relative frequency is the proportion of observations less than or equal to the current class. All the observations are less than or equal to the highest class, so the cumulative relative frequency of the last class must be 1, or 100%.

23. Time-series plots are drawn with quantitative variables. They are drawn to see trends in the data.

Section 2.4

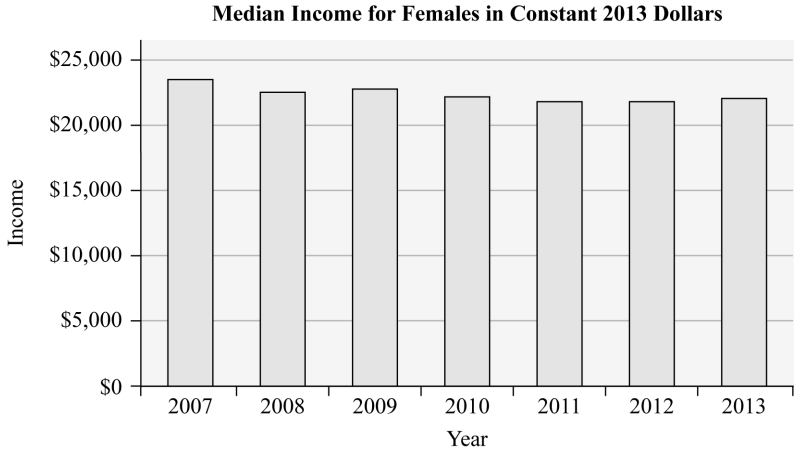
1. The lengths of the bars are not proportional. For example, the bar representing the cost of Clinton's inauguration should be slightly more than 9 times as long as the one for Carter's cost, and twice as long as the bar representing Reagan's cost.

2. **(a)** Answers will vary. The lengths of the bars are not proportional. For example, the bar for soda is 1/3 the size of the bar for a cheeseburger, but the number of steps for a cheeseburger is just over twice that for the soda. In addition, it is unclear where the graph begins: at the base of each figure or the bottom of the platform.

(b) Answers will vary. The pictures could be replaced by simple bars (of the same width) that are proportional in area.

3. **(a)** The vertical axis starts at \$21,500 instead of \$0. This tends to indicate that the median earnings for females changed at a faster rate than actually occurred.

(b) This graph indicates that the median earnings for females has decreased slightly over the given time period.



4. **(a)** The vertical axis starts at 4 instead of 0. This may lead the reader to conclude, for example, the percentage of employed people aged 55–64 who are members of a union is more than double the percentage of those aged 25–34 years.

(b)

5. The bar for 12p–6p covers twice as many hours as the other bars. By combining two 3-hour periods, this bar looks larger compared to the others, making afternoon hours look more dangerous. If this bar were split into two periods, the graph may give a different impression. For example, the graph may show that daylight hours are safer.

6. The article is basing its conclusion on a comparison of categories that do not cover the same number of years. A better comparison is the incidence rate (number of accidents per 100,000 licensed drivers). [Note: only about 14% of licensed drivers in 2005 were aged 24 years or younger.]

7. Answers will vary. This graph is misleading because it does not take into account the size of the population of each state. For example, Vermont is going to pay less in total taxes than California simply because its population is so much lower. There are many variables that should be considered on per capita basis. For example, this graph would be less misleading if it was drawn to represent taxes paid per capita (per person).

8. **(a)** The oil reserves in 2014 were 691.0 million barrels, whereas the oil reserves in 1977 were 7.5 million barrels. The oil reserves in 2014 were 92 times as large as in 1977 (e.g. 691/7.5=92.1). Thus, the graphic for 2014 should be roughly 92 times larger than the graphic for 1977.

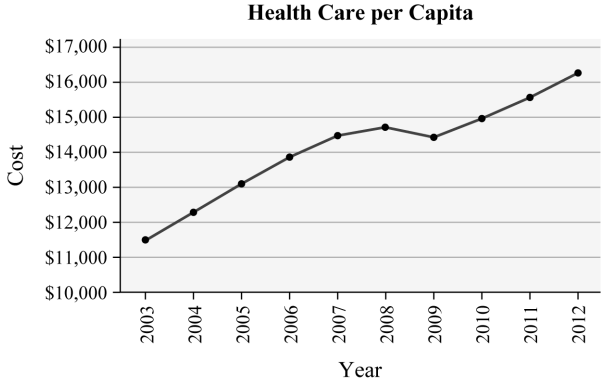
(b) Assuming no change in U.S. oil production, the U.S. strategic oil reserves would last approximately 90 days (e.g. 691/7.7 = 89.74 days).

9. **(a)** The graphic is misleading because the bars are not proportional. The bar for housing should be a little more than twice the length of the bar for transportation, but it is not.

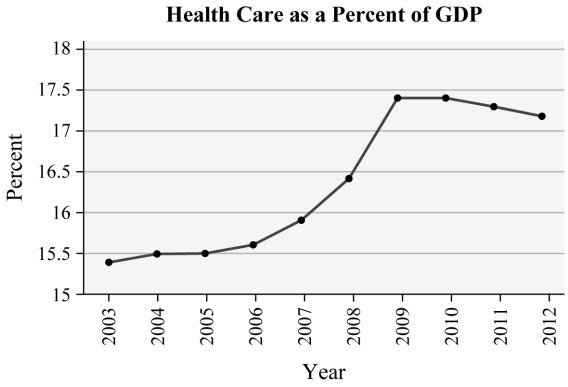
(b) The graphic could be improved by adjusting the bars so that their lengths are proportional.

10. The graph does not support the safety manager's claim. The vertical scale starts at 0.17 instead of 0, so the difference between the bars is distorted. While there was a decrease, it appears that the decrease is roughly 10% of the 1992 rate.

11. **(a)** Answers will vary. Here is a time-series plot that a politician might use to support the position that health care is increasing.

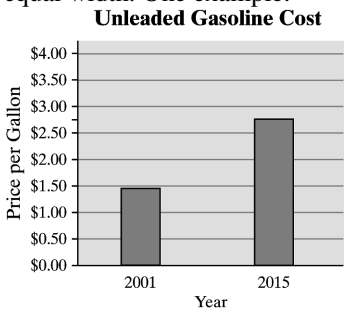


(b) Answers will vary. Here is a time-series plot that the health care industry might use to refute the opinion of the politician.

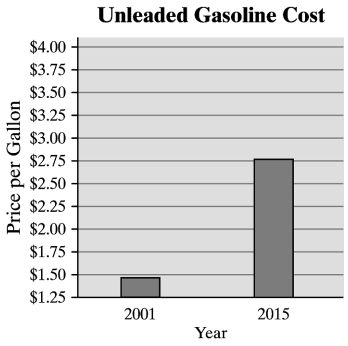


(c) Answers will vary. Changing the scale on the graph will affect the message. The message is also affected by using the variable "Health Care per Capita" rather than "Health Care as a Percent of GDP."

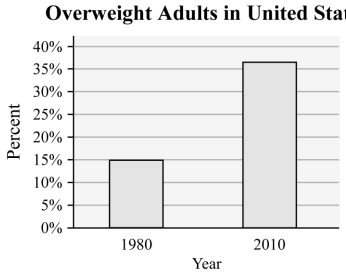
12. **(a)** A graph that is not misleading will use a vertical scale starting at \$0 and bars of equal width. One example:



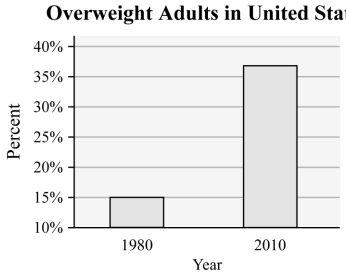
(b) A graph that is misleading might use bars of unequal width or will use a vertical scale that does not start at \$0. One example, as follows, is misleading because it starts at \$1.25 instead of 0 without indicating a gap. This might cause the reader to conclude that cost of unleaded gasoline has risen more sharply than actually occurred.



13. **(a)** A graph that is not misleading will use a vertical scale starting at 0% and bars of equal width. One example:



(b)

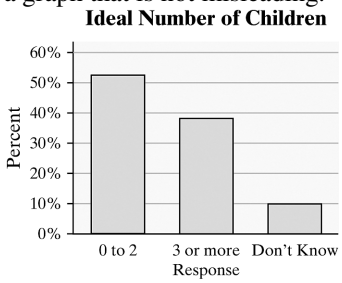


This graphic is misleading because the vertical scale starts at 10% instead of 0% without indicating a gap. This might cause the reader to think that the proportion of overweight adults in the United States is increasing more quickly than they really are.

14. **(a)** A bar graph

(b) A reader cannot tell whether the graph ends at the top of the nipple on the baby bottle, or at the end of the milk.

(c) Answers will vary. Here is an example of a graph that is not misleading.



15. Answers will vary. Three-dimensional graphs are deceptive because the pieces are not proportional. For example, the area for P (pitcher) looks substantially larger than the area for 3B (third base), even though both are the same percentage. Graphs should not be drawn using three dimensions. Instead, use two dimensions.

16. Answers will vary. This is a histogram so the bars should touch. In addition, there are no labels and no title.

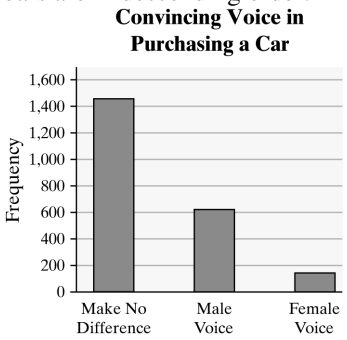
Chapter 2 Review Exercises

1. **(a)** There are 614 + 154 + 1448 = 2216 participants.

(b) The relative frequency of the respondents indicating that it makes no difference is

$$\frac{1448}{2216} \approx 0.653$$

(c) A Pareto chart is a bar chart where the bars are in descending order.



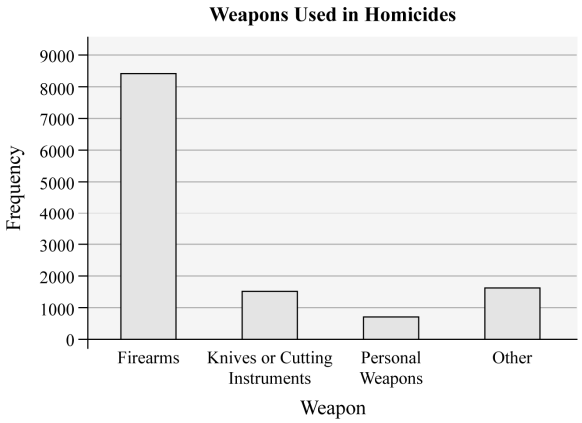
(d) Answers will vary.

2. (a) Total homicides = 8438 + 1486 + 685 + 1621 = 12230
Relative frequency for firearms is 8438/12230 = 0.6899, and so on.

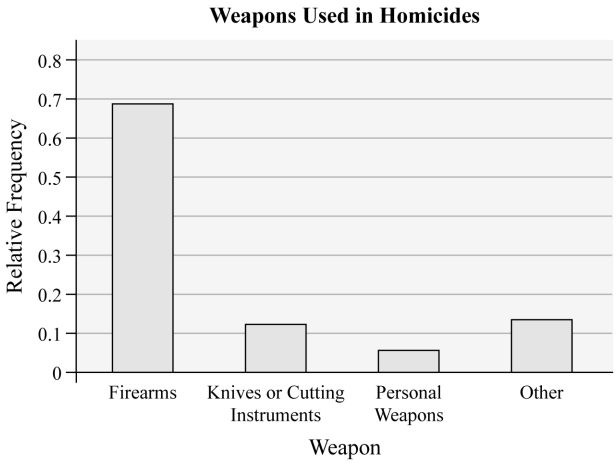
Type of Weapon	Relative Frequency
Firearms	0.6899
Knives or cutting instruments	0.1215
Personal weapons	0.0560
Other weapon	0.1325

(b) The relative frequency is 0.6899, so 68.99% of the homicides were committed using a firearm.

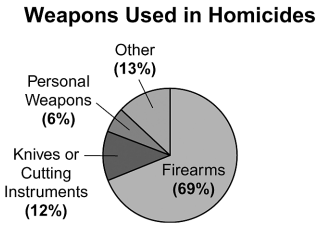
(c)



(d)



(e)



3. **(a)**, **(b)**, and **(c)**

Total births (in thousands) = 3 + 275 + 902 + 1128 + 1044 + 487 + 110 + 7 + 1 = 3957

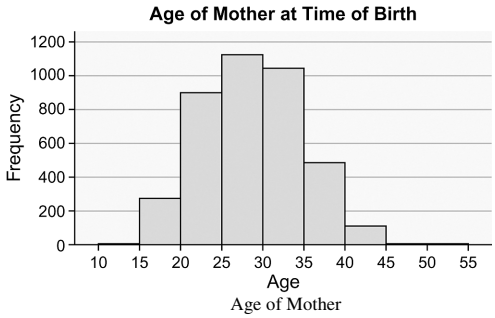
Relative frequency for 10–14 year old mothers = $3 / 3957 \approx 0.0008$, and so on.

Cumulative frequency for 15–19 year old mothers = 3 + 275 = 278, and so on.

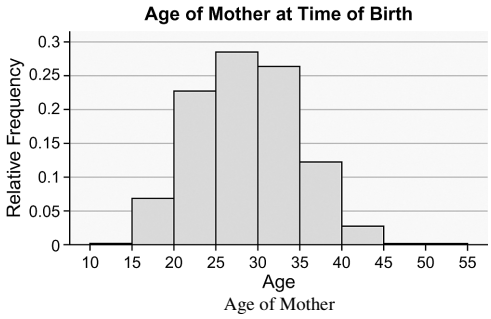
Cumulative relative frequency for 15–19 year old mothers = $278 / 3957 \approx 0.0703$, and so on.

Age of Mother	Rel. Freq.	Cumul. Freq.	Cumul. Rel. Freq.
10 – 14	0.0008	3	0.0008
15 – 19	0.0695	278	0.0703
20 – 24	0.2280	1180	0.2982
25 – 29	0.2851	2308	0.5833
30 – 34	0.2638	3352	0.8471
35 – 39	0.1231	3839	0.9702
40 – 44	0.0278	3949	0.9980
45 – 49	0.0018	3956	0.9997
50 – 54	0.0003	3957	1.0000

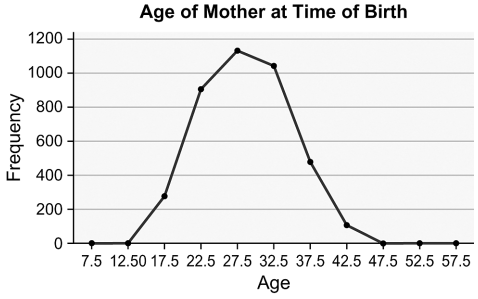
(d) The distribution is roughly symmetric and bell-shaped.



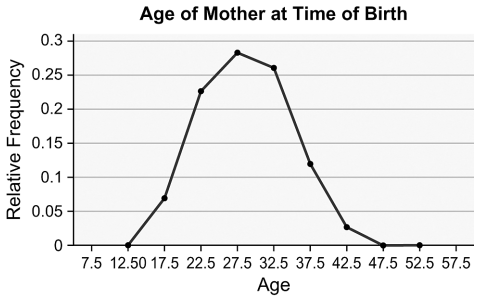
(e)



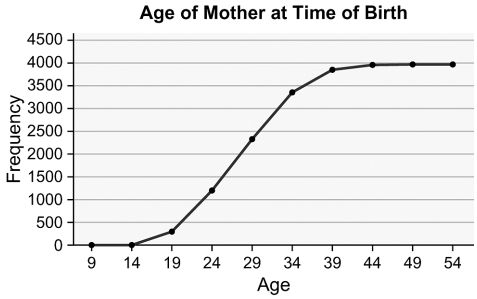
(f)



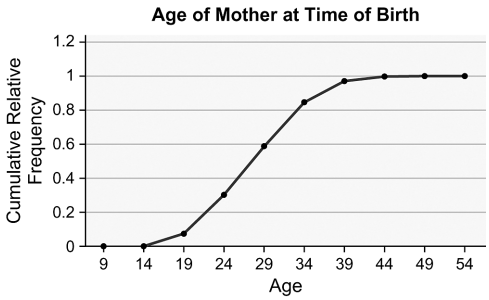
(g)



(h)



(i)



(j) From the relative frequency table, the relative frequency of 20–24 is 0.2280, so the percentage is 22.80%.

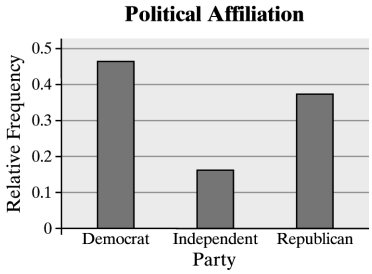
(k) $\dfrac{1044+487+110+7+1}{3957} = \dfrac{1649}{3957} \approx 0.4167$

41.67% of live births were to mothers aged 30 years or older.

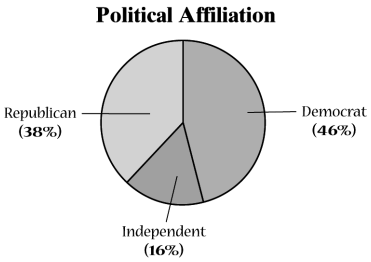
4. (a), (b)

Affiliation	Frequency	Relative Frequency
Democrat	46	0.46
Independent	16	0.16
Republican	38	0.38

(c)



(d)



(e) Democrat appears to be the most common affiliation in Naperville.

5. (a), (b), (c), and **(d)**

Family Size	Freq.	Rel. Freq.	Cumul. Freq.	Cumul. Rel. Freq.
0	7	0.1167	7	0.1167
1	7	0.1167	14	0.2333
2	18	0.3000	32	0.5333
3	20	0.3333	52	0.8667
4	7	0.1167	59	0.9833
5	1	0.0167	60	1.0000

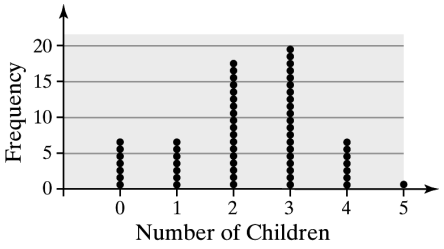
(e) The distribution is more or less symmetric.

(f)

(g) From the relative frequency table, the relative frequency of two children is 0.3000, so 30% of the couples have two children.

(h) From the frequency table, the relative frequency of at least two children (i.e. two or more) is $0.3000 + 0.3333 + 0.1167 + 0.0167 = 0.7667$ or 76.67%. So, 76.67% of the couples have at least two children.

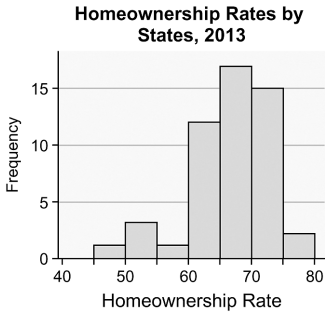
(i)



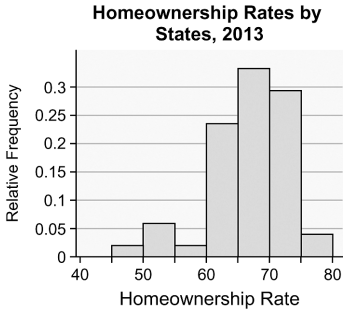
6. (a), (b)

Homeownership Rate	Frequency	Relative Frequency
45–49.9	1	0.0196
50–54.9	3	0.0588
55–59.9	1	0.0196
60–64.9	12	0.2353
65–69.9	17	0.3333
70–74.9	15	0.2941
75–75.9	2	0.0392

(c)



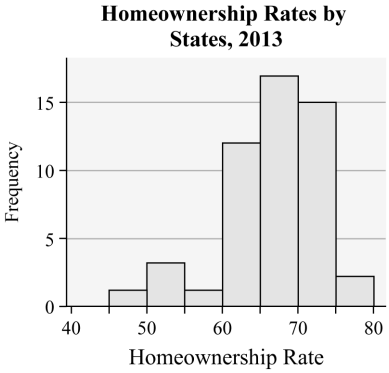
(d)

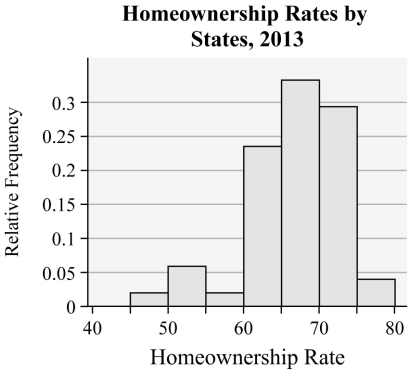


(e) The distribution is skewed left.

(f)

Homeownership Rate	Frequency	Relative Frequency
40–49.9	1	0.0196
50–59.9	4	0.0784
60–69.9	29	0.5686
70–79.9	17	0.3333





The distribution is skewed left.

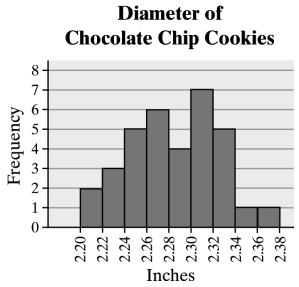
Answers will vary. Both class widths give a good overall picture of the distribution. The first class width provides a little more detail to the graph, but not necessarily enough to be worth the trouble. An intermediate value, say a width of 8, might be a reasonable compromise.

7. (a), **(b)**, **(c)**, and **(d)**

Answers will vary. Using 2.2000 as the lower class limit of the first class and 0.0200 as the class width, we obtain the following.

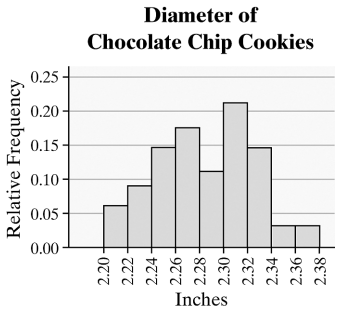
Class	Freq.	Rel. Freq.	Cumul. Freq.	Cumul. Rel.Freq.
2.2000 – 2.2199	2	0.0588	2	0.0588
2.2200 – 2.2399	3	0.0882	5	0.1471
2.2400 – 2.2599	5	0.1471	10	0.2941
2.2600 – 2.2799	6	0.1765	16	0.4706
2.2800 – 2.2999	4	0.1176	20	0.5882
2.3000 – 2.3199	7	0.2059	27	0.7941
2.3200 – 2.3399	5	0.1471	32	0.9412
2.3400 – 2.3599	1	0.0294	33	0.9706
2.3600 – 2.3799	1	0.0294	34	1

(e)



The distribution is roughly symmetric.

(f)



8.

Hours Spent Online

12	
13	467
14	05578
15	1236
16	456
17	113449
18	066889
19	2
20	168
21	119
22	29
23	48
24	4
25	7
26	

Legend: 13 | 4 = average 13.4 hours per week.

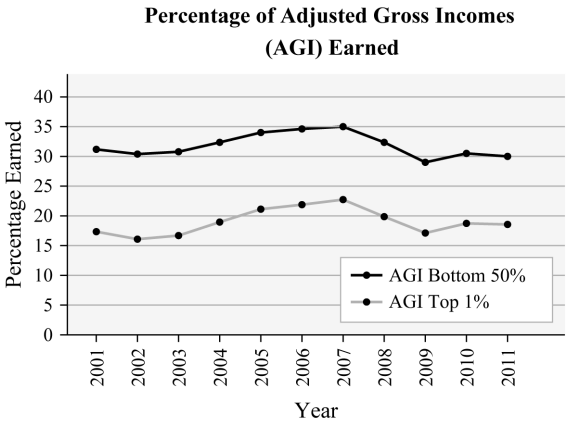
The distribution is slightly skewed right.

9. **(a)** Grade inflation seems to be happening in colleges. GPAs have increased every time period for all schools.

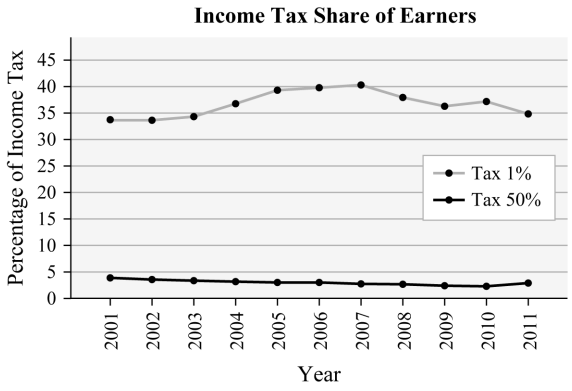
(b) Answers may vary. GPAs increased about 5.6% for public schools. GPAs increased about 6.8% for private schools. Private schools have higher grade inflation because the GPAs are higher and they are increasing faster.

(c) The graph is misleading because it starts at 2.6 on the vertical axis.

10. **(a)** Answers will vary. The adjusted gross income share of the top 1% of earners shows steady increases overall, with a few minor exceptions. The adjusted gross income share of the bottom 50% of earners shows steady decreases overall, with a few minor exceptions.

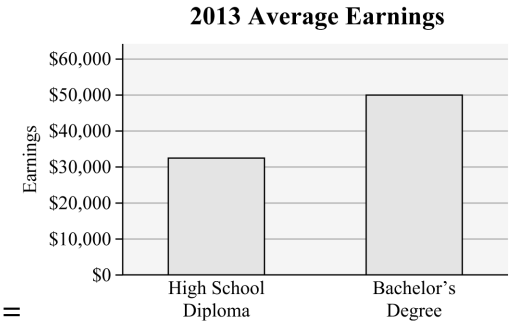


(b) Answers will vary. The income tax share of the top 1% of earners shows steady increases overall, with few exceptions, including a notable decrease from 2007 to 2008. The income tax share of the bottom 50% of earners shows steady decreases over time.



11. **(a)** Graphs will vary. One way to mislead would be to start the vertical scale at a value other than 0. For example, starting the vertical scale at $30,000 might make the reader believe that college graduates earn more than three times what a high school graduate earns (on average).

(b) A graph that does not mislead would use equal widths for the bars and would start the vertical scale at \$0. Here is an example of a graph that is not misleading:



=

12. (a) Flats are preferred the most (40%) and extra-high heels are preferred the least (1%).

(b) The graph is misleading because the bar heights and areas for each category are not proportional.

Chapter 2 Test

1. (a) A 5 Star rating was the most popular rating with 1675 votes.

(b) 35 + 67 +246 + 724 + 1675 = 2747 postings were posted on Yelp for Hot Doug's restaurant.

(c) $1675 - 724 = 951$
There were 951 more 5 Star ratings than 4 Star ratings.

(d) There were 1675 5 Star ratings out of a total of 2747 ratings. $\frac{1675}{2747} \approx 0.6098$
Approximately 61% of all ratings were 5 Star ratings.

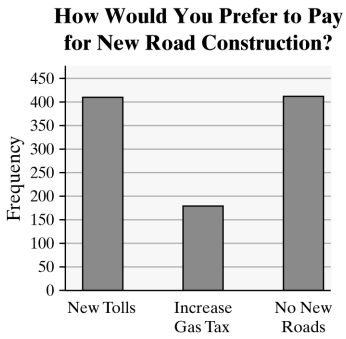
(e) No, it is not appropriate to describe the shape of the distribution as skewed right. The data represented by the graph are qualitative, so the bars in the graph could be placed in any order.

2. (a) There were 1005 responses. The relative frequency who indicated they preferred new tolls was $\frac{412}{1005} = 0.4100$, and so on.

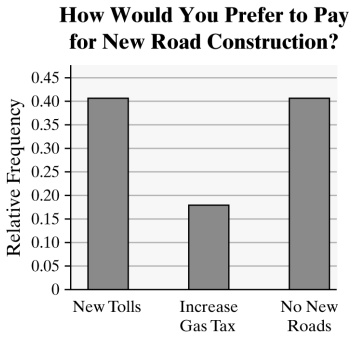
Response	Freq.	Rel. Freq.
New Tolls	412	0.4100
Inc. Gas Tax	181	0.1801
No New Roads	412	0.4100

(b) The relative frequency is 0.1801, so the percentage of respondents who would like to see an increase in gas taxes is 18.01%.

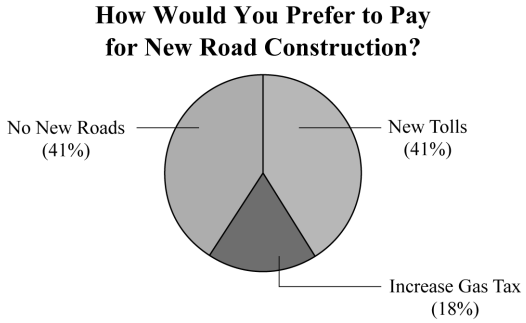
(c)



(d)



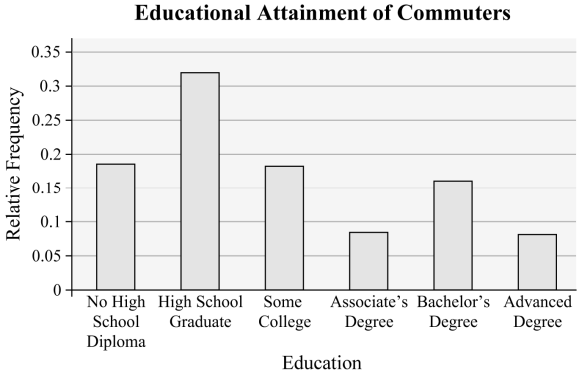
(e)



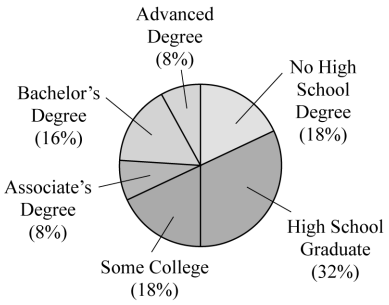
3. **(a)**, **(b)**

Education	Freq.	Rel. Freq.
No high school diploma	9	0.18
High school graduate	16	0.32
Some college	9	0.18
Associate's degree	4	0.08
Bachelor's degree	8	0.16
Advanced degree	4	0.08

(c)



(d) **Educational Attainment of Commuters**

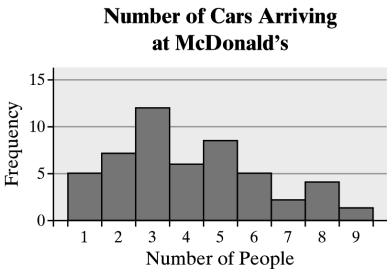


(e) The largest bar (and largest pie segment) corresponds to "High School Graduate," so high school graduate is the most common educational level of a commuter.

4. **(a)**, **(b)**, **(c)**, and **(d)**

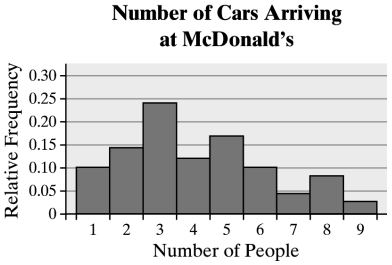
No. of Cars	Freq.	Rel. Freq.	Cumul. Freq.	Cumul. Rel. Freq.
1	5	0.10	5	0.10
2	7	0.14	12	0.24
3	12	0.24	24	0.48
4	6	0.12	30	0.60
5	8	0.16	38	0.76
6	5	0.10	43	0.86
7	2	0.04	45	0.90
8	4	0.08	49	0.98
9	1	0.02	50	1

(e)



The distribution is skewed right.

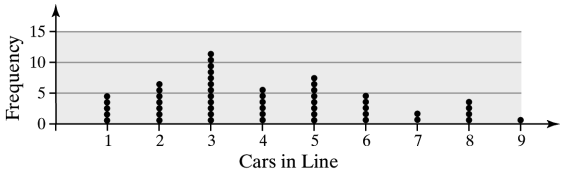
(f)



(g) The relative frequency of exactly 3 cars is 0.24. So, for 24% of the weeks, exactly three cars arrived between 11:50 am and 12:00 noon.

(h) The relative frequency of 3 or more cars
$= 0.24 + 0.12 + 0.16 + 0.10$
$+ 0.04 + 0.08 + 0.02 = 0.76$
So, for 76% of the weeks, three or more cars arrived between 11:50 am and 12:00 noon.

(i)



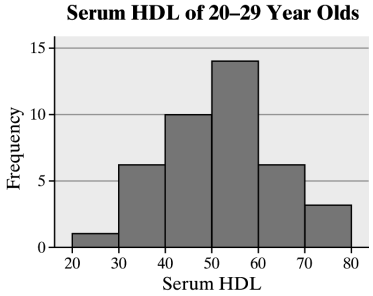
5. Answers may vary. One possibility follows:

(a), (b)

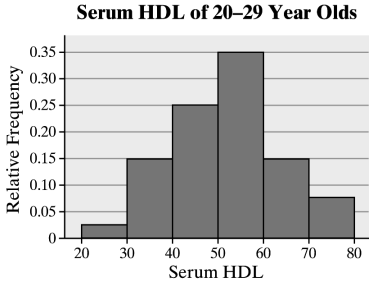
Using a lower class limit of the first class of 20 and a class width of 10:
Total number of data points = 40
Relative frequency of 20 – 29 = 1/40 = 0.025, and so on.

HDL Cholesterol	Frequency	Relative Frequency
20–29	1	0.025
30–39	6	0.150
40–49	10	0.250
50–59	14	0.350
60–69	6	0.150
70–79	3	0.075

(c)



(d)



(e) The distribution appears to be roughly bell-shaped.

6. (a), (b)

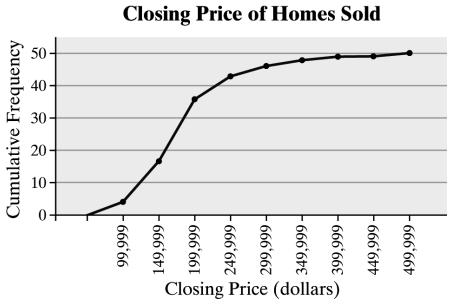
Closing Price	Cumul. Freq.	Cumul. Rel. Freq.
50,000 – 99,999	4	0.08
100,000 – 149,999	17	0.34
150,000 – 199,999	36	0.72
200,000 – 249,999	43	0.86
250,000 – 299,999	46	0.92
300,000 – 349,999	48	0.96
350,000 – 399,999	49	0.98
400,000 – 449,999	49	0.98
450,000 – 499,999	50	1

(c) The cumulative relative frequency for the class \$150,000–\$199,000 is 0.72. Therefore, 72% of the homes sold for less than \$200,000.

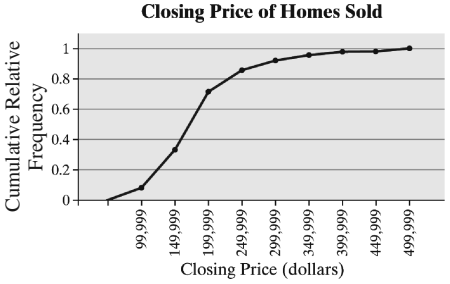
(d)

The distribution is skewed right.

(e)



(f)



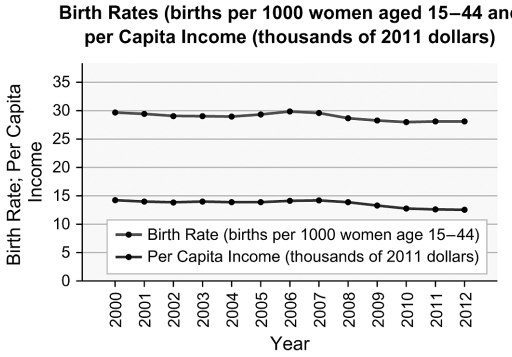
7. The stem-and-leaf diagram below shows an approximately uniform distribution.

Time Spent on Homework

4	0567
5	26
6	13
7	01338
8	59
9	1369
10	3899
11	0018
12	556

Legend: 4 | 0 represents 40 minutes.

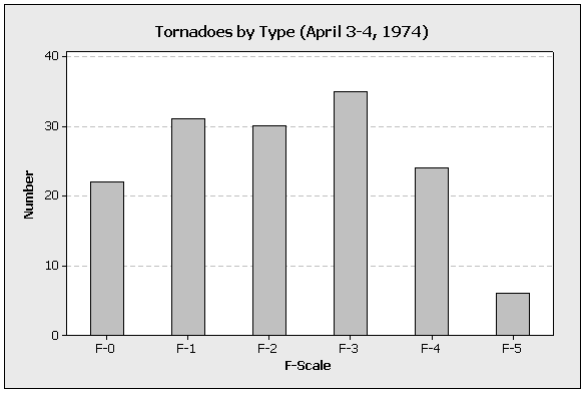
8. The curves in the figure below appear to follow the same trend. Birth rate increases as per capita income increases.

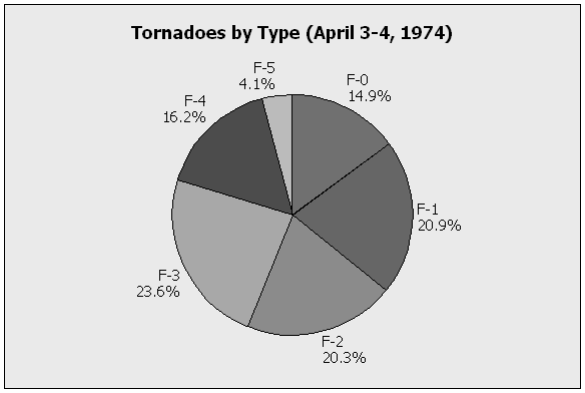


9. Answers may vary. It is difficult to interpret this graph because it is not clear whether the scale is represented by the height of the steps, the width of the steps, or by the graphics above the steps. The graphics are misleading because they must be increased in size both vertically and horizontally to avoid distorting the image. Thus, the resulting areas are not proportionally correct. The graph could be redrawn using bars whose widths are the same and whose heights are proportional based on the given percentages. The use of graphics should be avoided, or a standard size graphic representing a fixed value could be used and repeated as necessary to illustrate the given percentages.

Case Study: The Day the Sky Roared

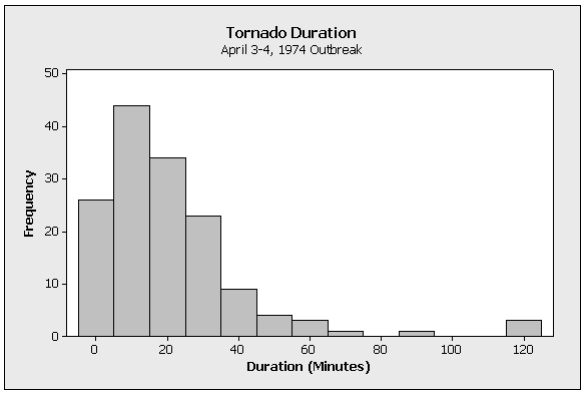
1.



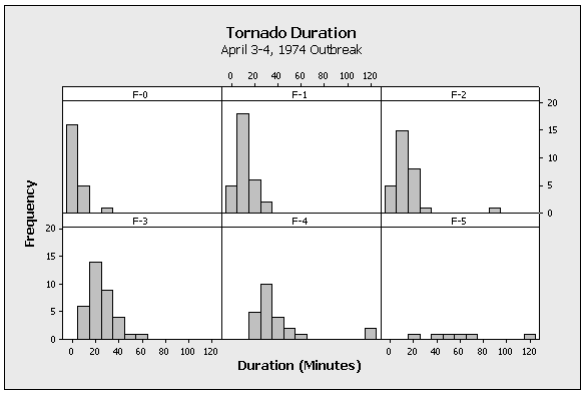


During the April 3-4, 1974 outbreak, about 44% of the tornadoes exceeded F-3 on the Fujita Wind Damage Scale. This was much greater than the 1% that typically occurs.

2. The histogram will vary depending on the class width.



3. Histograms may vary depending on class widths. For comparison purposes, the same class width was used for each histogram.

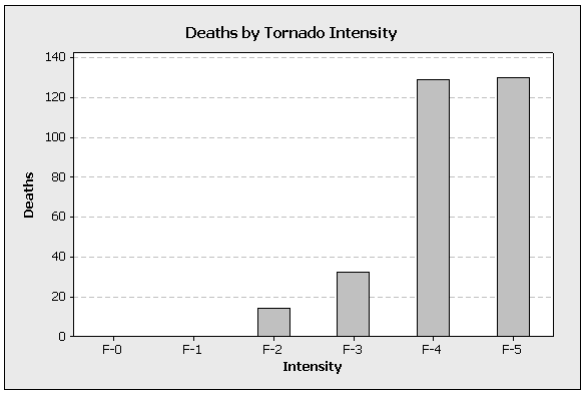


The distributions all appear to be skewed right, though the distribution for F-5 tornadoes is difficult to see due to the low sample size. There is an obvious shift in the distributions. As the strength of the tornado increases, the duration of the tornado increases.

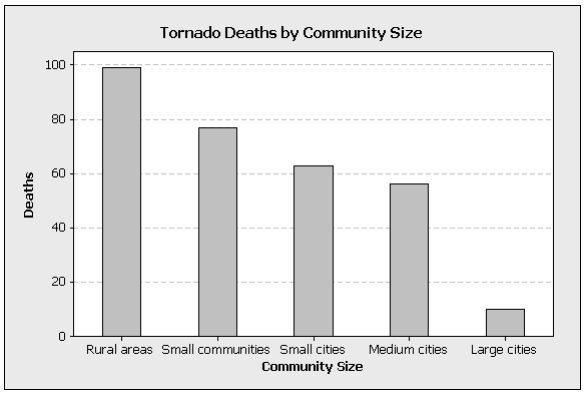
4. There were 305 deaths during the outbreak. Of these, 259 were due to the more severe tornadoes.

$$\frac{259}{305} \approx 0.8492$$

Roughly 85% of the deaths during the outbreak were due to the more severe tornadoes. This may be a little high, but it is consistent since it is greater than 70%.



5. The provided data is not sufficient to determine whether or not tornadoes are more likely to strike rural areas. Some research at Texas A&M University indicates that tornadoes are more likely to occur in urban or suburban areas, possibly due to greater temperature differences. The data does indicate that the number of deaths decreases as the population of the community increases. The higher the population density, the greater the chance that a tornado is detected and reported early, thereby providing more time for residents to take shelter.



6. Answers will vary. The outbreak of April 3-4, 1974 seemed to be more severe in intensity than usual with 20% of the tornadoes being classified as F-4 or F-5. While the shape of the duration distribution was roughly the same for each intensity level, the duration of a tornado increased with its intensity. The number of deaths decreased as the community size increased.

Chapter 2
Organizing and Summarizing Data

Overview

Remember, this course is divided into four parts that correspond to the four parts of our definition of Statistics. Chapters 2 through 4 represent the organizing and summarizing portion of the definition—descriptive statistics.

Chapter 2 discusses describing data through tables and graphs. We summarize raw qualitative data in Section 2.1 and raw quantitative data in Sections 2.2 and 2.3. Section 2.4 provides the opportunity to present misleading graphs.

What to Emphasize

The material in this chapter is elementary. Many of your students have likely seen much of this content in prior classes. Therefore, do not get bogged down in the details of construction of the graphs. Instead, focus on proper graphics construction and how to display qualitative and quantitative data in a fashion that clearly tells the story of the data.

- **Summarizing Qualitative Data -** We begin with a discussion of summarizing qualitative data in tables. This should not be a challenge for the students, so don't spend a lot of class time on this material. Ideally, you will use software such as StatCrunch to build frequency and relative frequency tables. The same goes for the construction of bar graphs and pie charts. Do not emphasize by-hand construction—nobody would ever be expected to construct one of these graphs by hand. Instead, emphasize when each type of graph should be constructed. Bar graphs are typically used when we wish to compare one value of a variable to another, while pie charts are useful when comparing a part to the whole. For example, what proportion of taxes collected by the federal government are from personal income taxes? This is easier to see from a pie chart than a bar graph. Bar graphs can be used to display ordinal data (Poor, Fair, Good, Excellent). Can the same information be displayed effectively in a pie chart?

  Also, emphasize side-by-side bar graphs and the fact that they should be constructed using relative frequencies (since it is not "fair" to compare frequencies when the sample/population sizes differ). Plus, a version of side-by-side bar graphs will be used in Chapters 4 and 12 when we study conditional distributions.
- **Summarizing Quantitative Data -** We decided to segment summarizing quantitative data into two sections—the popular displays (Section 2.2) and other graphs (Section 2.3). Section 2.3 is optional and can be skipped without loss of continuity. When considering which topics you might skip, ask yourself if the topic will be needed or revisited later in the course. For example, graphs such as frequency polygons are not utilized later in the course, so you might consider skipping this topic. That said, some graphs allow for interesting results, such as time series plots. So, judgment should be used when deciding what to skip.

In Section 2.2, do not get bogged down on by-hand creation of frequency or relative frequency distributions or the construction of histograms. Allow technology to do the work so that you may focus on the fact that there are many acceptable class widths that provide a nice summary of the data. However, be sure to emphasize that some class widths are really poor. In fact, spend time making sure students understand the results from the "Exploring Histograms with StatCrunch" activity in the Student Activity Notebook. In addition, spend time on distribution shape. Require students to justify their conclusions regarding shape, rather than simply claiming a distribution is skewed right.

Section 2.3 is entirely optional and may be skipped without loss of continuity. That said, you may consider at least requiring that students review time series graphs since they are popular in the media, and time series data is discussed at various points in the course (especially when we present correlation versus causation).

- **Graphical Misrepresentations of Data -** While this is an optional section, it does have merit. The media is full of examples where graphs mislead or misrepresent data. Be sure to alert students to be on the lookout for poor graphics and require students to clearly label each graph they create.

Ideas for Traditional/Online/Blended/Flipped

Again, real data should be utilized to illustrate concepts.

Whether you are in a traditional or online setting, require students to justify results. For example, create four histograms of the same set of data and ask students (either through discussion groups or small in-class groups) to rank the graphs from best to worst. Change things up by not including titles or labels. Ask students how a graphic might be improved. Activities such as these not only require students to know how the graphics are constructed, but also get them thinking about appropriate techniques. Another idea is to give various data sets to students and ask them to summarize the information. Don't tell them what type of graphic to create—let them decide. Mix the data up between qualitative and quantitative. For example, include a small data set with discrete data (collect data from your class on number of siblings, which is likely best summarized using a dot plot) along with larger data sets that might be better summarized using histograms (use data from the 2014 World Cup at http://www.statcrunch.com/app/index.php?dataid=1130049). If a student or group chooses a histogram, ask for justification of the choice. In a classroom, perhaps the students could be required to present the graphic to the class and explain what message the graphic conveys.

A great **applet** that may be used to emphasize that there is no such thing as the "correct" class width when constructing a histogram is the Histogram with Sliders applet in StatCrunch. Use a data set to build the applet and allow students to experiment with various starting points and class widths.

Look through the **Student Activity Workbook** for in-class activities.

Classroom Examples

Section 2.1

1. The 2010 Census results include a summary of the racial composition of the population. The races reported by the populations of the United States and the State of California are summarized below. The data are given in millions. (Source: census.gov)

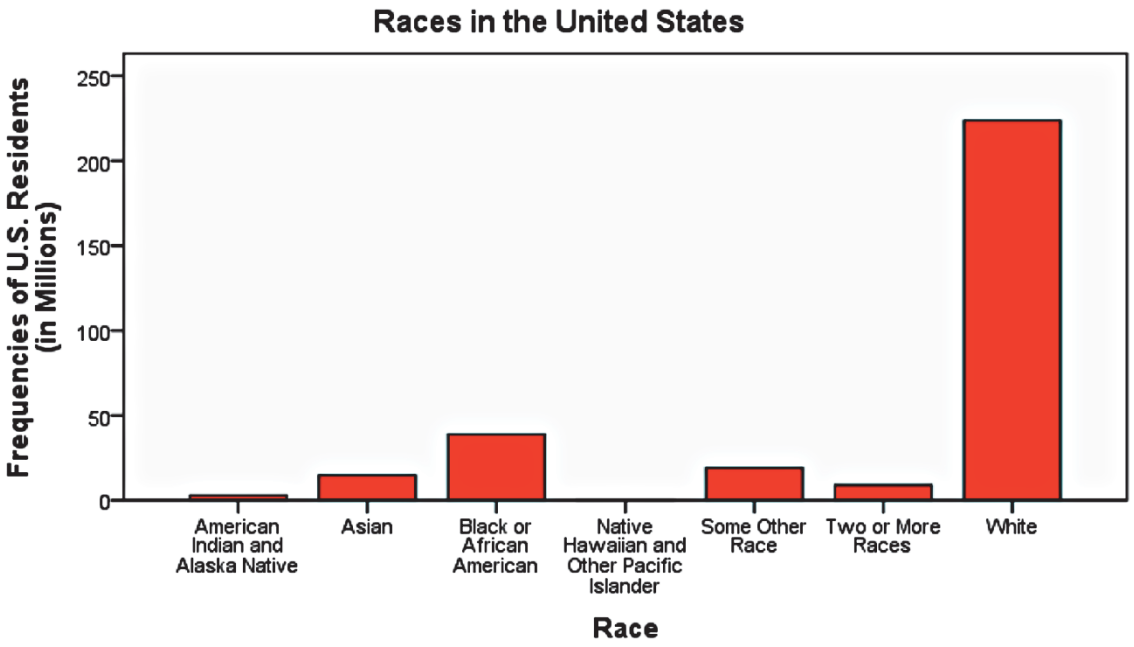
Race	U.S.	California
White	223.6	21.5
Black or African American	38.9	2.3
American Indian and Alaska Native	2.9	0.4
Asian	14.7	4.9
Native Hawaiian and Other Pacific Islander	0.5	0.1
Some Other Race	19.1	6.3
Two or More Races	9	1.8

a. Construct a relative frequency distribution for races in the U.S.

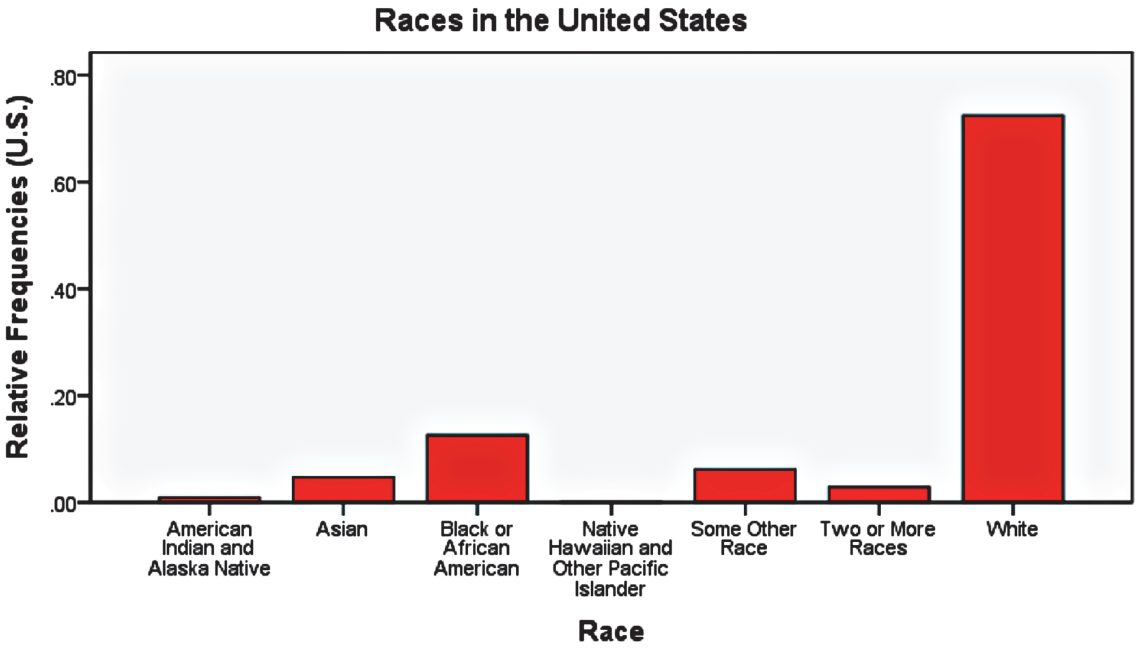
Race	Relative Frequency (U.S.)
White	0.7241
Black or African American	0.1261
American Indian and Alaska Native	0.0095
Asian	0.0475
Native Hawaiian and Other Pacific Islander	0.0017
Some Other Race	0.0619
Two or More Races	0.0292

b. What percentage of U.S. residents claims two or more races? 2.92%
c. What percentage of U.S. residents is not Asian? 95.25%

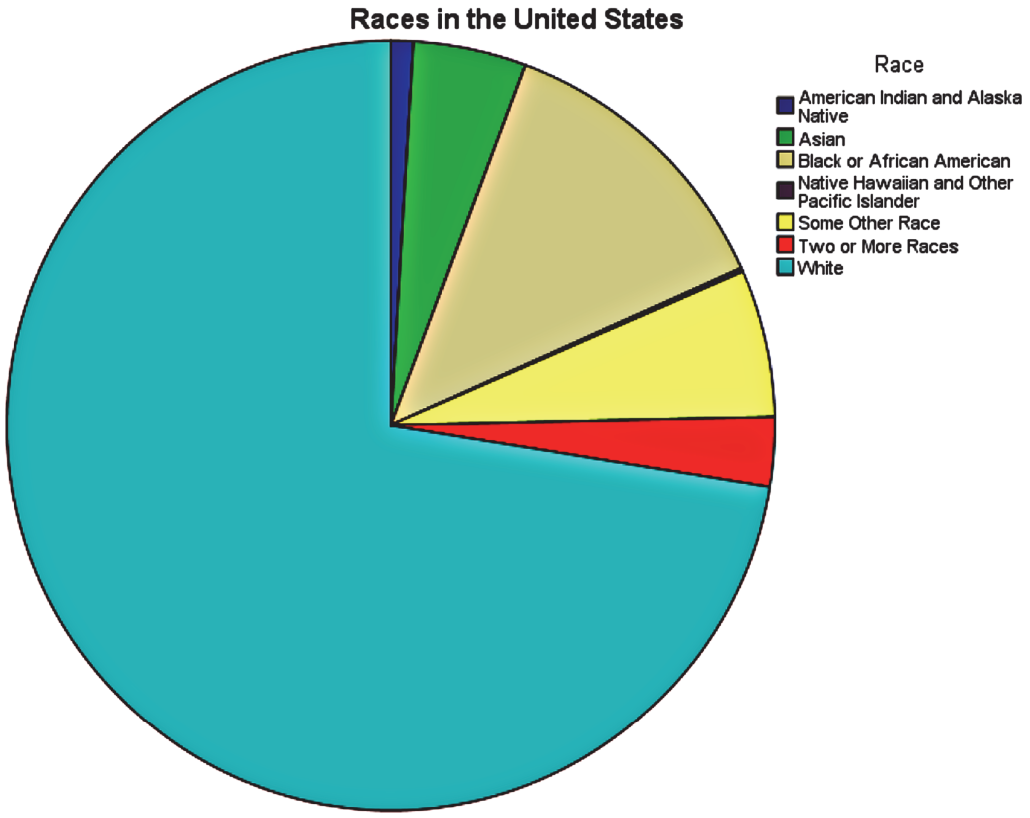
d. Construct a frequency bar graph for races in the U.S.



e. Construct a relative frequency bar graph for races in the U.S.



f. Construct a pie chart for races in the U.S.

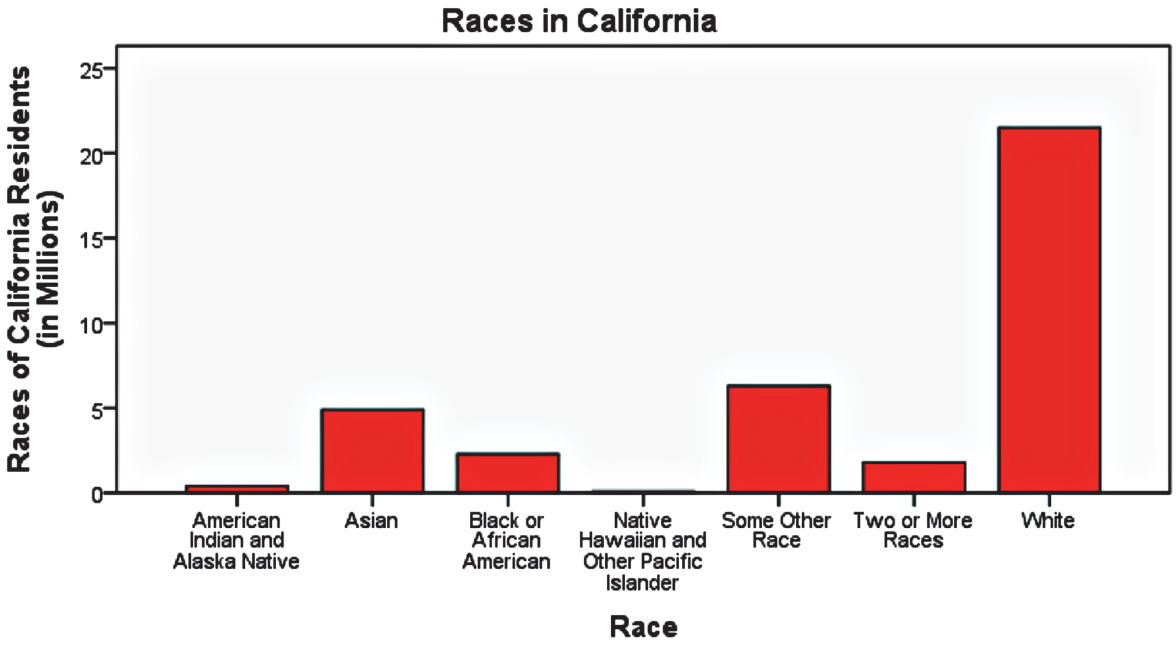


2. Use the data in Problem 1 to answer this question.
a. Construct a relative frequency distribution for the races in California.

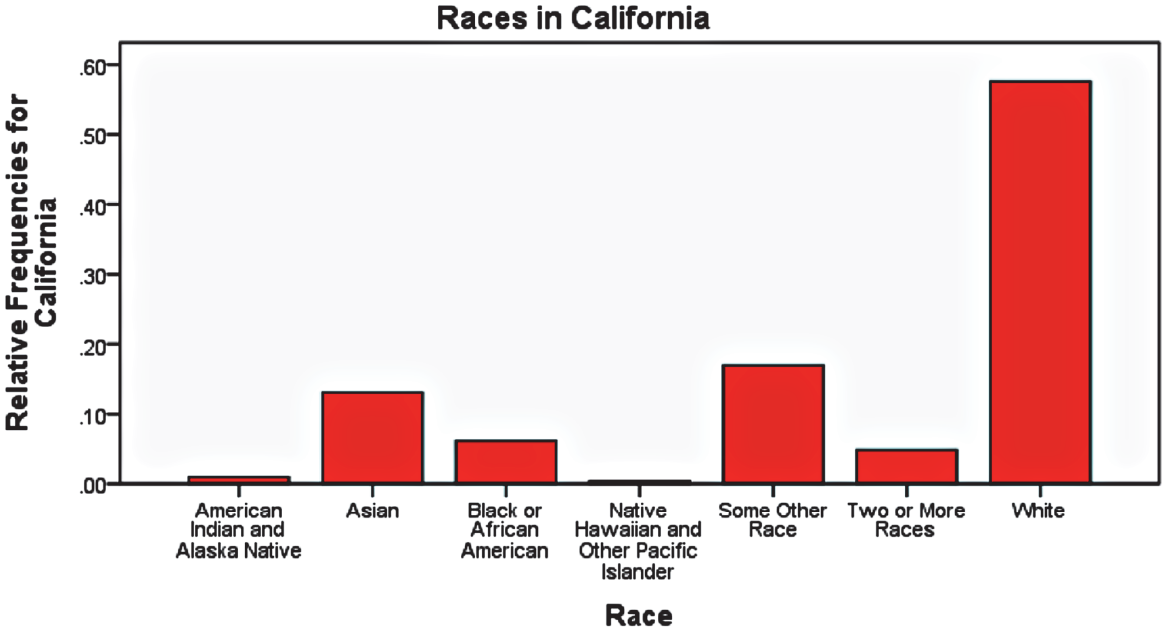
Race	Relative Frequency (California)
White	0.5759
Black or African American	0.0617
American Indian and Alaska Native	0.0097
Asian	0.1305
Native Hawaiian and Other Pacific Islander	0.0039
Some Other Race	0.1696
Two or More Races	0.0487

b. What percentage of California residents claims two or more races? 4.87%
c. What percentage of U.S. residents is not Asian? 86.95%

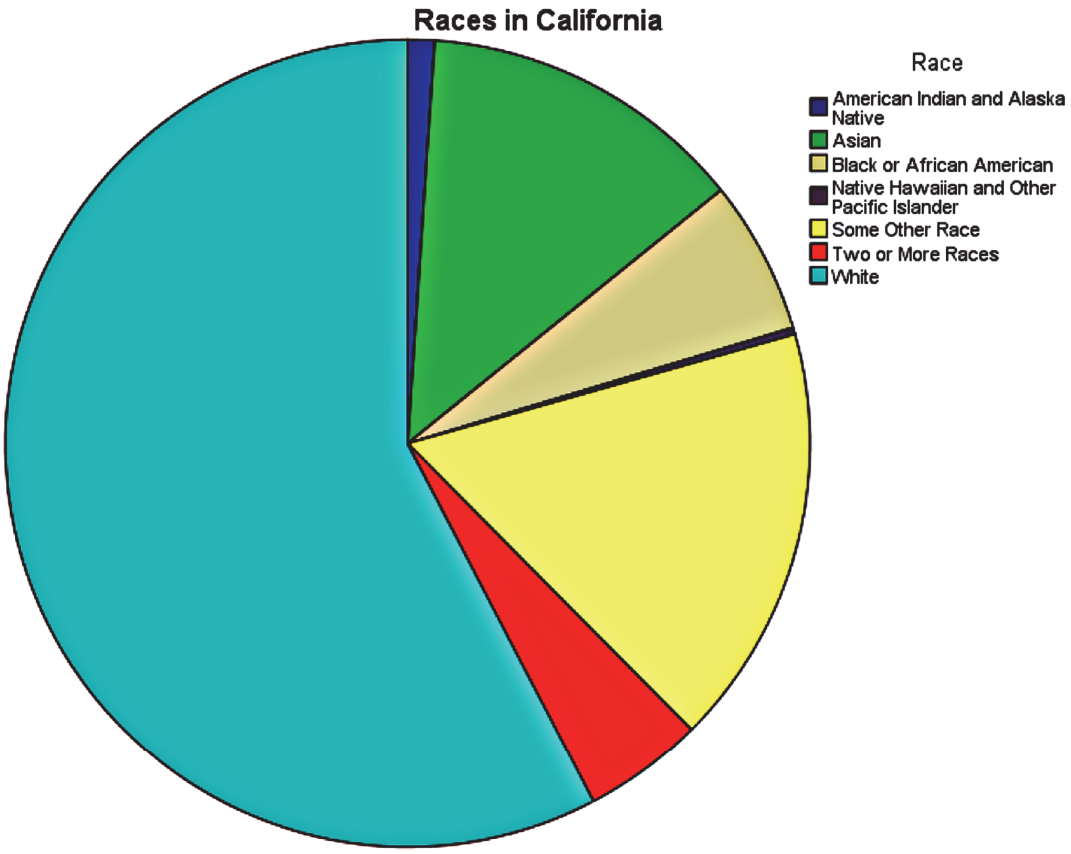
d. Construct a frequency bar graph for California.



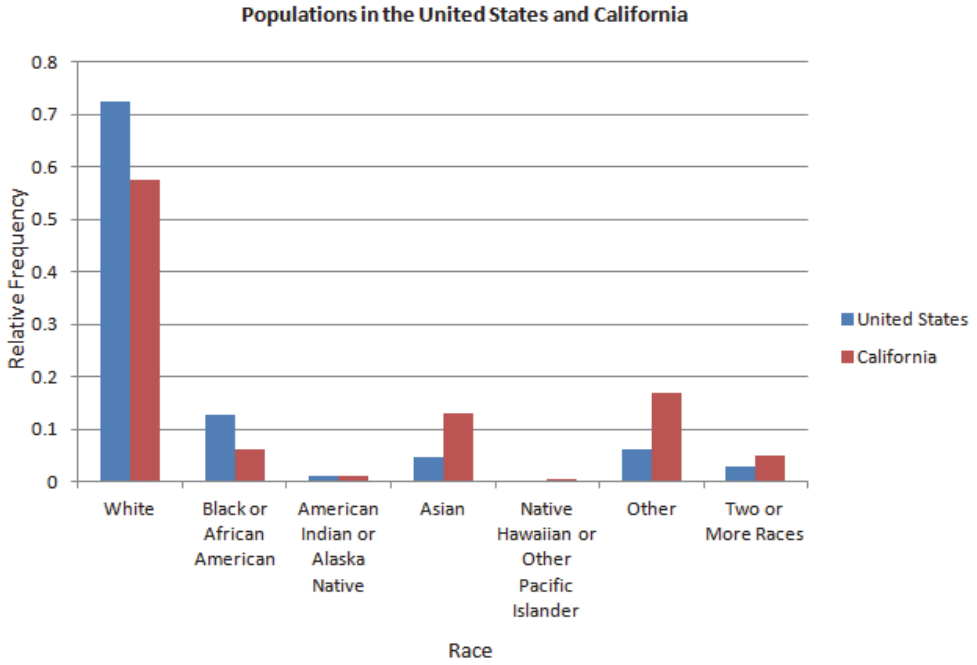
e. Construct a relative frequency bar graph for California.



f. Construct a pie chart for California.



3. Draw a side-by-side relative frequency bar graph of the populations of the United States and California. What do you notice?



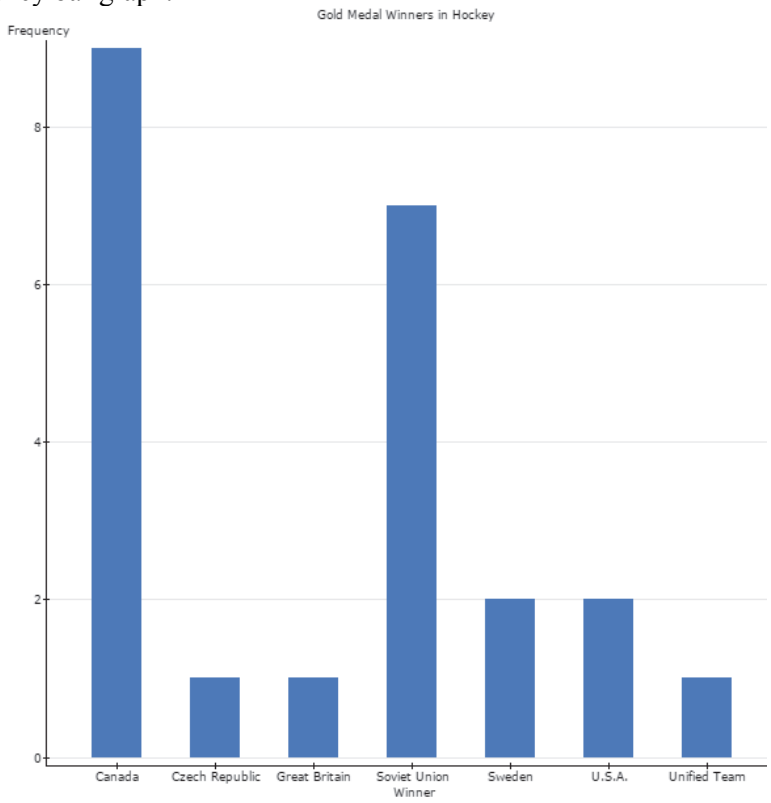
4. The table shows the gold medal winners in hockey in the Winter Olympics since 1920.

Year	Winner		Year	Winner		Year	Winner
1920	Canada		1960	U.S.A.		1992	Unified Team
1924	Canada		1964	Soviet Union		1994	Czech Republic
1928	Canada		1968	Soviet Union		1998	Sweden
1932	Canada		1972	Soviet Union		2002	Canada
1936	Great Britain		1976	Soviet Union		2006	Sweden
1948	Canada		1980	U.S.A.		2010	Canada
1952	Canada		1984	Soviet Union		2014	Canada
1956	Soviet Union		1988	Soviet Union			

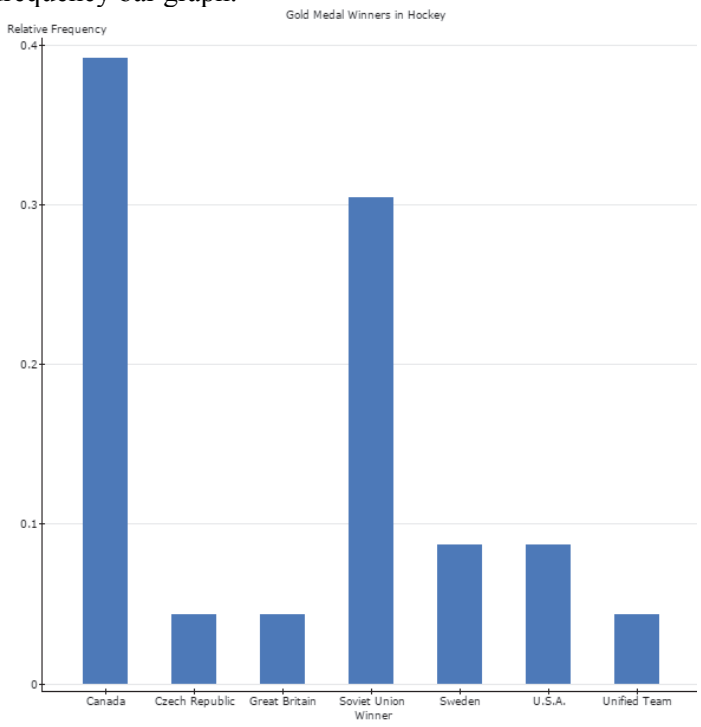
a. Construct a frequency distribution.
b. Construct a relative frequency distribution.

Winner	Frequency	Relative Frequency
Canada	9	0.39130435
Czech Republic	1	0.043478261
Great Britain	1	0.043478261
Soviet Union	7	0.30434783
Sweden	2	0.086956522
U.S.A.	2	0.086956522
Unified Team	1	0.043478261

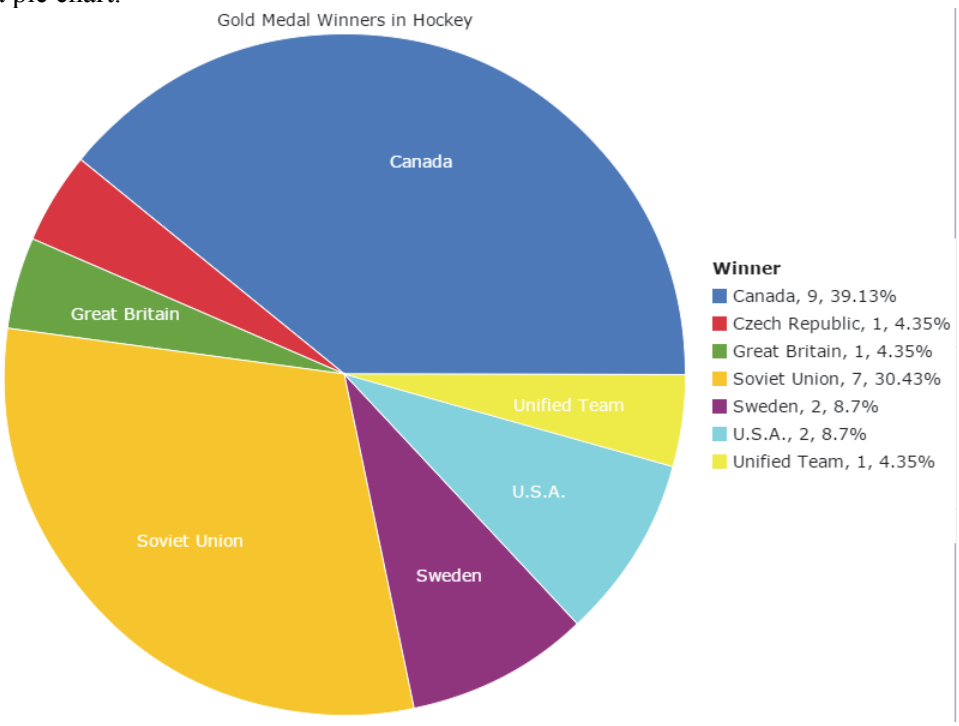
c. Construct a frequency bar graph.



d. Construct a relative frequency bar graph.



e. Construct a pie chart.



5. In-Class Activity (Requires one bag of multicolored candy, such as M&M's or Skittles for each student.) Ask the students bring one bag of candy, such as M&M's or Skittles to class. Have them create a frequency distribution of the counts of each color of candy in their bag. Have the students create a bar graph, relative frequency bar graph, and a pie chart illustrating the number of each type they observed.

Section 2.2

1. The following data represent the number of live births in the United States in 2012 for women 15 to 49 years of age.

Age	Live Births
15-19	10,908
20-24	916,811
25-29	1,123,900
30-34	1,013,416
35-39	472,318
40-44	109,579
45-49	7157
50-54	600

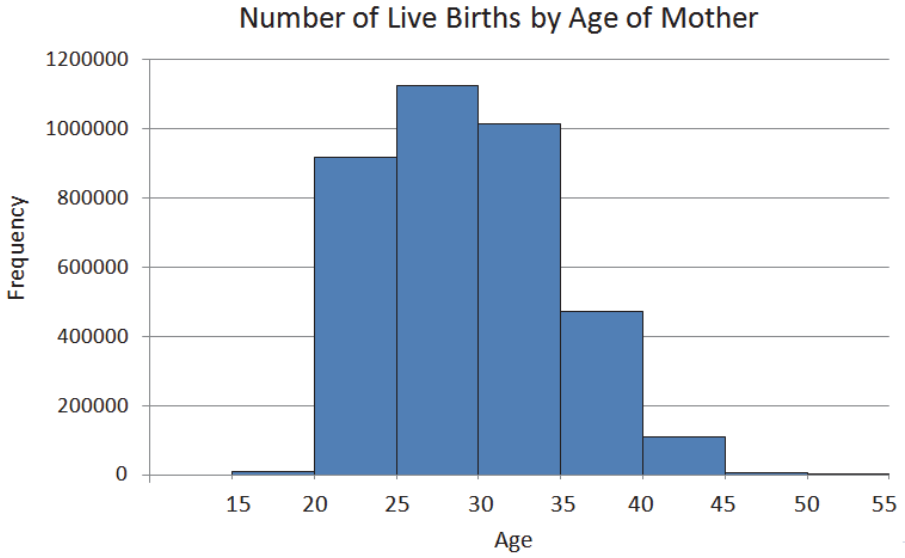
Source: National Vital Statistics Report

a. Determine the number of classes. 8
b. Determine the class limits. 15-19;20-24;25-29;30-34;35-39;40-44;45-49;50-54
c. Determine the class width. 5

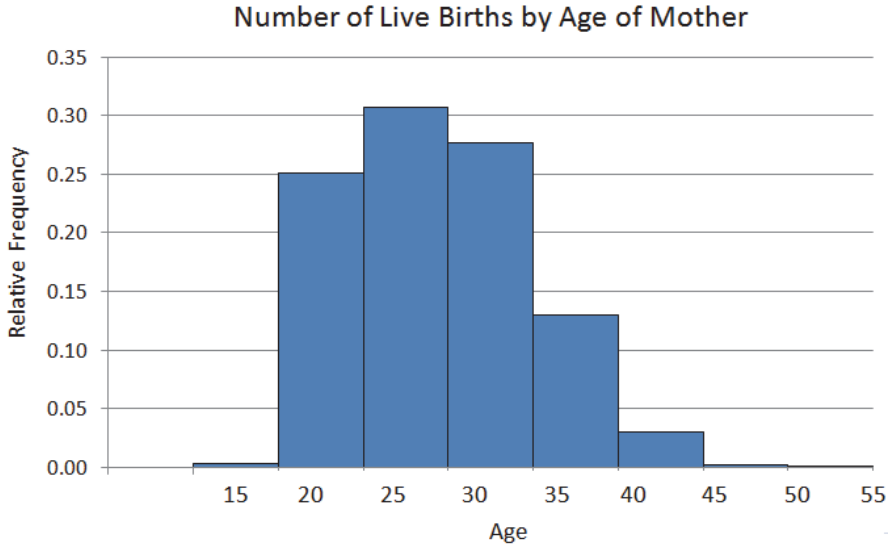
d. Construct a relative frequency distribution.

Age	Live Births
15-19	0.0030
20-24	0.2509
25-29	0.3075
30-34	0.2773
35-39	0.1292
40-44	0.0300
45-49	0.0020
50-54	0.0002

e. Construct a frequency histogram.



f. Construct a relative frequency histogram.



g. What percent of live births were to mothers 50-54 years of age? 0.02%

2. The following data represent the graduation rate for a random sample of 60 colleges and universities in the United States. Data from www.payscale.com.

86	59	65	55	59	40
90	48	37	65	67	67
39	53	82	57	71	46
83	77	40	52	92	38
56	57	36	61	34	54
69	35	73	29	92	39
24	48	41	46	79	41
44	43	28	61	49	65
48	42	72	35	58	39
80	75	44	52	52	47

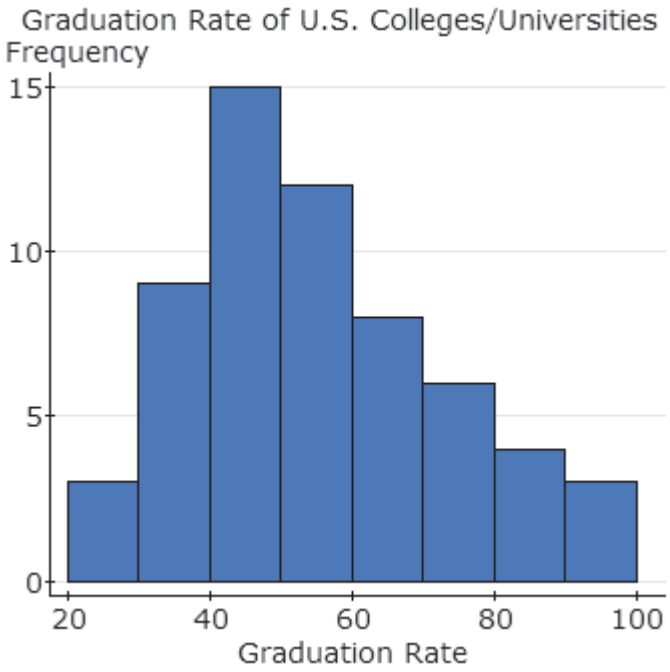
With a first class having lower class limit of 20 and a class width of 10:
a. Construct a frequency distribution.

Graduation Rate	Frequency
20 - 29	3
30 - 39	9
40 - 49	15
50 - 59	12
60 - 69	8
70 - 79	6
80 - 89	4
90 - 99	3

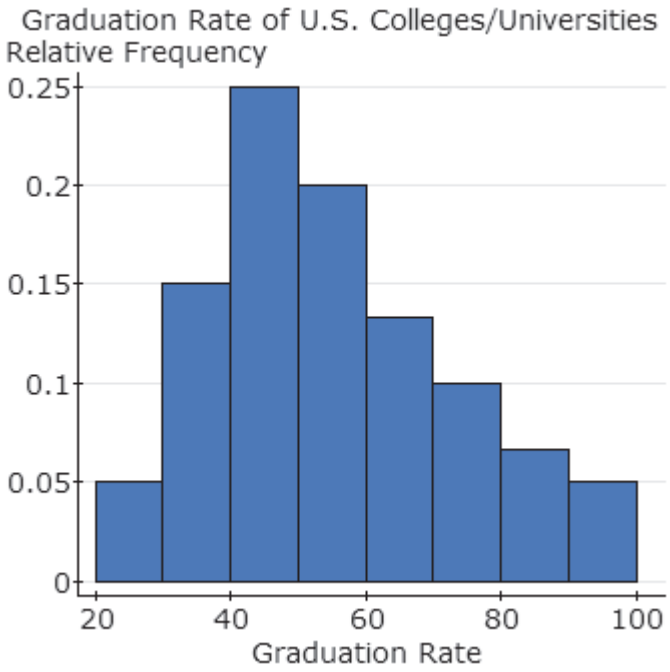
b. Construct a relative frequency distribution.

Graduation Rate	Relative Frequency
20 - 29	0.05
30 - 39	0.15
40 - 49	0.25
50 - 59	0.2
60 - 69	0.1333
70 - 79	0.1
80 - 89	0.0667
90 - 99	0.05

c. Construct a frequency histogram of the data.



d. Construct a relative frequency histogram of the data.



e. Describe the shape of the distribution. The distribution is skewed right.

2. Draw a stem-and-leaf plot of the data from Problem 1.

```
2 : 489
3 : 455678999
4 : 001123446678889
5 : 222345677899
6 : 11555779
7 : 123579
8 : 0236
9 : 022
```

Legend 2|4 represents 24

3. Go to http://en.wikipedia.org/wiki/United_States_congressional_apportionment. The site includes data on the number of representatives in the House of Representatives for each census year. Create a frequency and relative frequency histogram of the data for the latest census. Draw a dot plot of the data. Note: The data can be easily extracted using StatCrunch This! in StatCrunch.

4. Ask the students to compute their heights (in inches). Randomly select some students to report their heights. Create a histogram of the students' heights. It is a good idea to make a histogram for men and a histogram for women. Ask the students to identify the shape of the distribution.

Section 2.3

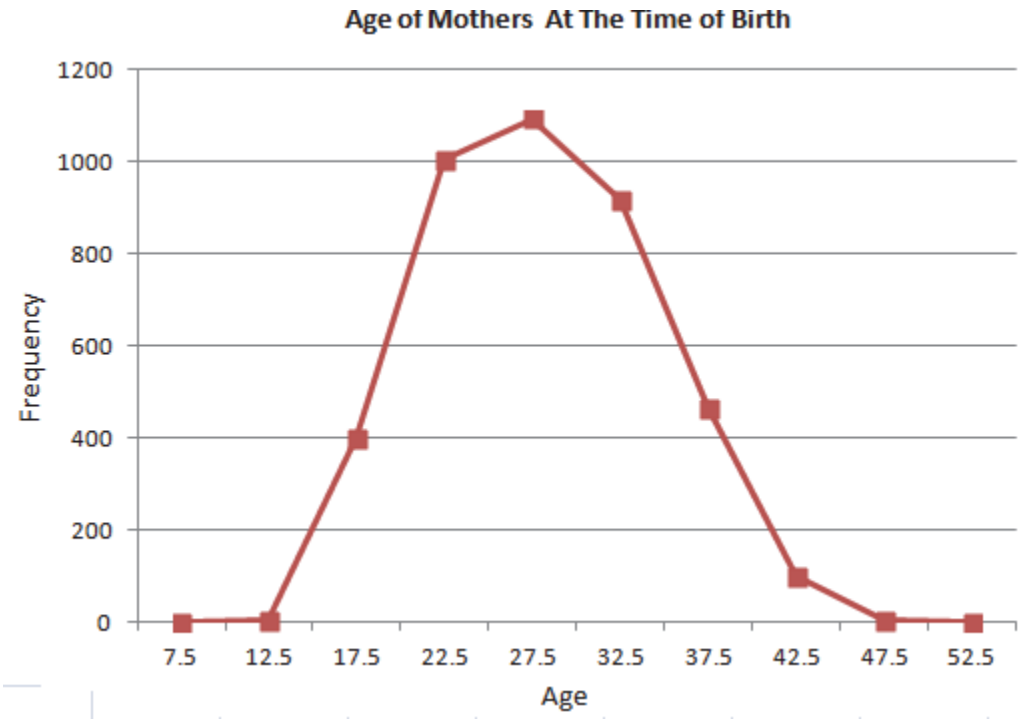
1. A simple random sample of 4,000 women who gave birth was collected. The following table summarizes the mother's ages at the time they gave birth. (Based on data at: www.infoplease.com/ipa/A0005074.html)

Age of Mother	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Frequency	9	401	1006	1094	919	467	101	6

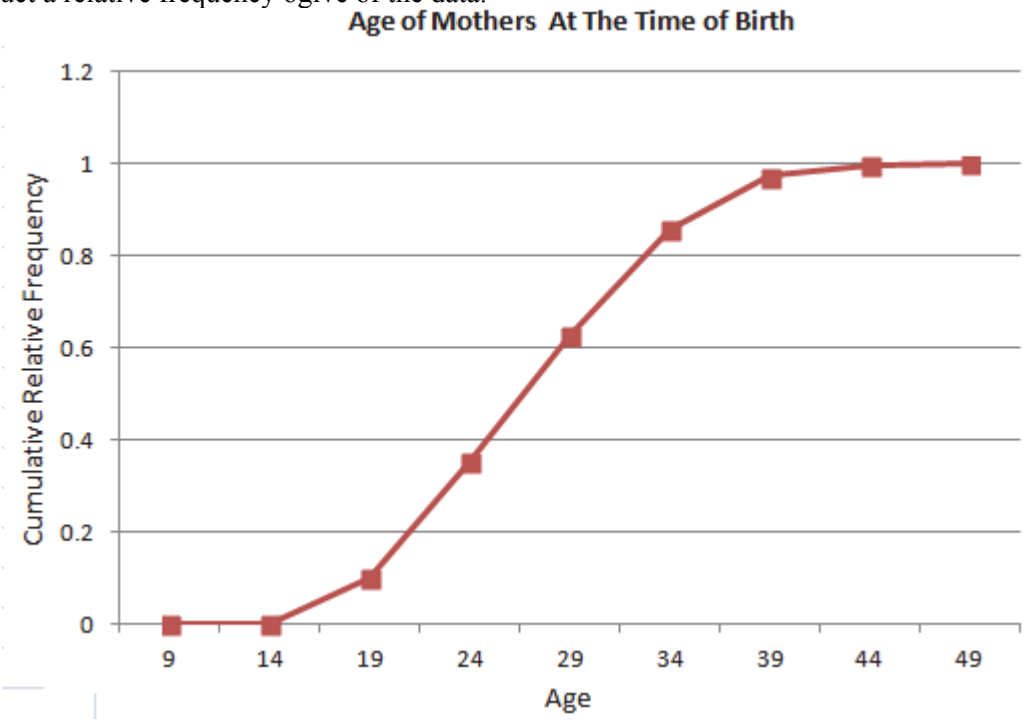
a. Create a cumulative frequency table summarizing the data. See below
b. Create a relative frequency table of the data. See below
c. Create a cumulative relative frequency table for the data. Note: Answers may vary slightly due to rounding.

Age of Mother	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
10-14	6	6	0.0015	0.0015
15-19	401	407	0.1003	0.1018
20-24	1006	1413	0.2515	0.3533
25-29	1094	2507	0.2735	0.6268
30-34	919	3426	0.2298	0.8565
35-39	467	3893	0.1168	0.9733
40-44	101	3994	0.0253	0.9985
45-54	6	4000	0.0015	1.0000

d. Construct a frequency polygon of the data.



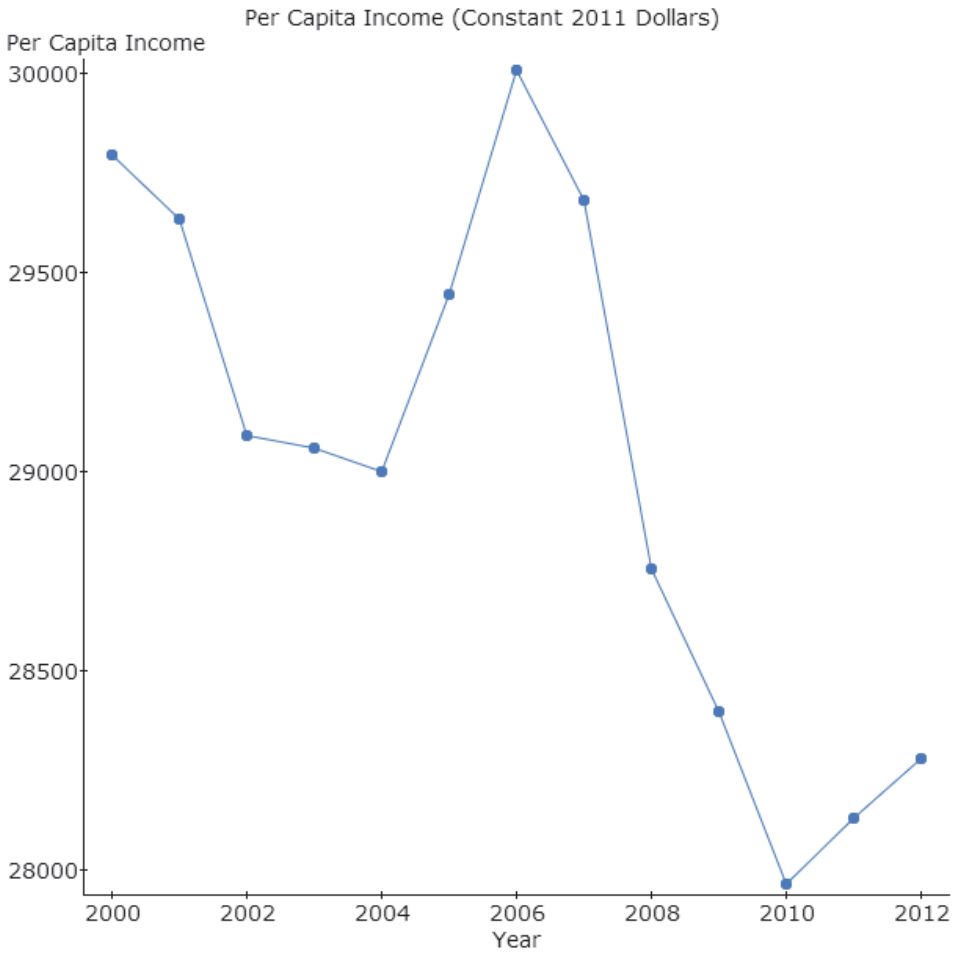
e. Construct a relative frequency ogive of the data.



2. The following data represent the per capita income in the United States from 2000 - 2012 in constant 2011 dollars (that is, adjusted for inflation). Draw a time series graph of the data.

Year	Per Capita Income (2011 Dollars)
2000	29,795
2001	29,636
2002	29,092
2003	29,058
2004	29,000
2005	29,446
2006	30,010
2007	29,682
2008	28,755
2009	28,400
2010	27,968
2011	28,130
2012	28,281

Source: United States Census Bureau

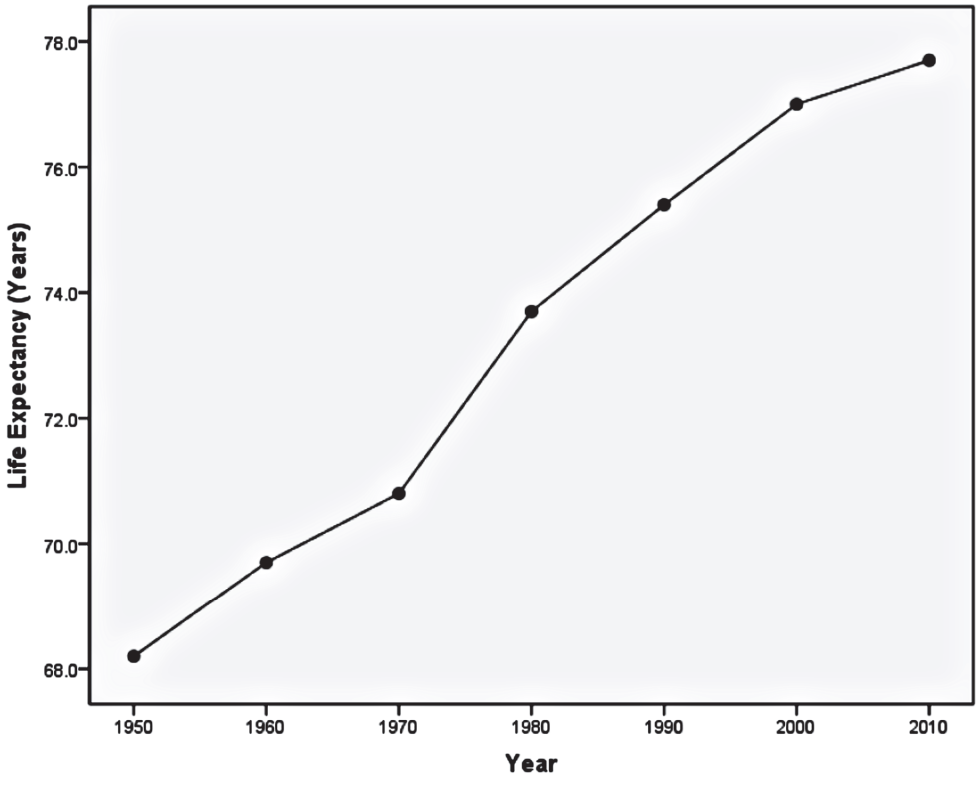


Section 2.4

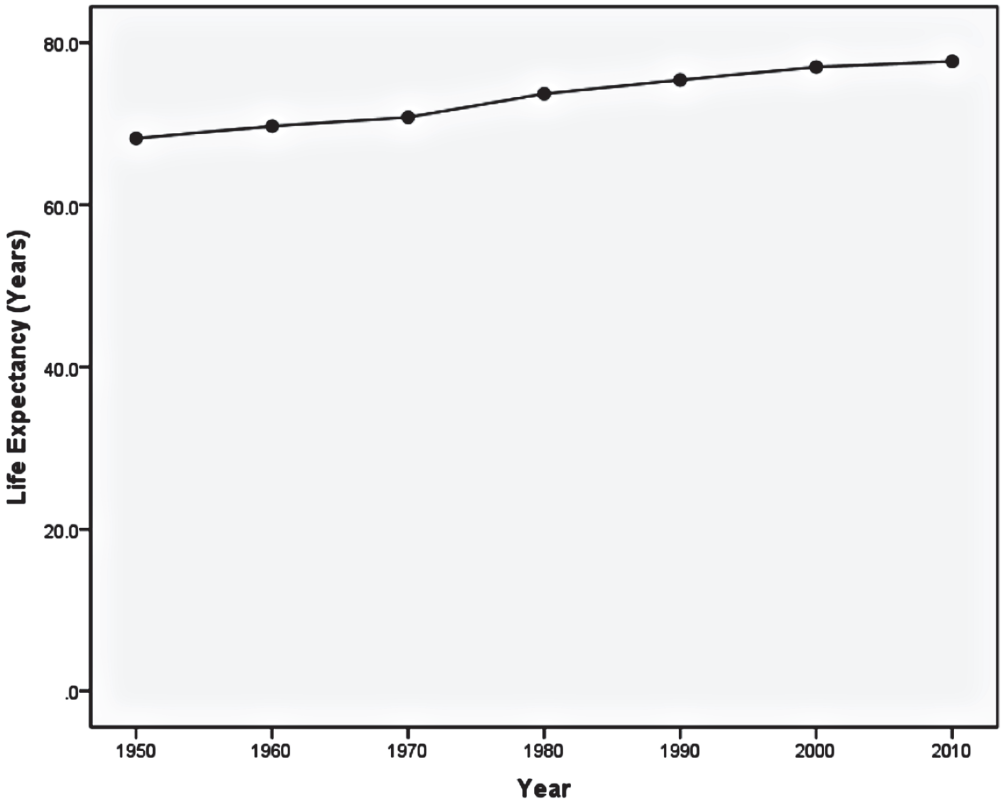
1. The data in the table below represent the historical life expectancies (in years) of residents of the United States. (Source: National Center for Health Statistics, *National Vital Statistics Reports,* www.cdc.gov/nchs and http://www.cdc.gov/nchs/data/hus/hus10.pdf#022)

Year, x	Life Expectancy, y
1950	68.2
1960	69.7
1970	70.8
1980	73.7
1990	75.4
2000	77.0
2010	77.7

a. Construct a misleading time-series plot that indicates that the life expectancy has risen sharply over time.



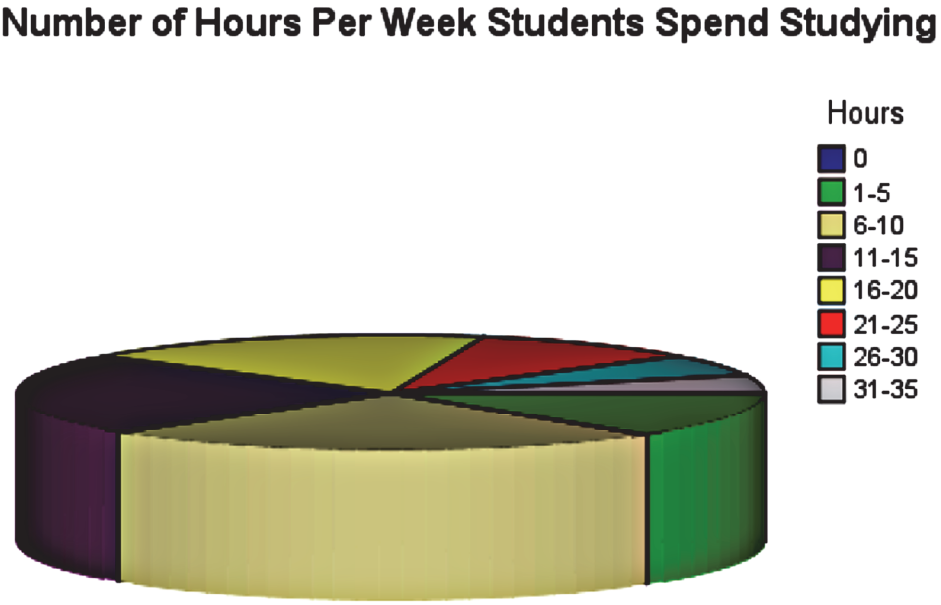
b. Construct a time-series plot that is not misleading.



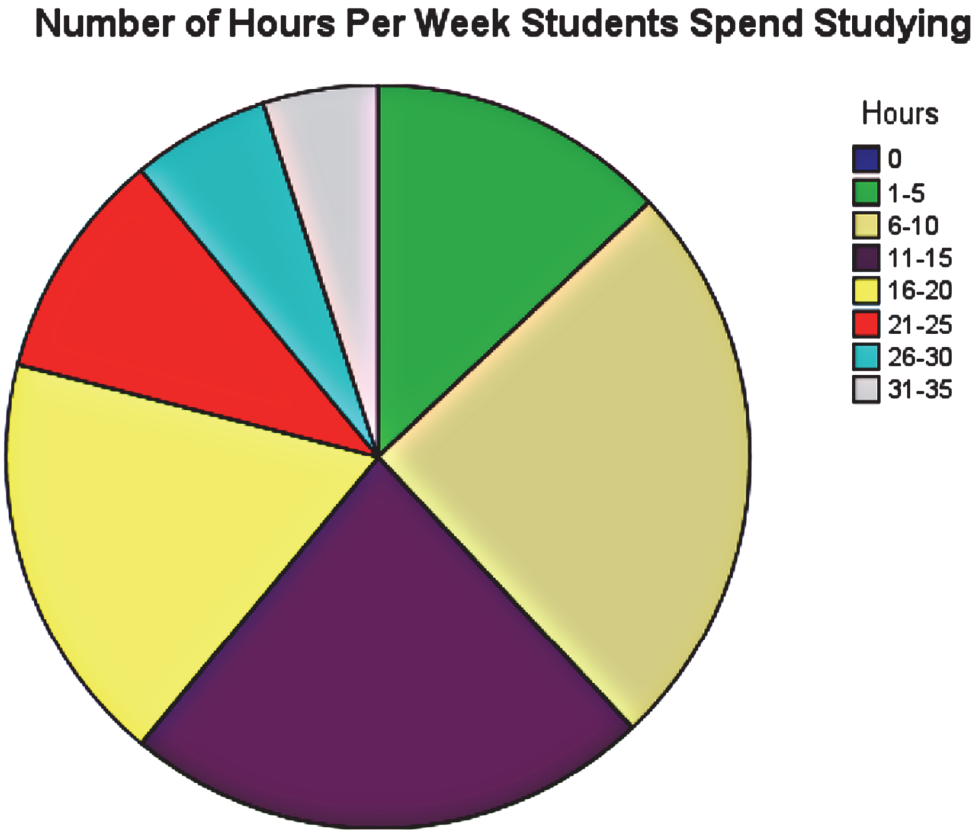
2. The National Survey of Student Engagement is a survey that (among other things) asks first year students at liberal arts colleges how much time they spend preparing for class each week. The results are summarized below. (Source: NSSE)

Hours	0	1-5	6-10	11-15	16-20	21-25	26-30	31-35
Percentage of 1st year students	0%	13%	25%	23%	18%	10%	6%	5%

a. Construct a pie chart that exaggerates the percentage of students who spend between 6 and 10 hours preparing for class each week.



b. Construct a pie chart that is not misleading.



Note: If you use Excel to create 3-D pie charts, you can rotate the graph so students may visualize the distortion created by the three-dimensional art.