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)** $\frac{128,000}{1,350,000} = \frac{64}{675} \approx 0.0948$; approximately 9.48 % of cosmetic surgeries in 2009 were for tummy tucks.

(b) $\frac{138,000}{1,350,000} = \frac{23}{225} \approx 0.102$; 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 largest bar corresponds to China, so China had the most internet users in 2010.

(b) The bar for the United Kingdom appears to reach the line for 50. Thus, we estimate that there were 50 million internet users in the United Kingdom in 2007.

(c) The bar for China appears to reach 420 on the vertical axis. The bar for Germany appears to reach 70. Since, 420-70=350, we estimate that there were about 350 million more internet users in China than in Germany during 2007.

(d) This graph should use relative frequencies, rather than frequencies.

8. **(a)** 29,830,000 whites were living in poverty.

(b) 12350/(29830+9944+12350+1746) = 0.229 = 22.9%
In 2009, 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 2008 involved credit card fraud. So,
10 million * 0.26 = 2.6 million cases of credit card fraud occurred in 2008.

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.

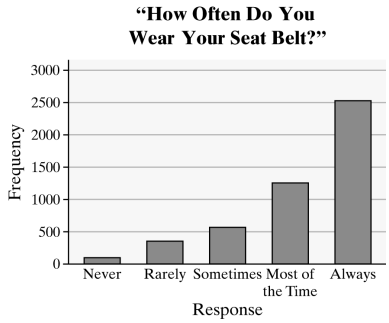
13. **(a)** Total students surveyed = 125 + 324 + 552 + 1257 + 2518 = 4776
Relative frequency of "Never"
$= 125/4776 \approx 0.0262$ and so on.

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

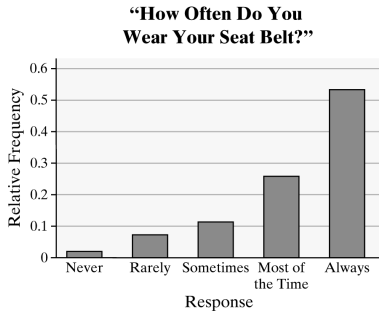
(b) 52.72%

(c) $0.0262 + 0.0678 = 0.0940$ or 9.40%

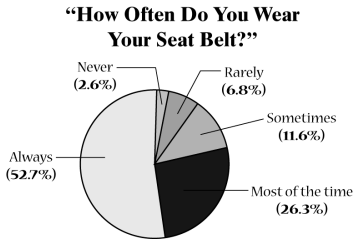
(d)



(e)



(f)



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

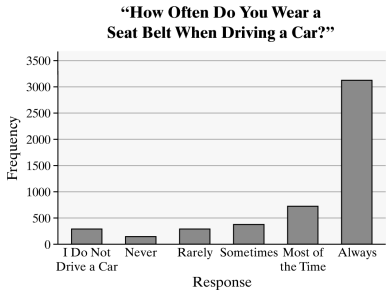
14. **(a)** Total students surveyed = 249 + 118 + 249 + 345 + 716 + 3093 = 4770
Relative frequency of " I do not drive"
$= \frac{249}{4770} \approx 0.0522$ and so on.

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%

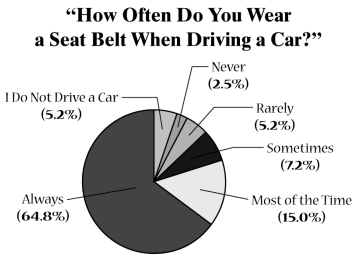
(c) $0.0247 + 0.0522 = 0.0769$ or 7.7%

(d)



(e)

(f)



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

Relative frequency of "Never"

$= \frac{118}{4521} \approx 0.0261$ 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 than when riding in a car driven by another.

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

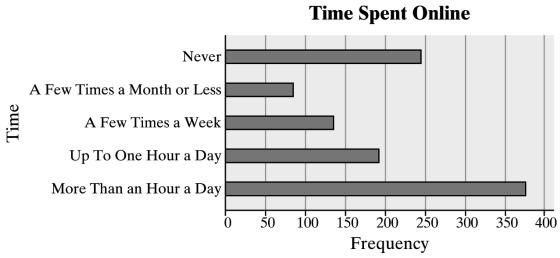
15. (a) Total adults surveyed = 377 + 192 + 132 + 81 + 243 = 1025

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

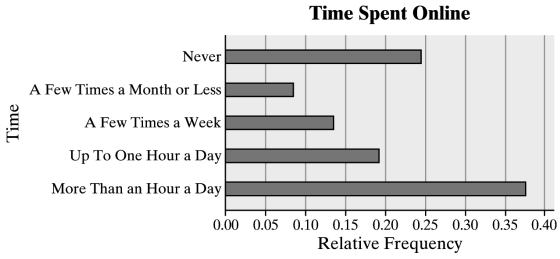
Response	Relative Frequency
More than 1 hr a day	0.3678
Up to 1 hr a day	0.1873
A few times a week	0.1288
A few times a month or less	0.0790
Never	0.2371

(b) 0.2371 (about 24%)

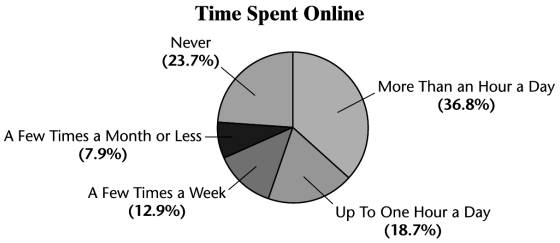
(c)



(d)



(e)



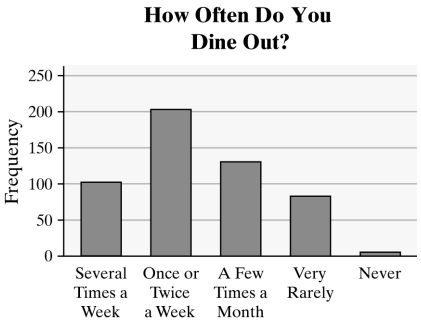
(f) The statement provides an estimate, but no level of confidence is given.

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.

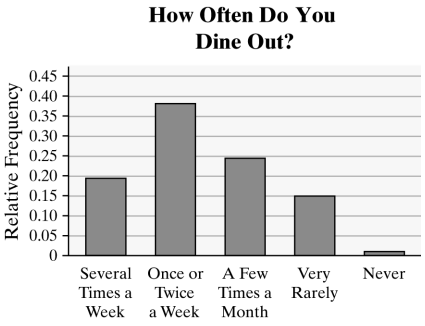
Response	Relative Frequency
Several times a week	0.197
Once or twice a week	0.392
A few times a month	0.250
Vary rarely	0.152
Never	0.010

(b) The proportion surveyed who dine out once or twice a week is 204/(103+204+130+79+5)=0.396

(c)



(d)



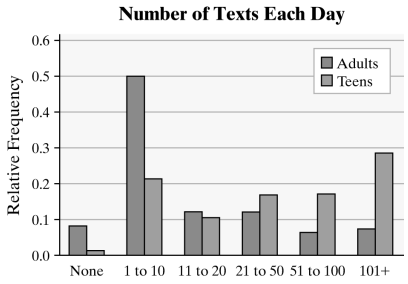
17. **(a)** Total adults = 1936
Relative frequency for "none" is: 173/1936=0.09, and so on.

Number of Texts	Rel. Freq. (Adults)
None	0.09
1 to 10	0.51
11 to 20	0.13
21 to 50	0.13
51 to 100	0.07
101+	0.08

(b) Total teens = 627
Relative frequency for "none" is: 13/627=0.02, and so on.

Number of Texts	Rel. Freq. (Teens)
None	0.02
1 to 10	0.22
11 to 20	0.11
21 to 50	0.18
51 to 100	0.18
101+	0.29

(c)



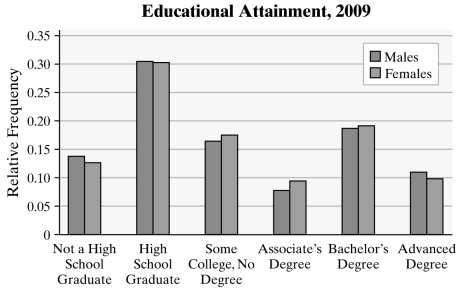
(d) Answers will vary. Adults are much more likely to do fewer texts per day, while teens are much more likely to do more texting.

18. **(a), (b)**
Total males = 94.5 million
Relative frequency for "Not HS graduate" is 13.3/94.5 = 0.141 and so on.

Total females = 102.0 million
Relative frequency for "Not HS graduate" is 13/100.5 = 0.130 and so on.

Educational Attainment	Males	Females
Not a HS graduate	0.141	0.127
High school graduate	0.312	0.311
Some college, no degree	0.167	0.176
Associate's degree	0.079	0.096
Bachelor's degree	0.190	0.192
Advanced degree	0.110	0.097

(c)



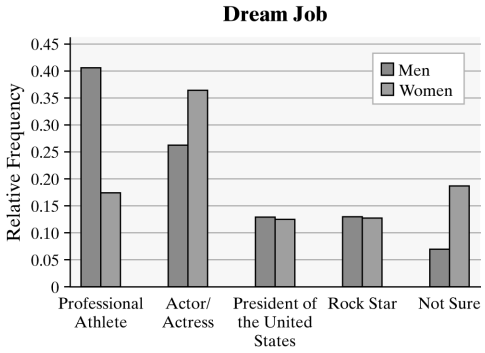
(d) Answers will vary. It appears that females are slightly more likely to start, but not finish college. Males appear to be slightly more likely to attain an advanced degree.

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 U.S.	0.131	0.130
Rock Star	0.131	0.130
Not Sure	0.071	0.190

(b)



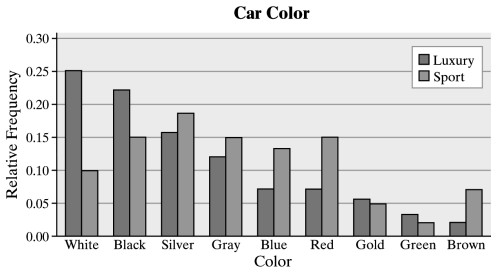
(c) Answers will vary. Males are much more likely to want to be a professional athlete. Women are more likely to aspire to a career in acting than men. Men's desire to become athletes may be influenced by the prominence of male sporting figures in popular culture. Women may aspire to careers in acting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

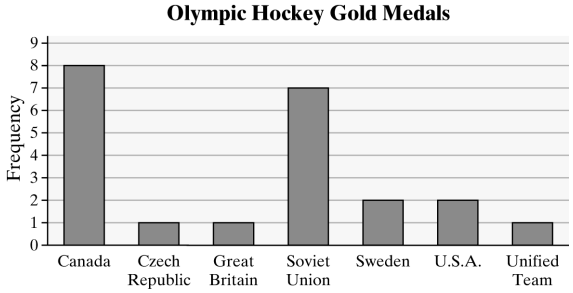


(c) Answers will vary. White is the most popular color for luxury cars, while silver is the most popular for sports cars. People who drive luxury cars may enjoy the clean look of a white vehicle. People who drive sports cars may prefer the flashier look of silver.

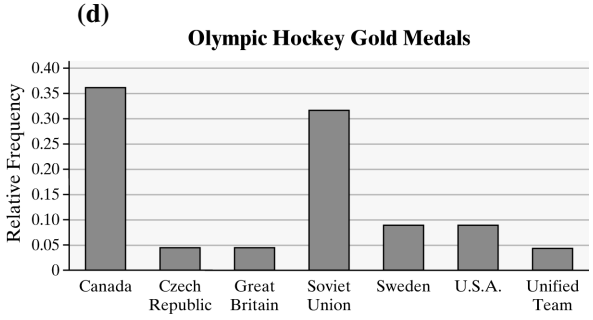
21. **(a), (b)**
Total number of Winter Olympics = 22; relative frequency for Canada is 8/22=0.364.

Winner	Freq.	Rel. Freq.
Canada	8	0.364
Czech Republic	1	0.045
Great Britain	1	0.045
Soviet Union	7	0.318
Sweden	2	0.091
U.S.A.	2	0.091
Unified Team	1	0.045

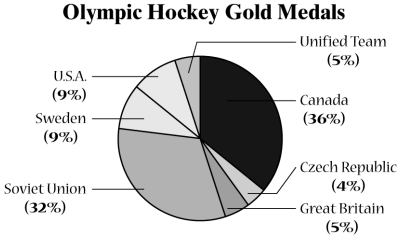
(c)



(d)



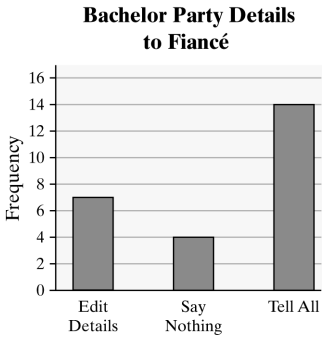
(e)



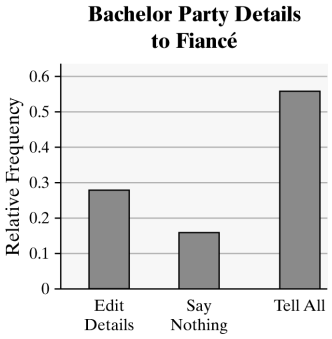
22. (a), (b)
Total number of responses = 25; relative frequency for "edit details" is 7/25=0.28.

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

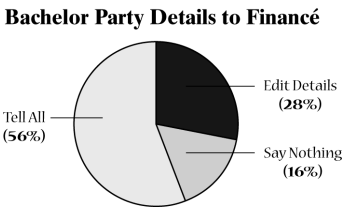
(c)



(d)



(e)

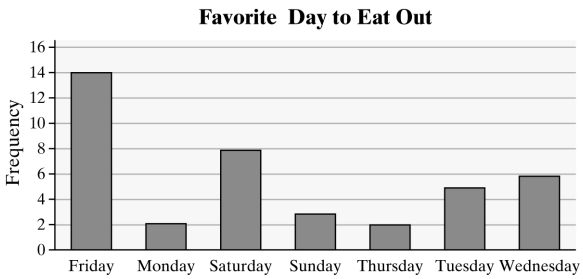


23. (a), (b)
Total number of responses = 40; relative frequency for "Sunday" is 3/40=0.075.

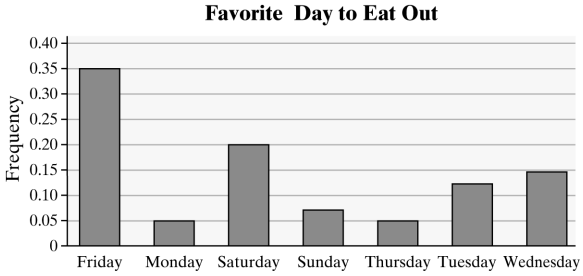
Response	Freq.	Rel. Freq.
Sunday	3	0.075
Monday	2	0.05
Tuesday	5	0.125
Wednesday	6	0.15
Thursday	2	0.05
Friday	14	0.35
Saturday	8	0.2

(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.

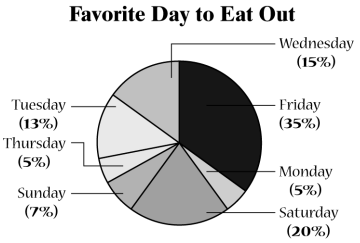
(d)



(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

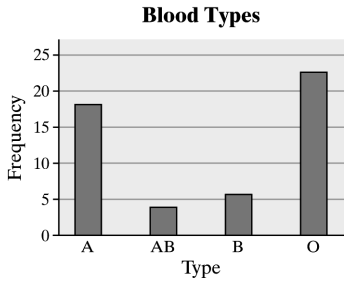
(c) Type O is the most common.

(d) Type AB is the least common.

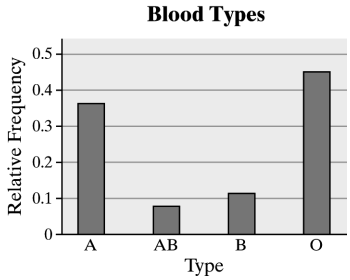
(e) We estimate that 44% of the population has type O blood. This is considered inferential statistics because a conclusion about the population is being drawn based on sample data.

(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.

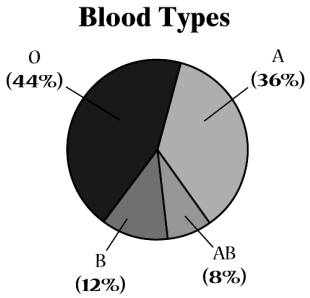
(g)



(h)



(i)



25. (a), (b)

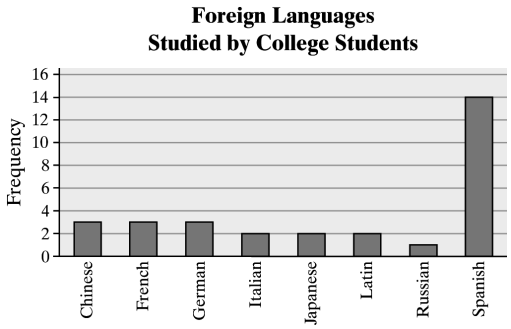
Total number of students = 30

Relative frequency for "Chinese"

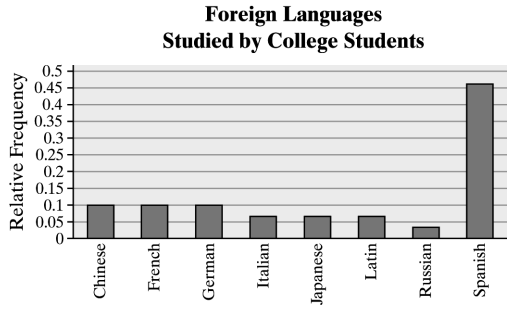
$= \frac{3}{30} = 0.100$ and so on.

Language	Freq.	Rel. Frequency
Chinese	3	0.100
French	3	0.100
German	3	0.100
Italian	2	0.067
Japanese	2	0.067
Latin	2	0.067
Russian	1	0.033
Spanish	14	0.467

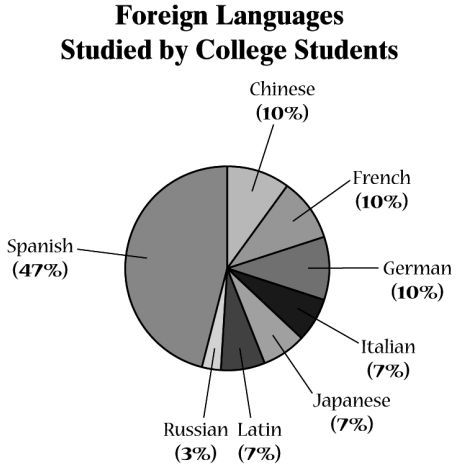
(c)



(d)



(e)



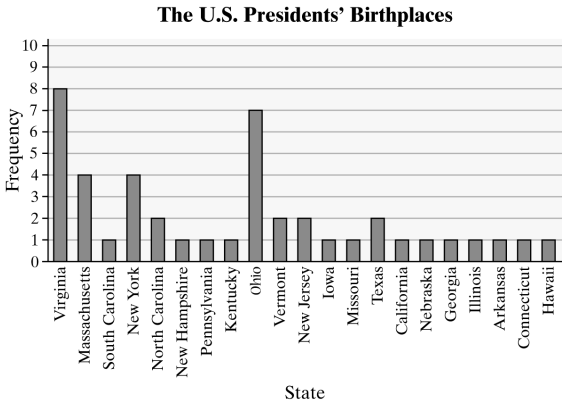
26. (a)

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

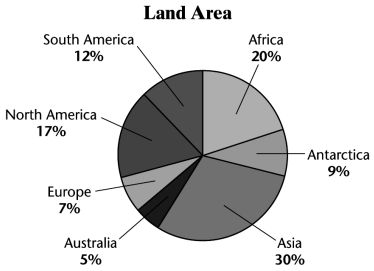


(b) More presidents were born in Virginia than in any other state.

(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 U.S. 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.

27. (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.

28. Answers will vary.

29. Answers will vary.

30. **(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.

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

32. 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.

33. 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.

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

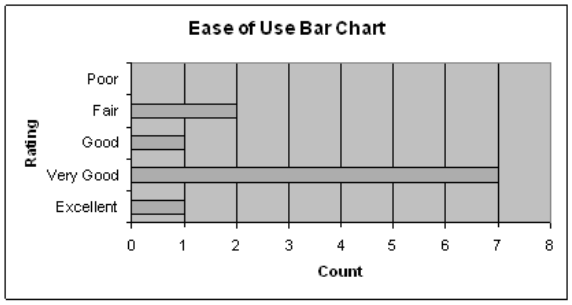
Consumer Reports®: Consumer Reports Rates Treadmills

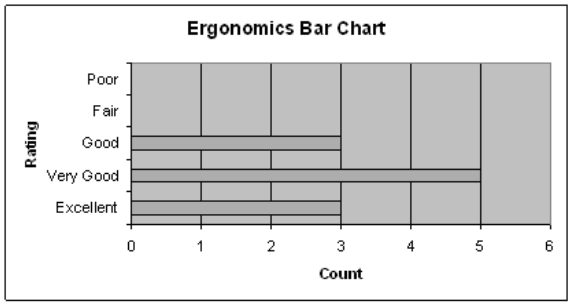
(a) A bar chart is used to display the overall scores. Because the bars are in decreasing order, this is an example of a Pareto chart.

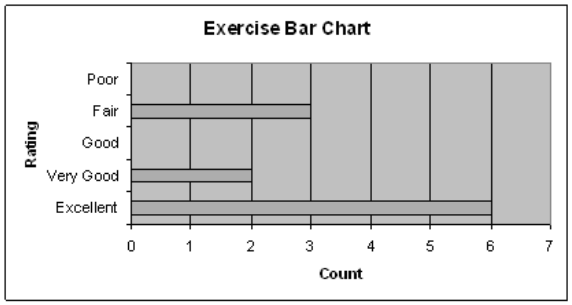
(b) The Precor M9.33 has the highest construction score since it was the only model receiving an excellent rating. Two models, the Tunturi J6F and the ProForm 525E received a Fair rating, making them the models with the lowest ease of use score.

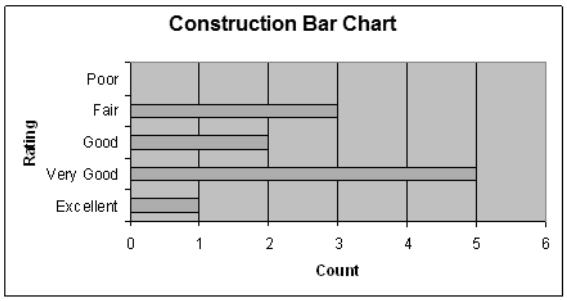
(c) 1 model was rated Excellent, 7 models were rated Very Good, 1 model was rated Good, and 2 models were rated Fair. No models were rated Poor for ease of use.

(d) The following bar charts were created in Microsoft® Excel:









(e) The following scatterplot was obtained by eyeballing the value of the scores from the Overall Score Pareto chart. Although there is a great deal of scatter in the data, even within a similar price range, there appears to be a relationship between score and price. The more expensive models tested by Consumer Reports in March 2002 tended to score higher in overall performance. (One should be cautious about generalizing the conclusions to the universe of treadmills since only a small sample of treadmills have been tested here.)

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

8. False

9. **(a)** 8

(b) 2

(c) 15

(d) $11 - 7 = 4$

(e) $\frac{15}{100} = 0.15$ or 15%

(f) The distribution is bell shaped.

10. **(a)** 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)** 200 (e.g. 200 – 0 = 200)

(b) 0-199, 200-399, 400-599, 600-799, 800-999, 1000-1199, 1200-1399, 1400-1599

(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)** 8 classes

(b) Lower class limits: 775, 800, 825, 850, 875, 900, 925, 950
Upper class limits: 799, 824, 849, 874, 899, 924, 949, 974

(c) The class width is found by subtracting consecutive lower class limits. For example, 800 –775 = 25. Therefore, the class width is 25(dollars).

22. **(a)** 8 classes

(b) Lower class limits: 0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0
Upper class limits: 0.9, 1.9, 2.9, 3.9, 4.9, 5.9, 6.9, 7.9, 8.9

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

23. **(a)** 7 classes

(b) Lower class limits: 15, 20, 25, 30, 35, 40, 45; Upper class limits: 19, 24, 29, 34, 39, 44, 49

(c) The class width is found by subtracting consecutive lower class limits. For example, $20 - 15 = 5$. Therefore, the class width is 5 (years).

24. **(a)** 6 classes

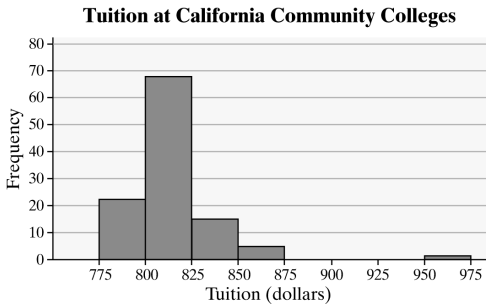
(b) Lower class limits: 0, 5000, 10,000, 15,000, 20,000, 25,000
Upper class limits: 4999, 9999, 19,999, 24,999, 29,999

(c) The class width is found by subtracting consecutive lower class limits. For example, $10,000 - 5000 = 5000$. Therefore, the class width is 5000 (students).

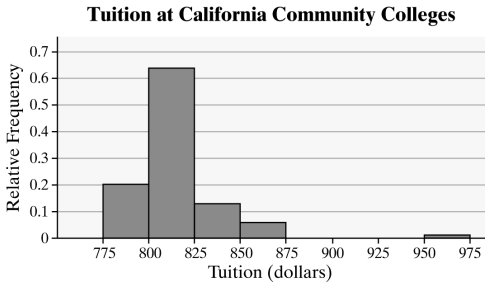
25. **(a)** Total frequency is:
22 + 68 + 15 + 5+ 0+ 0+ 0+ 1 = 111
Relative frequency for 775-779 is
22/111 = 0.1982 and so on.

Tuition ($)	Relative Frequency
775-799	0.1982
800-825	0.6126
825-849	0.1351
850-874	0.0450
875-899	0.0000
900-924	0.0000
925-949	0.0000
950-974	0.0090

(b)



(c)



Total number of California community colleges with tuition less than $800 is 22.

$\frac{22}{111} \cdot 100\% \approx 19.82\%$ of California community colleges had tuition of less than $800.

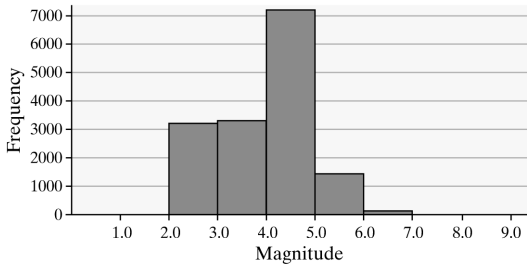
Total number of colleges with tuition of $850 or more = 5 + 1 = 6

$\frac{6}{111} \cdot 100\% \approx 5.41\%$ of California community colleges had tuition of $850 or more.

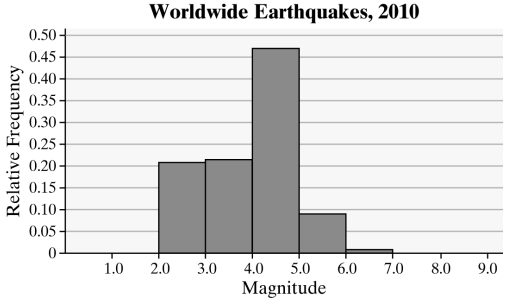
26. **(a)** Total number of earthquakes is:
22 + 22 + 3201 + 3332 + 7276 + 1430 + 130 + 18+ 1 = 15,342
Relative frequency for 0-0.9 is
22/15,342 = 0.0014 and so on.

Magnitude	Relative Frequency
0-0.9	0.0014
1.0-1.9	0.0014
2.0-2.9	0.2086
3.0-3.9	0.2172
4.0-4.9	0.4743
5.0-5.9	0.0932
6.0-6.9	0.0085
7.0-7.9	0.0012
8.0-8.9	0.0001

(b)



(c)



$\frac{7,276}{15,432} \cdot 100\% = 47.15\%$ of earthquakes registered 4.0-4.9 in magnitude, and

$\frac{22+22+3201+3332+7276}{15,432} \cdot 100\%$

$= \frac{13,853}{15,432} \cdot 100\% = 89.77\%$

of earthquakes in 2007 registered 4.9 or less in magnitude.

27. (a) Total births = 445,045+ 1,082,837+ 1,208,405+ 962,179+ 499,816+ 105,071+ 7,349 = 4,310,802
Relative frequency for 15-19 = 445,045/4,310,802 = 0.1032 and so on.

Age	Relative Frequency
15-19	0.1032
20-24	0.2512
25-29	0.2803
30-34	0.2232
35-39	0.1160
40-44	0.0244
45-49	0.0017

(b)

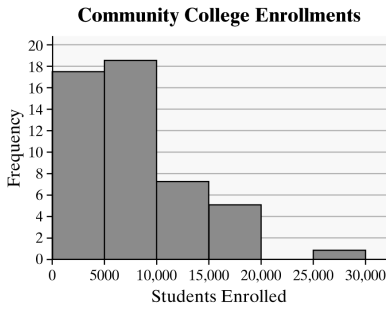
(c)

The relative frequency is 0.0244, so 2.44% of live births were to women 40-44 years of age.
$0.1032 + 0.2512 = 0.3544$ so 35.44% of live births were to women 24 years of age or younger.

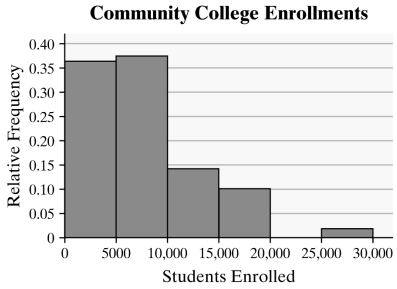
28. (a) Total frequency:
17 + 18 + 7 + 5 + 0 + 1 = 48
Relative frequency for 0-4999 is 17/45 = 0.3542 and so on.

Enrollment	Relative Frequency
0-4999	0.3542
5000-9999	0.3750
10,000-14,999	0.1458
15,000-19,999	0.1042
20,000-24,999	0.0000
25,000-29,999	0.0208

(b)



(c)



The relative frequency is 0.3750, so 37.50% of public Illinois community colleges enrolled between 5000 and 9999 students.

$0.1042 + 0 + 0.0208 = 0.1250$ so 12.50% of public Illinois community colleges enrolled 15,000 or more students.

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

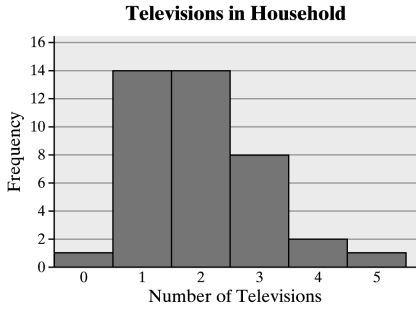
(b), **(c)**
The relative frequency for 0 color televisions is 1/40 = 0.025, and so on.

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

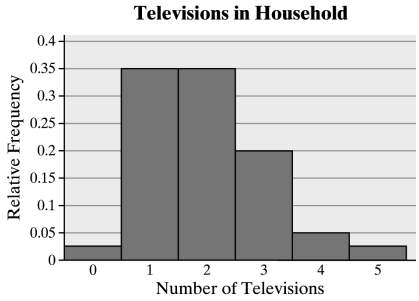
(d) The relative frequency is 0.2 so 20% of the households surveyed had 3 color televisions.

(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.

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

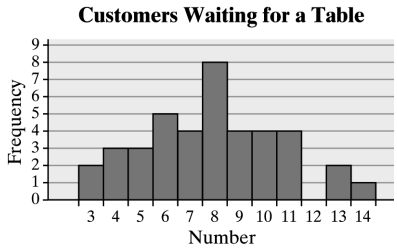
(b) and **(c)**
Relative frequency of 3 customers waiting = 2/40 = 0.05, and so on.

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

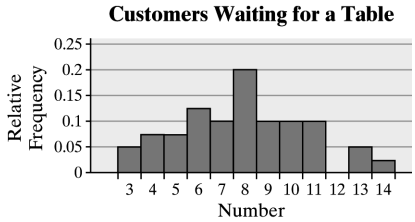
(d) $10.0 + 10.0 + 0.0 + 5.0 + 2.5 = 27.5\%$ of the Saturdays had 10 or more customers waiting for a table at 6 p.m.

(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)

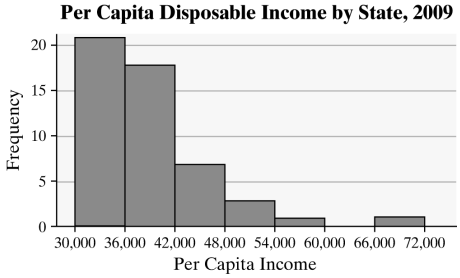


(h) The distribution is more or less symmetric.

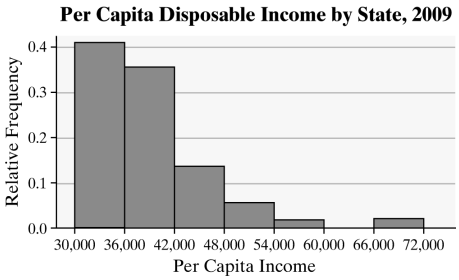
31. **(a)** and **(b)**
Relative frequency for 24,000-26,999 = 8/51 = 0.1569 and so on.

Disposable Income ($)	Freq.	Rel. Freq.
30,000 – 35,999	21	0.4118
36,000 – 41,999	18	0.3529
42,000 – 47,999	7	0.1373
48,000 – 53,999	3	0.0588
54,000 – 59,999	1	0.0196
60,000 – 65,999	0	0.0000
66,000 – 71,999	1	0.0196

(c)



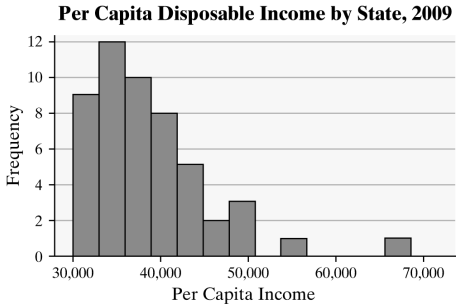
(d)

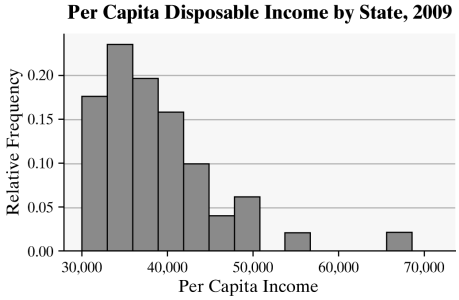


(e) The distribution appears to be skewed right.

(f) Relative frequency for 30,000-32,999 = 10/51 = 0.1961 and so on.

Disposable Income ($)	Freq.	Rel. Freq.
30,000 – 32,999	9	0.1765
33,000 – 35,999	12	0.2353
36,000 – 38,999	10	0.1961
39,000 – 41,999	8	0.1569
42,000 – 44,999	5	0.0980
45,000 – 47,999	2	0.0392
48,000 – 50,999	3	0.0588
51,000 – 53,999	0	0.0000
54,000 – 56,999	1	0.0196
57,000 – 59,999	0	0.0000
60,000 – 62,999	0	0.0000
63,000 – 65,999	0	0.0000
66,000 – 68,999	1	0.0196





The distribution appears to be skewed right.

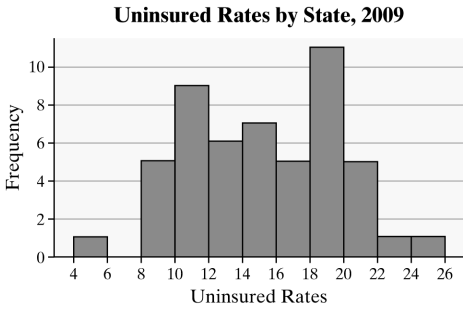
(g) Answers will vary. While both distributions indicate the data are skewed right, the first distribution provides a more detailed look at the data. The second distribution has a larger width of the bars, which can potential obscure details in the data.

32. **(a)** and **(b)**

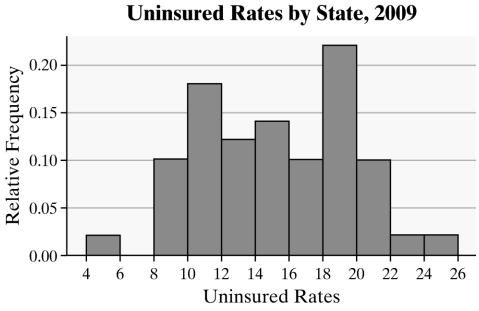
Relative frequency for 4.0-5.9 = 1/51 = 0.0196 and so on.

% Uninsured	Freq.	Rel. Freq.
4.0 – 5.9	1	0.0196
6.0 – 7.9	0	0.0000
8.0 – 9.9	5	0.0980
10.0 – 11.9	9	0.1765
12.0 – 13.9	6	0.1176
14.0 – 15.9	7	0.1373
16.0 – 17.9	5	0.0980
18.0 – 19.9	11	0.2157
20.0 – 21.9	5	0.0980
22.0 – 23.9	1	0.0196
24.0 – 25.9	1	0.0196

(c)



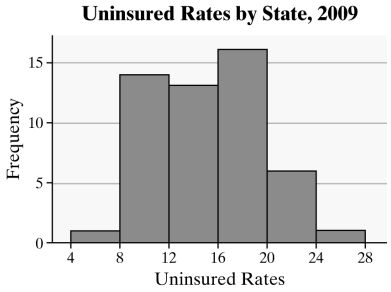
(d)

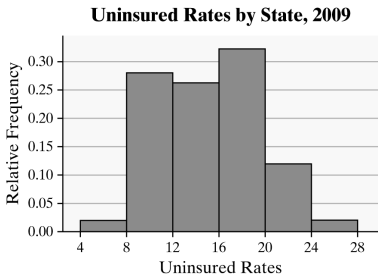


(e) The distribution is symmetric.

(f) Relative frequency for 4.0-7.9 = 1/51 = 0.0196 and so on.

% Uninsured	Freq.	Rel. Freq.
4.0 – 7.9	1	0.0196
8.0 – 11.9	14	0.2745
12.0 – 15.9	13	0.2549
16.0 – 19.9	16	0.3137
20.0 – 23.9	6	0.1176
24.0 – 25.9	1	0.0196





The distribution is symmetric.

(g) Answers will vary. Both frequency distributions show the data are skewed right. The number of classes in the first distribution gives more detail, but this makes the graph a bit more jagged. The second distribution gives a cleaner view of the data.

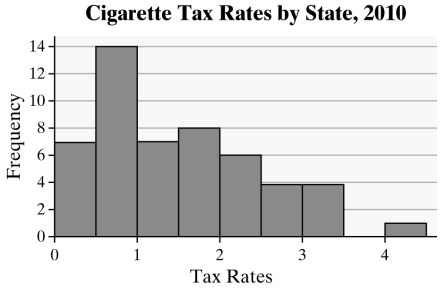
33. **(a)** and **(b)**

Total number of data points = 51

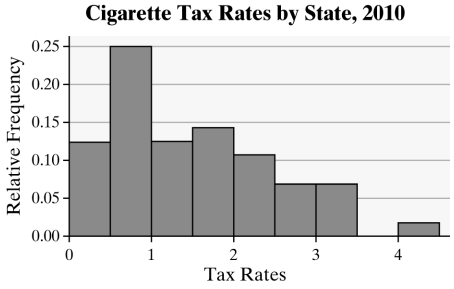
Relative frequency of 0-0.49 is: 7/51 = 0.1373, and so on.

Cigarette Tax	Frequency	Relative Frequency
0.00-0.49	7	0.1373
0.50-0.99	14	0.2745
1.00-1.49	7	0.1373
1.50-1.99	8	0.1569
2.00-2.49	6	0.1176
2.50-2.99	4	0.0784
3.00-3.49	4	0.0784
3.50-3.99	0	0.0000
4.00-4.49	1	0.0196

(c)



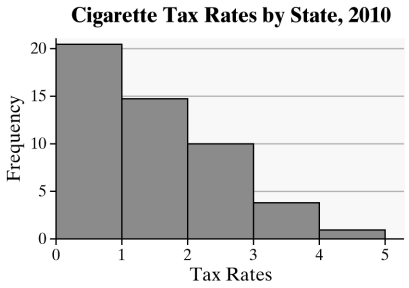
(d)

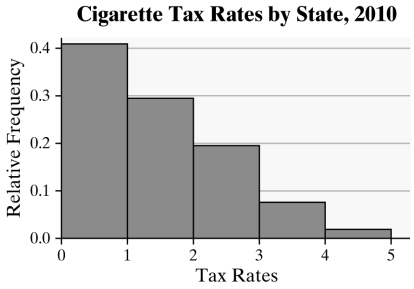


(e) The distribution appears to be right skewed.

(f) Relative frequency of 0 – 0.99is: 21/51 = 0.4118, and so on.

Cigarette Tax	Frequency	Relative Frequency
0.00-0.99	21	0.4118
1.00-1.99	15	0.2941
2.00-2.99	10	0.1961
3.00-3.99	4	0.0784
4.00-4.99	1	0.0196





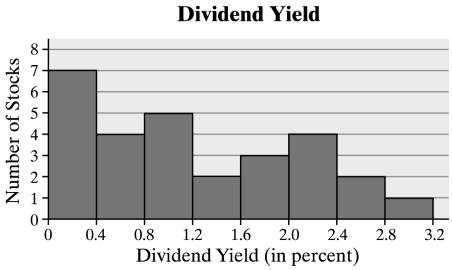
The distribution is right skewed.

(g) Answers will vary. The first distribution gives a more detailed pattern.

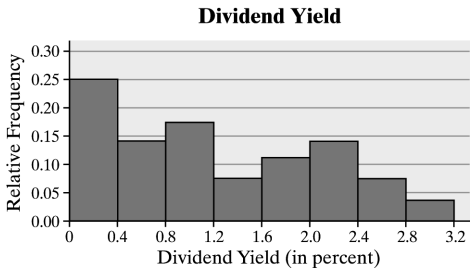
34. **(a)** and **(b)**
Relative frequency for 0.00-0.39 = 7/28 = 0.2500, and so on.

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)



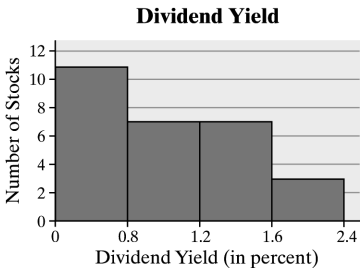
(d)

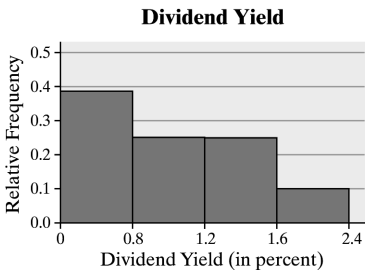


(e) The distribution is skewed right.

(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.

35. Answers will vary. One possibility follows.

(a) Lower class limit of first class: 100; 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,

$$\text{Class width} \approx \frac{1345.9 - 119.8}{13} = 94.3 \to 100$$

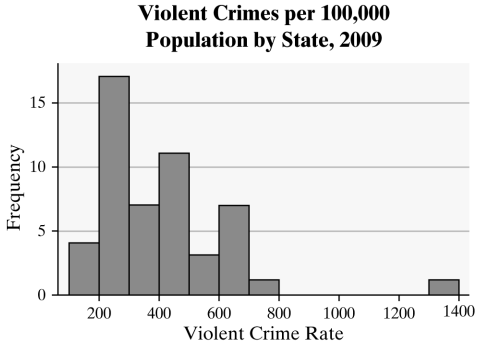
So, a class width 100 should suffice.

(b), (c)

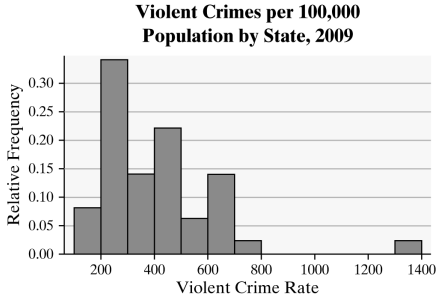
Relative frequency for 100-199 is: 4/51 = 0.0784, and so on.

Violent Crimes per 100,000	Freq.	Rel. Freq.
100 – 199	4	0.0784
200 – 299	17	0.3333
300 – 399	7	0.1373
400 – 499	11	0.2157
500 – 599	3	0.0588
600 – 699	7	0.1373
700 – 799	1	0.0196
800 – 899	0	0.0000
900 – 999	0	0.0000
1000 – 1099	0	0.0000
1100 – 1199	0	0.0000
1200 – 1299	0	0.0000
1300 – 1399	1	0.0196

(d)



(e)



(f) The distribution is skewed right.

36. 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,

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

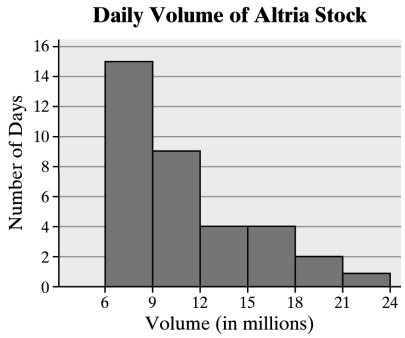
Our first lower class limit should be a nice number below the smallest data value. In this case, 6 is a good first lower limit since it is the nearest whole number below the smallest data value. Thus, we will have a class width of 3 and the first class will have a lower limit of 6.

(b), (c)

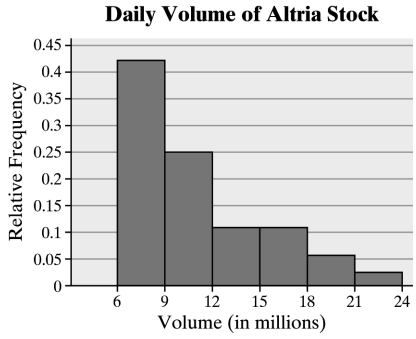
Relative frequency for 6-8.99 = 15/35 = 0.4286 and so on.

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.

37. **(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.

38. **(a)**

Divorce Rates by State, 2006

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

Legend: 2 | 1 represents 2.1 divorces per 1000 population.

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

39. **(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.

40. **(a)**

Gasoline Mileages

2 | 2233
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.

41. **(a)** Rounded data:

17.3	10.5	8.7	9.4	8.8	6.7	6.5
12.3	9.6	8.9	13.1	8.8	7.5	15.4
14.5	7.5	8.1	8.6	7.8	10.2	25.3
14.6	10.4	7.7	8.6	7.6	9.0	
14.6	9.4	8.1	8.8	8.1	7.6	
13.2	10.0	12.3	7.1	9.8	6.1	
15.7	7.8	13.9	9.2	10.5	14.4	
17.2	8.6	10.7	6.9	10.0	7.7	

Electric Rates by State, 2010

6	1579
7	155667788
8	11166678889
9	024468
10	0024557
11	
12	33
13	129
14	4566
15	47
16	
17	23
18	
19	
20	
21	
22	
23	
24	
25	3

Legend: 6 | 1 represents 6.1 cents per kWh.

(b) The distribution is skewed right.

(c) Hawaii's average retail price is 25.33 cents/kWh. Hawaii's rate may be so much higher because it is an island far away from the mainland. Resources on the island are limited and importing resources increases the overall cost. (Answers will vary.)

42. (a) Rounded data:

68	100	94	222	121	121	97
118	86	96	131	94	112	170
91	89	87	123	193	80	93
59	92	124	126	59	202	87
118	119	75	71	90	89	113
180	80	108	138	86	162	85
119	98	49	106	85	33	91
89	111					

Home Appreciation from 1991 to 2010

3	3
4	9
5	99
6	8
7	15
8	00556677999
9	0112344678
10	068
11	1238899
12	11346
13	18
14	
15	
16	2
17	0
18	0
19	3
20	2
21	
22	2

Legend: 8 | 3 represents 33 percent price increase.

(b) The distribution is skewed right.

(c) Answers will vary.

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

450	1350	400	280	380	670	300
630	610	260	490	400	190	260
410	430	620	250	200	670	230
520	270	120	280	330	490	
470	230	590	700	500	210	
340	500	460	160	250	130	
300	330	500	310	380	230	
640	280	240	620	250	330	

(b)

Violent Crime Rates by State, 2009

1	2369
2	013334555667888
3	001333488
4	001356799
5	00029
6	1223477
7	0
8	
9	
10	
11	
12	
13	5

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

(c)

Violent Crime Rates by State, 2009

1	23
1	69
2	013334
2	555667888
3	0013334
3	88
4	0013
4	56799
5	0002
5	9
6	12234
6	77
7	0
7	
8	
8	
9	
9	
10	
10	
11	
11	
12	
12	
13	
13	5

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.

44. (a)

Ages of Academy Award Winners

Best Actor		Best Actress
9	2	1566899
98877766220	3	012233333455689
876555332200	4	1125599
43210	5	
200	6	11
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.

45. (a)

Home Run Distances

McGwire		Bonds
	32	00
	33	
10	34	7
00	35	0
9000	36	00015
70000	37	005555
85500000	38	000005
80000	39	00146
900	40	000045
00000	41	0000000000155677
5300000	42	000000009
0000000	43	00000556
000	44	00002
820000	45	04
100	46	
8000	47	
0	48	8
	49	
0	50	
00	51	
7	52	
	53	
	54	
0	55	

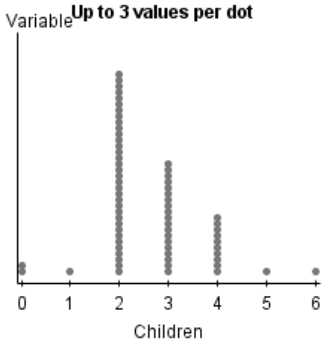
Legend: 0 | 34 | 7 represents 340 feet for McGwire and 347 feet for Bonds.

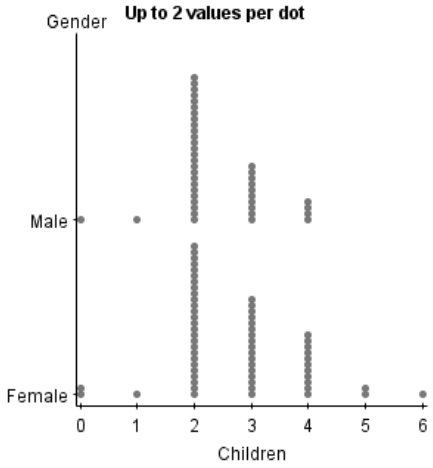
(b) Answers will vary. For both players, the distances of homeruns mainly fall from 360 to 450 feet. McGwire has quite a few extremely long distances.

46. Answers will vary.

47. Answers will vary.

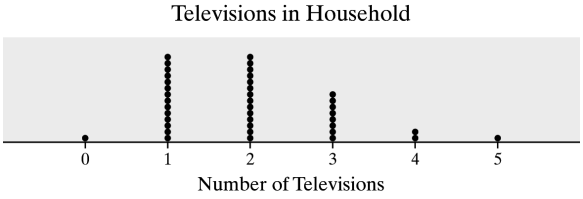
48.



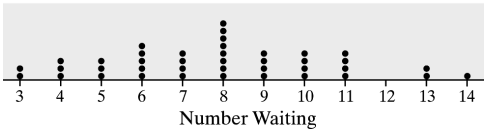


There are several similarities in the distribution of the ideal number of children, as reported by males and females. However, females seem more likely to esteem larger families as ideal.
A histogram would better serve us in comparing the preferences between males and females.

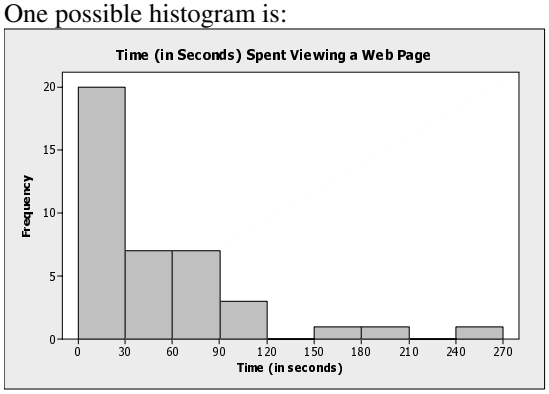
49.



50.



51. 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.

52. 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

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

54. 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.

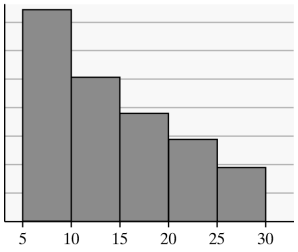
55. There is no such thing as *the* correct choice for a class width, however some choices are better than others. For example, if the class width is too small, the histogram will show many gaps between the bars.

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

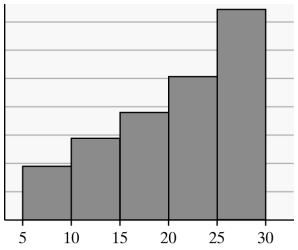
57. Answers will vary.

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

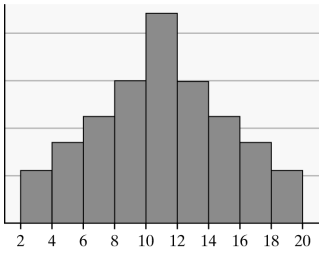
Skewed Right



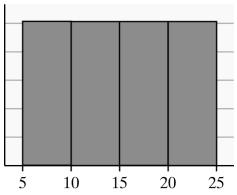
Skewed Left



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. A histogram is uniform if all the bars are about the same height.

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)** 10; The class width is the difference between successive class midpoints. e.g. $15-5=10$
There are 10 classes represented in the graph. The 12 plotted points are for the 10 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 5. Assuming data values are given in whole numbers, the lower limit of the first class is 0 and the upper limit is 9.

(c) The midpoint of the last class is 95. Assuming data values are given in whole numbers, the lower limit of the last class is 90 and the upper limit is 99.

(d) The highest population occurs in the age group $40-49$; this population is about 44 million.

(e) The lowest population occurs in the age group $90-99$; this population is about 2 million.

(f) We estimate the total population by adding the estimates for the 10 classes.
$42 + 41 + 43 + 40 + 44 + 41 + 28 + 16 + 9 + 2 = 306$
The approximate total number of U.S. residents below 100 years of age is 306 million.

(g) The population change is most extreme where the line segment is the longest. This occurs between the classes $50-59$ and $60-69$.
The population change is least extreme where the line segment is the shortest. This occurs between the classes $10-19$ and $20-29$.

6. **(a)** 10; The class width is the difference between successive class midpoints. e.g. $15-5=10$
There are 9 classes represented in the graph. The 11 plotted points are for the 9 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 5. Assuming data values are given in whole numbers, the lower limit of the first class is 0 and the upper limit is 9.

(c) The midpoint of the last class is 85. Assuming data values are given in whole numbers, the lower limit of the last class is 80 and the upper limit is 89.

(d) The highest number of deaths due to legal intervention occurs in the age group $40-49$.

(e) The highest number of deaths due to legal intervention occurs in the age group $20-29$ and $30-39$. There are about 46 deaths in these two groups.

(f) The frequency of the $20-29$ class is 46. There are about 175 observations in all. Therefore, the relative frequency is $\frac{46}{175} \approx 0.263$.

7. **(a)** From the graph, it appears that about 20% of students had an ACT composite score of 16 or below.

(b) From the graph, it appears that about 60% of students had an ACT composite score of 19 or above.

(c) About $(0.20)(1{,}568{,}835) \approx 313{,}767$ students had ACT composite scores of 16 or below.
About 21% scored above 25.
About $(0.21)(1{,}568{,}835) \approx 329{,}455$ students had ACT composite scores above 25.

8. **(a)** 50; The class width is the difference between successive class upper limits. e.g. $300 - 250 = 50$
There are 12 classes represented in the graph. The 13 plotted points are for the 12 upper class limits plus one additional point to connect the graph to the horizontal axis on the left.

(b) Since the points of an ogive are plotted at the upper class limits, the lower limit of the first class is 201 and the upper limit is 250.

(c) Since the points of an ogive are plotted at the upper class limits, the lower limit of the last class is 751 and the upper limit is 800.

(d) From the graph, it appears that 28% of the students had a score of 450 or below.

(e) From the graph, it appears that about 56% of the students had a score above 500.

(f) We can locate the class with the most students by locating the longest line segment in the graph. The longest line segment occurs for the class $501 - 550$.

9. **(a)** The HOI in the first quarter of 1999 is about 70%.

(b) The lowest value of the HOI was about 40%. This occurred in 2006.

(c) The highest value of the HOI was about 73%. This occurred in 2009.

(d) $\frac{40 - 70}{70} = \frac{-30}{70} \approx -0.43$
The HOI decreased by about 43% from the first quarter of 1999 to the third quarter of 2006.

(e) There is an increase of about 82.5%.

10. **(a)** About 8.8 million motor vehicles were produced in the United States in 1991.

(b) About 13.0 million motor vehicles were produced in the United States in 1999.

(c) $\frac{13000 - 8800}{8800} = \frac{4200}{8800} \approx 0.477$
The number of vehicles produced increased by about 47.7% between 1991 and 1999.

(d) $\frac{8600 - 13000}{13000} = \frac{-4400}{13000} \approx -0.338$
The number of vehicles produced decreased by about 33.8% between 1999 and 2009.

11. **(a)** For 1992, the unemployment rate was about 7.5% and the inflation rate was about 3.0%.

(b) For 2009, the unemployment rate was about 9.2% and the inflation rate was about -0.4%.

(c) $7.5\% + 3.0\% = 10.5\%$
The misery index for 1992 was 10.5%.
$4.6\% + 3.4\% = 8.0\%$
The misery index for 2009 was 8.8%.

(d) Answers may vary. One possibility: An increase in the inflation rate seems to be followed by an increase in the unemployment rate. Likewise, a decrease in the inflation rate seems to be followed by a decrease in the unemployment rate.

12. **(a)** In 1996, the men's prize money was £400,000 and the ladies' prize money was £350,000.

(b) In 2006, the men's prize money was £655,000 and the ladies' prize money was £625,000.

(c) Answers may vary. One possibility: Until 2007, the prize money for men's singles is higher than the prize money for ladies' singles. Both prizes increase over time at similar rates.

(d) In 2007, the prize money for men's and ladies' singles was the same for the first time. The prize money for each was £700,000.

(e) From 2009 to 2010, the prize money increased from £850,000 to £1,000,000 for both the men and the women. This is a relative increase of
$\frac{1,000,000 - 850,000}{850,000} \approx 0.176$
or 17.6%.

13. (a), (b)

Second class cumulative frequency = 22 + 58 = 90 and so on.

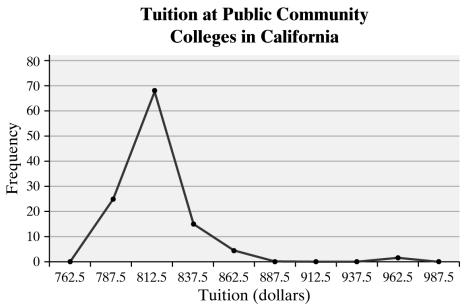
Total frequency is:

22 + 68 + 15 + 5+ 0+ 0+ 0+ 1 = 111

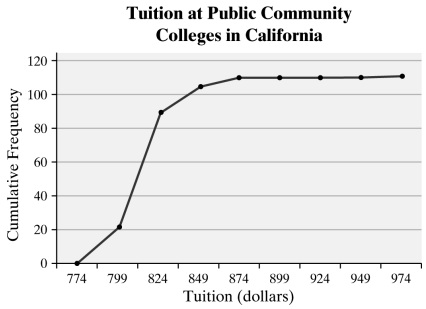
Second class cumulative relative frequency = $\frac{90}{111} = 0.8108$ and so on.

Tuition ($)	Cumul. Freq.	Cumul. Rel. Freq.
775-799	22	0.1982
800-824	90	0.8108
825-849	105	0.9459
850-874	110	0. 9910
875-899	110	0. 9910
900-924	110	0. 9910
925-949	110	0.9910
950-999	111	1

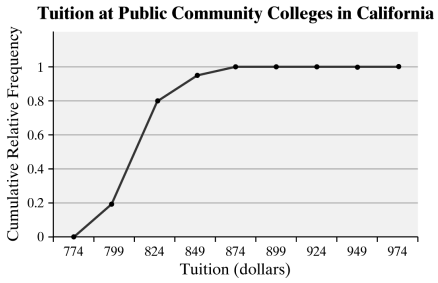
(c) The class midpoint is computed dividing the sum of successive lower class limits by 2. So, for example, the first class midpoint is: $\frac{775+800}{2} \approx 787.5$. The next class midpoint is: $\frac{800+825}{2} \approx 812.5$, and so on. We connect the graph with the x-axis at (762.5,0) and (987.5,0).



(d)



(e)



14. (a), (b)

Total number of earthquakes is:

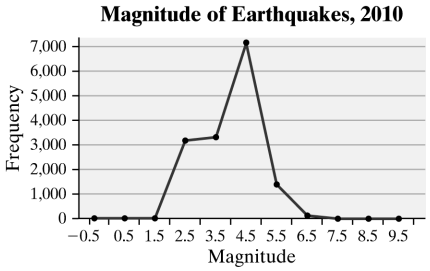
22 + 22 + 3201 + 3332 + 7276 + 1430 + 130 + 18+ 1 = 15,432

Second class cumulative frequency is 22 + 22 = 44 and so on.

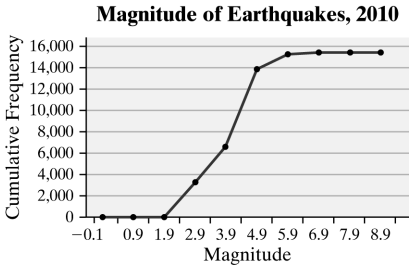
Second class cumulative relative frequency = $\frac{44}{15,432} \approx 0.0029$ and so on.

Magnitude	Cumul. Freq.	Cumul. Rel. Freq.
0-0.9	22	0.0014
1.0-1.9	44	0.0029
2.0-2.9	3,245	0.2103
3.0-3.9	6,577	0.4262
4.0-4.9	13,853	0.8977
5.0-5.9	15,283	0.9903
6.0-6.9	15,413	0.9988
7.0-7.9	15,431	0.9999
8.0-8.9	15,432	1.0000

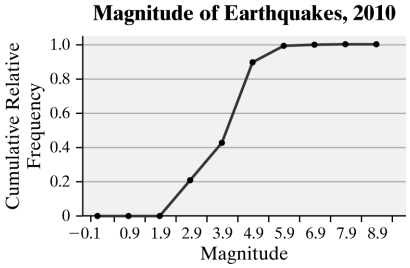
(c) The class midpoint is computed dividing the sum of successive lower class limits by 2. For example, the first class midpoint is: $\frac{0+1}{2} \approx 0.5$. The next class midpoint is: $\frac{1+2}{2} \approx 1.5$, and so on. We connect the graph with the x-axis at $(-0.5, 0)$ and $(9.5, 0)$.



(d)



(e)

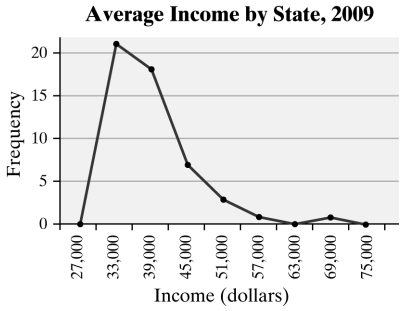


15. (a), (b)

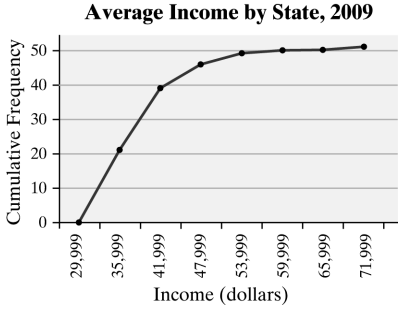
Second class cumulative frequency is 21 + 18 = 39 and so on.
Second class cumulative relative frequency
$= \frac{39}{51} \approx 0.7647$ and so on.

Disposable Income ($)	Cumul. Freq.	Cumul. Rel. Freq.
30,000 – 35,999	21	0.4118
36,000 – 41,999	39	0.7647
42,000 – 47,999	46	0.9020
48,000 – 53,999	49	0.9608
54,000 – 59,999	50	0.9804
60,000 – 65,999	50	0.9804
66,000 – 71,999	51	1

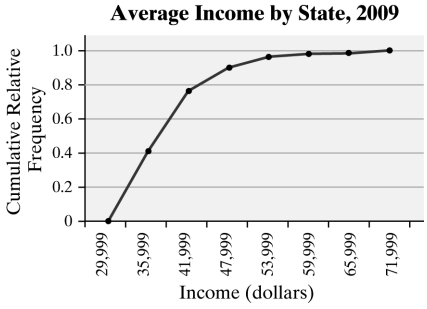
(c) The class midpoint is computed dividing the sum of successive lower class limits by 2. So, for example, the first class midpoint is: $\frac{30,000+36,000}{2} \approx 33,000$.
The next class midpoint is:
$\frac{36,000+42,000}{2} \approx 39,000$, and so on.
We connect the graph with the x-axis at $(27000, 0)$ and $(75000, 0)$.



(d)



(e)

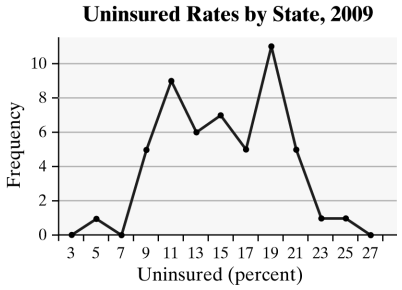


16. (a), (b)

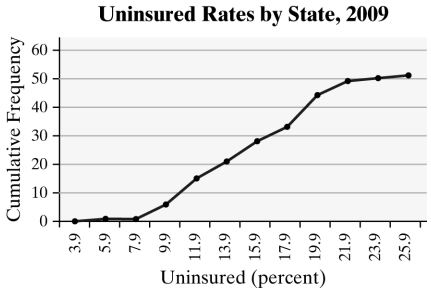
Second class cumulative frequency = 1 + 0 = 1 and so on.
Second class cumulative relative frequency = $\frac{1}{51} \approx 0.0196$ and so on.

% Uninsured	Cumul. Freq.	Cumul. Rel. Freq.
4.0 – 5.9	1	0.0196
6.0 – 7.9	1	0.0196
8.0 – 9.9	6	0.1176
10.0 – 11.9	15	0.2941
12.0 – 13.9	21	0.4118
14.0 – 15.9	28	0.5940
16.0 – 17.9	33	0.6471
18.0 – 19.9	44	0.8627
20.0 – 21.9	49	0.9608
22.0 – 23.9	50	0.9804
24.0 – 25.9	51	1.0000

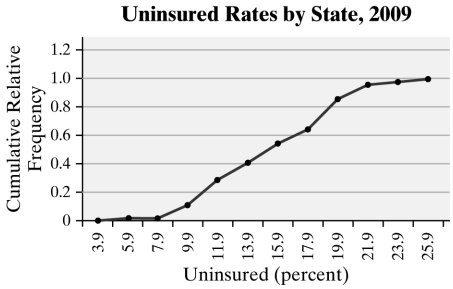
(c)



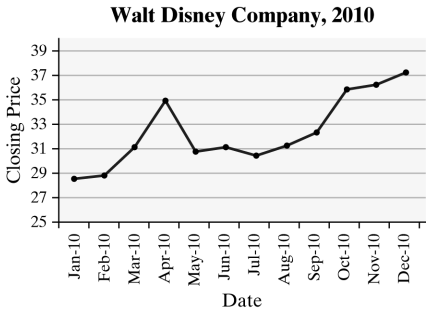
(d)



(e)



17. The price of Disney stock over the year seemed to fluctuate somewhat, with a general upward trend.

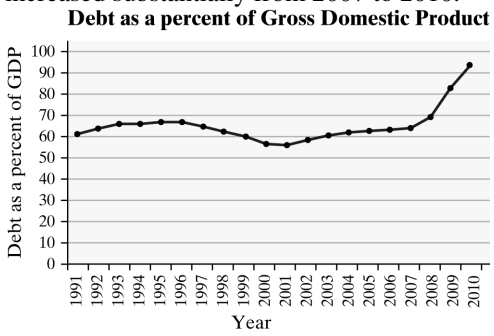


The percent change in the stock price of Disney from January 2010 to December 1010 is 30.7%.

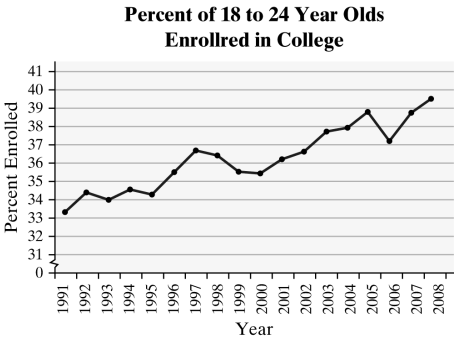
18. Since going public, the price of Google stock has increased rapidly at the onset with a sharp decrease in 2008, and then a recovery in 2009 and a slight decrease in 2010.

The percent change in the stock price of Google from 2009 to 2010 is -4.2%.

19. During the late 1990s debt as a percent of GDP was decreasing. It was increasing slightly during the early to mid 2000s. It has increased substantially from 2007 to 2010.



20. The percent of recent high school graduates enrolling in college seems to have increased slightly over the given time period amid a variety of fluctuations. The early 1990s showed no increase, but this was followed by an unusual jump and decline in the mid-to-late 1990s. There has been a steady increase since 2000 (with the exception of 2005 to 2006.)



21. (a) Total frequency for financial stocks = 35

Total frequency for energy stocks = 35

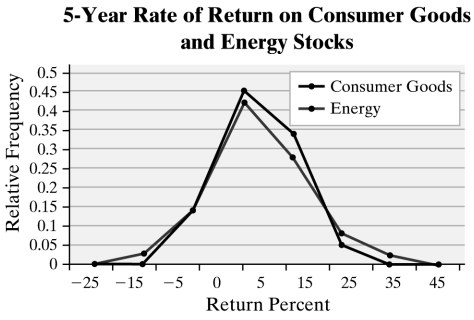
Rate of Return	Frequency (Consumer Goods)	Frequency (Energy)
−20 to −10.01	1	0
−10 to −0.01	5	5
0 – 9.99	15	16
10 – 19.99	10	12
20 – 29.99	3	2
30 – 39.99	1	0

(b) Relative frequency of −20 to −10.01 for consumer goods stocks = $\frac{1}{35} \approx 0.0286$, and so on.

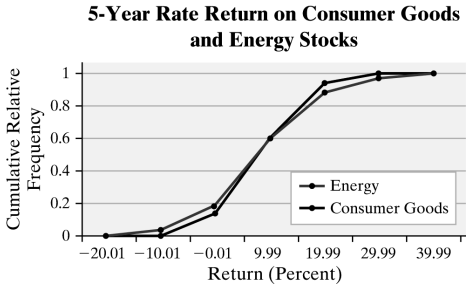
Relative frequency of −20 to −10.01 for energy stocks = $\frac{0}{35} \approx 0.0000$, and so on.

Rate of Return	Rel. Freq. (Consumer Goods)	Rel. Freq. (Energy)
−20 to −10.01	0.0286	0.0000
−10 to −0.01	0.1429	0.1429
0 – 9.99	0.4286	0.4571
10 – 19.99	0.2857	0.3429
20 – 29.99	0.0857	0.0571
30 – 39.99	0.0286	0.0000

(c) The class midpoint is computed dividing the sum of successive lower class limits by 2. So, for example, the first class midpoint is: $\frac{-20+(-10)}{2} \approx -15$. The next class midpoint is: $\frac{-10+0}{2} \approx -5$, and so on. We connect the graph with the x-axis at $(-25,0)$ and $(45,0)$.



(d)



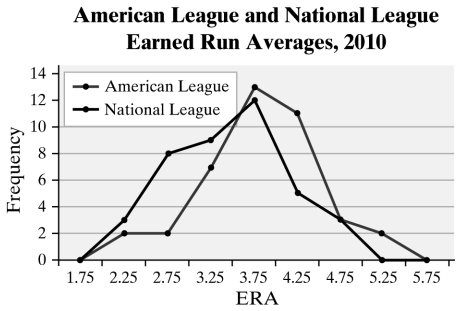
(e) Answers will vary.

22. (a) Since the sample sizes are the same, we can use frequencies to compare the two leagues.

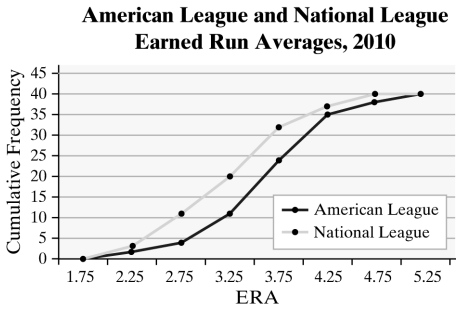
(b)

ERA	Freq. (AL)	Freq. (NL)
2.00 – 2.49	2	3
2.50 – 2.99	2	8
3.00 – 3.49	7	9
3.50 – 3.99	13	12
4.00 – 4.49	11	5
4.50 – 4.99	3	3
5.00 – 5.49	2	0

(c) The class midpoint is computed dividing the sum of successive lower class limits by 2. So, for example, the first class midpoint is: $\frac{2+2.5}{2} \approx 2.25$. The next class midpoint is: $\frac{2.5+3}{2} \approx 2.75$, and so on. We connect the graph with the x-axis at $(1.75,0)$ and $(5.75,0)$.



(d)



(e) Answers will vary. Overall the two leagues seem similar, though the NL appears to have more variability. This could be because pitchers (who tend to be poor batters) must bat in the NL. Designated hitters are allowed in the AL, so this may cause an increase in the AL ERA values, and more uniformity, since the DH is generally a better hitter than the pitcher.

23. 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.

24. Answers will vary.

25. 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%.

26. Answers will vary. Time series plots are drawn with quantitative variables. They are drawn to see trends in the data.

Section 2.4

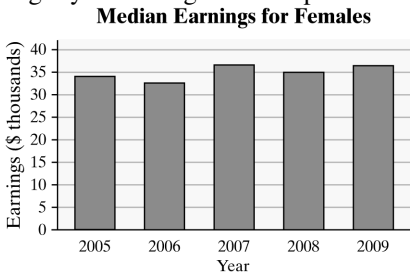
1. The lengths of the bars are not proportional. For example, the bar representing the cost of Clinton's inauguration should be slightly more than 9 times as long as the one for Carter's cost, and twice as long as the bar representing Reagan's cost.

2. **(a)** The lengths of the bars are not proportional. For example, the bar for soda is 1/3 the size of the bar for 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 34,500 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 might cause 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.

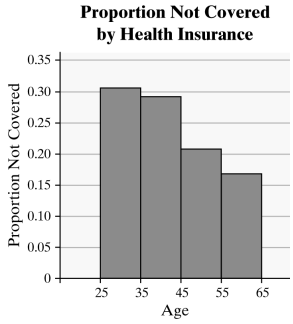
(b)

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. **(a)** The vertical axis starts at 0.1 instead of 0. This might cause the reader to conclude, for example, that the proportion of people aged 25-34 years old who are not covered by any type of health insurance is more than 3 times the proportion for those aged 55-64 years old.

(b)



8. **(a)** The vertical axis starts at 50 without indicating a gap.

(b) The graph may be trying to convey that new home construction in the Midwest increased significantly between 2000 and 2004, then declined at roughly twice the rate between 2004 and 2009.

(c) The graph does support this view. The sharp drop after 2005 indicates a market cooling and less demand for new construction. The National Association of Realtors was not correct in their assessment of the new home market.

9. This graph is misleading because it does not take into account the size of the population of each state. Certainly, Vermont is going to pay less in total taxes than California simply because its population is so much lower. The moral of the story here is that many variables should be considered on a per capita (per person) bias. For example, this graph should be drawn with taxes paid per capita (per person).

10. **(a)** $\dfrac{726.6}{7.5} = 96.88 \rightarrow 97$

Since the oil reserve in 2009 was about 97 times the oil reserve in 1977, the graphic for 2009 should be 97 times the size (i.e. area) of the graphic for 1977.

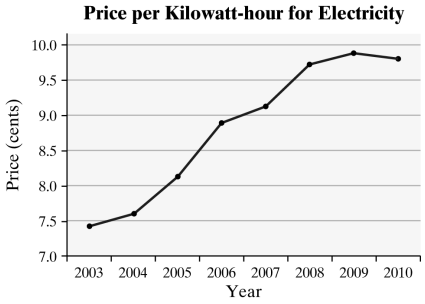
(b) $\dfrac{726.6 \text{ million}}{10 \text{ million}} = \dfrac{726.6}{10} \approx 72.7$

Assuming no change in production, the 2009 oil reserves would last about 73 days.

11. **(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.

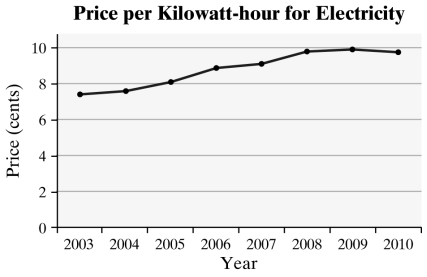
(b) The graphic could be improved by adjusting the bars so that their lengths are proportional.

12. **(a)**

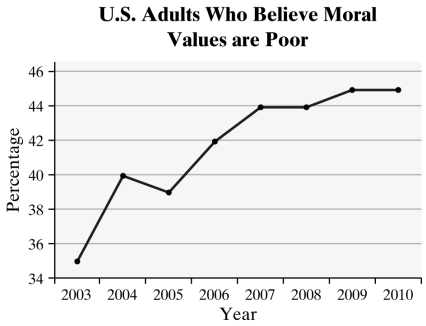


The graphic is misleading because it starts at 7 instead of 0 without indicating a gap. This might cause the reader to conclude that cost of electricity has risen more sharply than actually occurred.

(b)

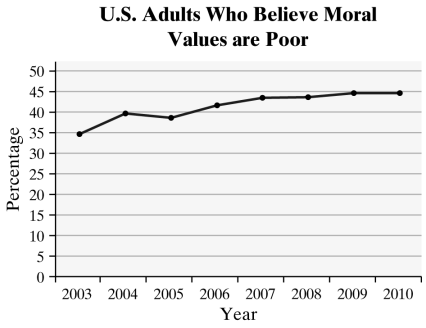


13. (a)



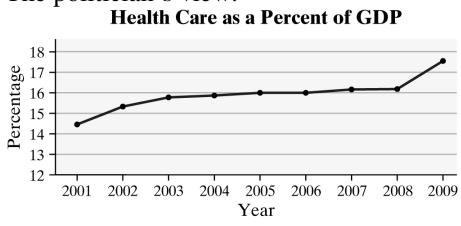
This graphic is misleading because the vertical scale starts at 34 instead of 0 without indicating a gap. This might cause the reader to think that the proportion of U.S. adults who believe moral values are poor is increasing more quickly than they really are.

(b)

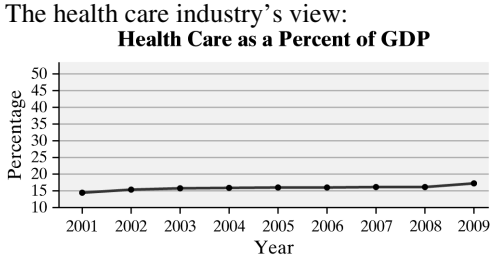


14. The graphic does not support the safety manager's claim. The vertical scale starts at 0.17 instead of 0, so the difference between the bars is distorted. While there has been a decrease in the proportion of workers injured, it appears that the decrease is only about 10% of the 1992 rate.

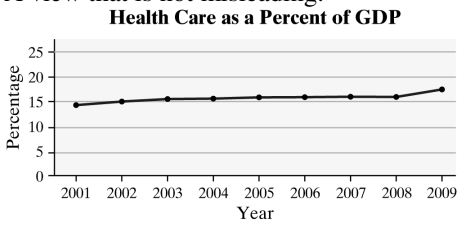
15. (a) The politician's view:



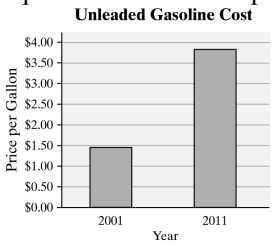
(b) The health care industry's view:



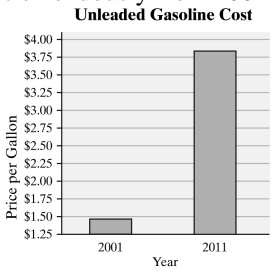
(c) A view that is not misleading:



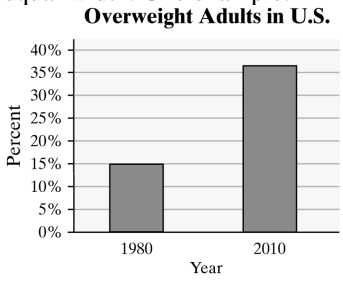
16. (a) A graph that is not misleading will use a vertical scale starting at $0 and bars of equal width. One example:



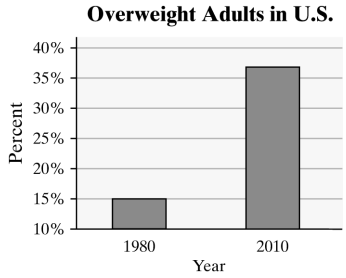
(b) A graph that is misleading might use bars of unequal width or will use a vertical scale that does not start at $0. One example, as follows, indicates that the average price for regular unleaded gasoline has increased tremendously from 2001 to 2011.



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



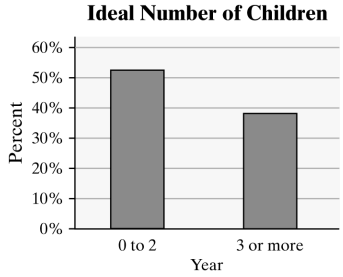
(b) A graph that is misleading might use bars of unequal width or will use a vertical scale that does not start at 0%. One example, as follows, implies that the percent of overweight adults in the U.S. has more than quadrupled between 1980 and 2010.



18. **(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)



19. This is a histogram and the bars do not touch. In addition, there are no labels on the axes and there is not title on the graph.

20. **(a)** The two variables being graphed are Cornell's tuition and Cornell's ranking.

Tuition: quantitative, discrete, interval level of measurement

Ranking: qualitative, ordinal level of measurement

(b) The graph was likely generated from existing data sources such as the U.S. Department of Education.

(c) The graph is a time series plot because the data are plotted in time order.

(d) Answers may vary. The graph appears to show that the cost to attend Cornell is increasing while its ranking is decreasing. The graph is misleading because the ranking graph may imply that ranking is getting worse (the graph goes down) when in fact a lower value means a better ranking.

(e) Answers may vary. No horizontal or vertical scale is provided; the horizontal scale is inconsistent; the vertical scale is inconsistent between graphs; the reader's attention is drawn more to the picture than the data.

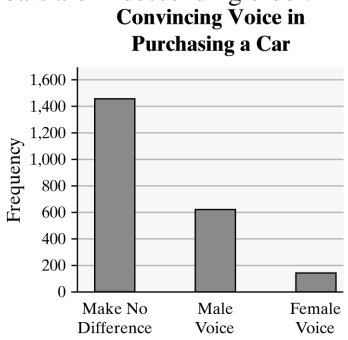
Chapter 2 Review Exercises

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.

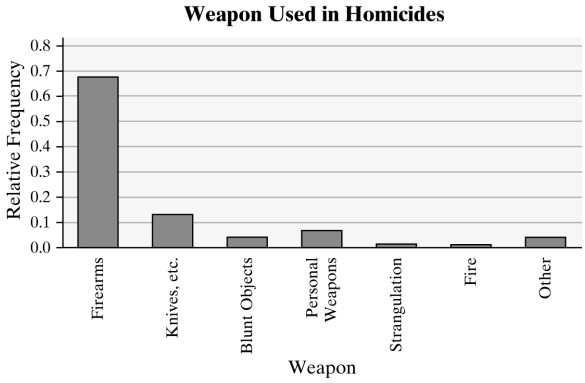
2. **(a)** Total homicides = 9,146 + 1,825 + 611 + 801 + 121 + 99 + 895 = 13,498
Relative frequency for Firearms = $\frac{9,146}{13,498} \approx 0.6776$ and so on.

Type of Weapon	Relative Frequency
Firearms	0.6776
Knives or cutting intstruments	0.1352
Blunt objects (clubs, hammers, etc.)	0.0453
Personal weapons (hands, fists, etc.)	0.0593
Strangulation	0.0090
Fire	0.0073
Other weapon or not stated	0.0663

(b) The relative frequency is 0.0453, so 4.5% of the homicides were due to blunt objects.

(c)

(d)



(e)

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

Total births (in thousands) = 6 + 435 + 1053 + 1197 + 958 + 489 + 106 = 4244

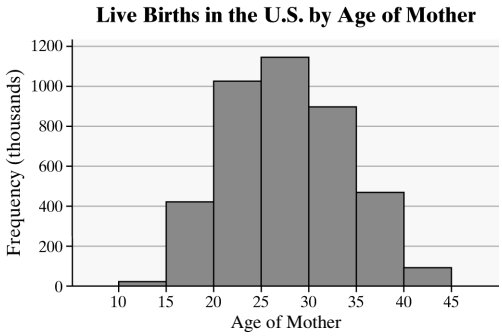
Relative frequency for 10-14 year old mothers = $6 / 4244 \approx 0.0014$ and so on.

Cumulative frequency for 15-19 year old mothers = 6 + 435 = 441 and so on.

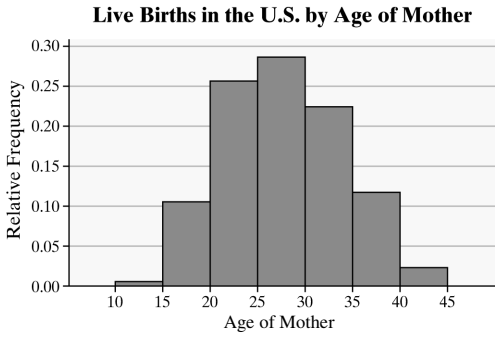
Cumulative relative frequency for 15-19 year old mothers = $441 / 4244 \approx 0.1039$ and so on.

Age of Mother	Rel. Freq.	Cumul. Freq.	Cumul. Rel. Freq.
10 – 14	0.0014	6	0.0014
15 – 19	0.1025	441	0.1039
20 – 24	0.2481	1494	0.3520
25 – 29	0.2820	2691	0.6341
30 – 34	0.2257	3649	0.8598
35 – 39	0.1152	4138	0.9750
40 – 44	0.0250	4244	1.0000

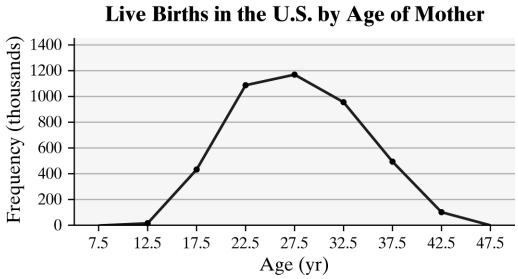
(d) The distribution is roughly symmetric and bell shaped.



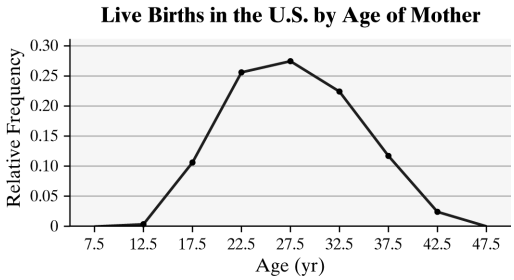
(e)



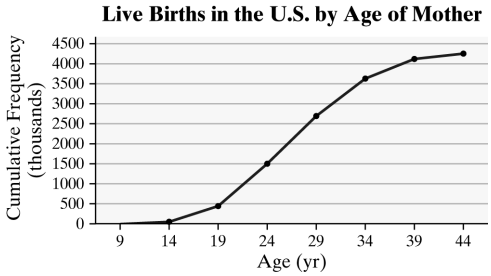
(f)



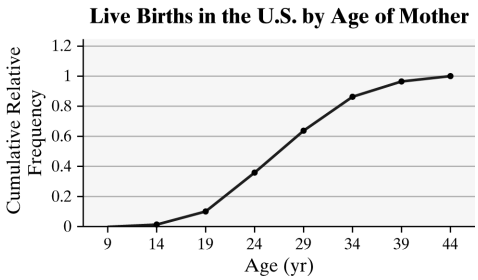
(g)



(h)



(i)



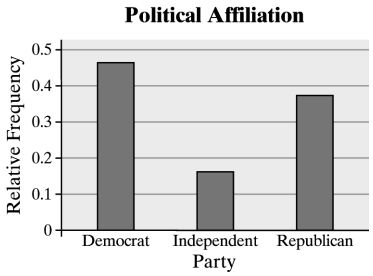
(j) From the relative frequency table, the relative frequency of 20-24 is 0.2481 and so the percentage is 24.81%.

(k) $\frac{958+489+106}{4244} = \frac{1553}{4244} \approx 0.3659$
36.59% of live births were to mothers aged 30 years or older.

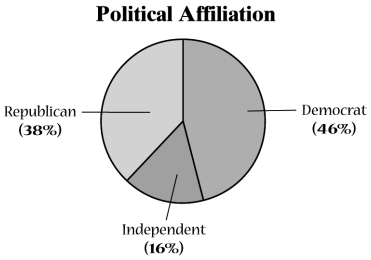
4. **(a)** and **(b)**

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

(c)



(d)

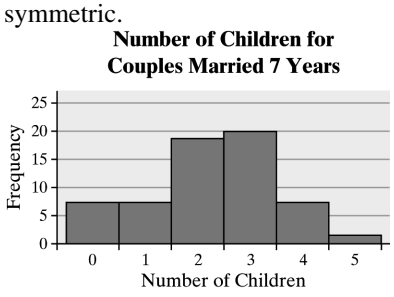


(e) Democrat appears to be the most common affiliation in Naperville.

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.

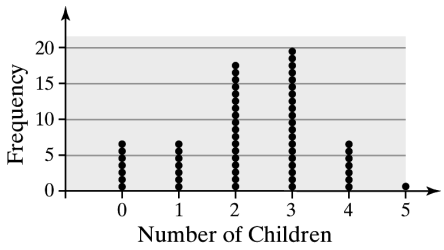


(f)

(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.

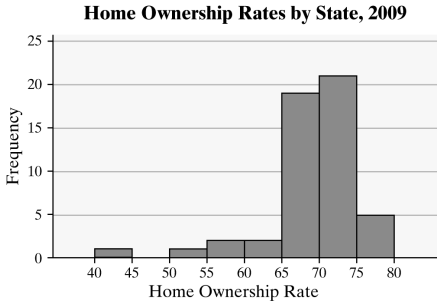
(i)



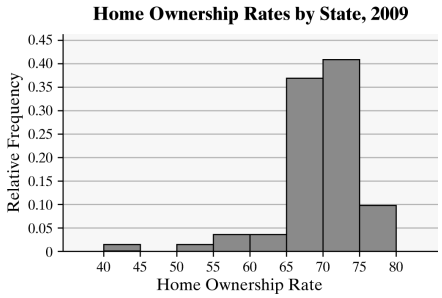
6. (a), (b)

Ownership Rate	Freq.	Rel. Freq.
40 – 44.9	1	0.0196
45 – 49.9	0	0.0000
50 – 54.9	1	0.0196
55 – 59.9	2	0.0392
60 – 64.9	2	0.0392
65 – 69.9	19	0.3725
70 – 74.9	21	0.4118
75 – 79.9	5	0.0980

(c)



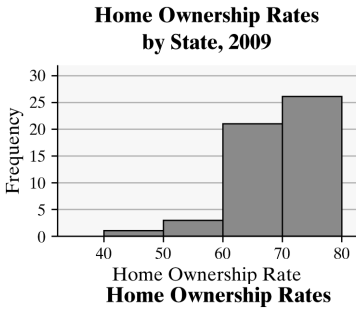
(d)

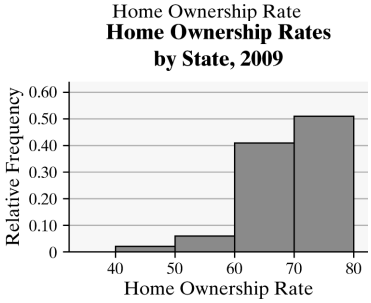


(e) The distribution is skewed left.

(f)

Ownership Rate	Freq.	Rel. Freq.
40 – 49.9	1	0.0196
50 – 59.9	3	0.0588
60 – 69.9	21	0.4118
70 – 79.9	26	0.5098





The distribution is skewed left.

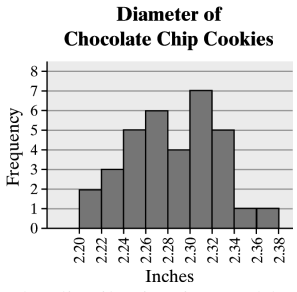
Answers will vary. Both class widths give a good overall picture of the distribution. The first class width provides a little more detail to the graph, but not necessarily enough to be worth the trouble. An intermediate value, say a width of 500, might be a reasonable compromise.

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.

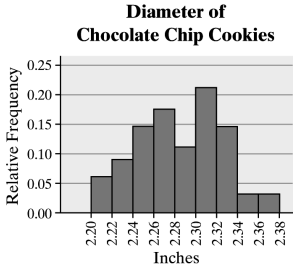
Class	Freq.	Rel. Freq.	Cumul. Freq.	Cumul. Rel.Freq.
2.2000 – 2.2199	2	0.0588	2	0.0588
2.2200 – 2.2399	3	0.0882	5	0.1471
2.2400 – 2.2599	5	0.1471	10	0.2941
2.2600 – 2.2799	6	0.1765	16	0.4706
2.2800 – 2.2999	4	0.1176	20	0.5882
2.3000 – 2.3199	7	0.2059	27	0.7941
2.3200 – 2.3399	5	0.1471	32	0.9412
2.3400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

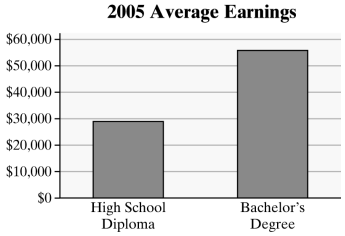
9. **(a)** Grade inflation seems to be happening in colleges. GPAs have increased every time period for all schools.

(b) GPAs increased 5.6% for public schools. GPAs increased 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.

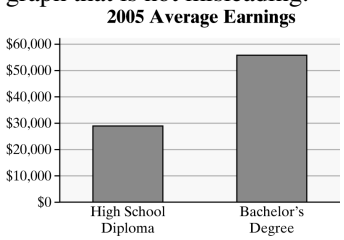
10. **(a)** Answers will vary. The adjusted gross income share of the top 1% of earners shows steady increases overall, with a few minor exceptions. The adjusted gross income share of the bottom 50% of earners shows steady decreases overall, with a few minor exceptions.

(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) The United States won the most men's singles championships between 1968 and 2010 with 15 wins.

(b) $6 - 4 = 2$
Representatives from Australia have won 2 more championships than representatives from Germany.

(c) $15 + 8 + 6 + 5 + 4 + 2 + 1 + 1 + 1 = 43$ championships between 1968 and 2010.
$\frac{8}{43} = 0.186$
Representatives of Sweden won 18.6% of the championships.

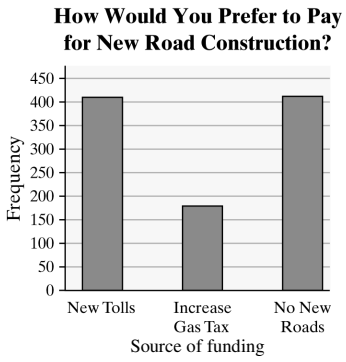
(d) No, it is not appropriate to describe the shape of the distribution as skewed right. The data represented by the graph are qualitative so the bars in the graph could be placed in any order.

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

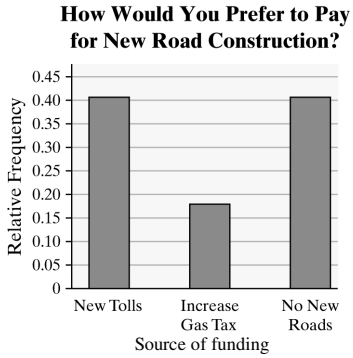
Response	Freq.	Rel. Freq.
New Tolls	412	0.4100
Inc. Gas Tax	181	0.1801
No New Roads	412	0.4100

(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%.

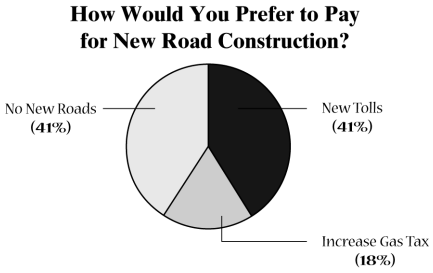
(c)



(d)



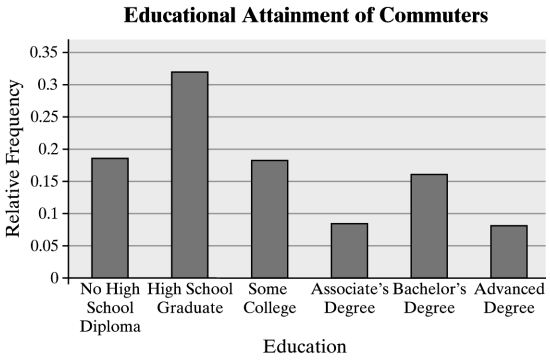
(e)



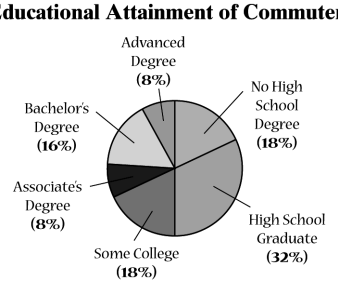
3. (a), (b)

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)

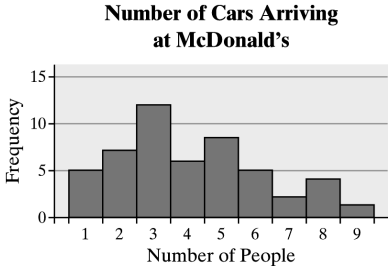


(e) The largest bar (and largest pie segment) corresponds to 'High School Graduate', so high school graduate is the most common educational level of a commuter.

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

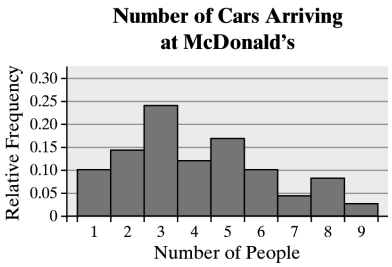
No. of Cars	Freq.	Rel. Freq.	Cumul. Freq.	Cumul. Rel. Freq.
1	5	0.10	5	0.10
2	7	0.14	12	0.24
3	12	0.24	24	0.48
4	6	0.12	30	0.60
5	8	0.16	38	0.76
6	5	0.10	43	0.86
7	2	0.04	45	0.90
8	4	0.08	49	0.98
9	1	0.02	50	1

(e)



The distribution is skewed right.

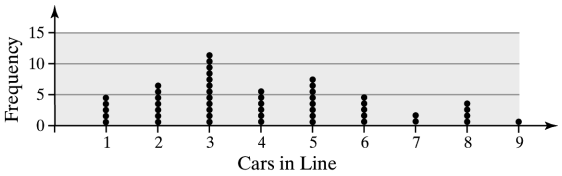
(f)



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

(i)

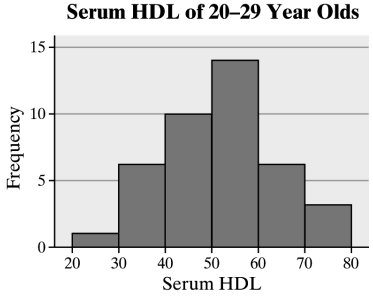


5. Answers may vary. One possibility follows:

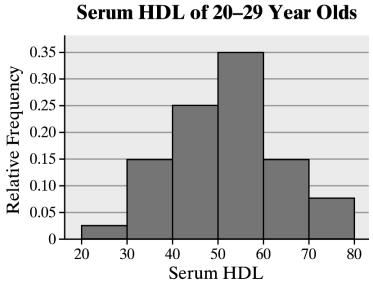
(a), (b)
Using a lower class limit of the first class of 20 and a class width of 10:
Total number of data points = 40
Relative frequency of 20 – 29 = 1/40 = 0.025, and so on.

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)



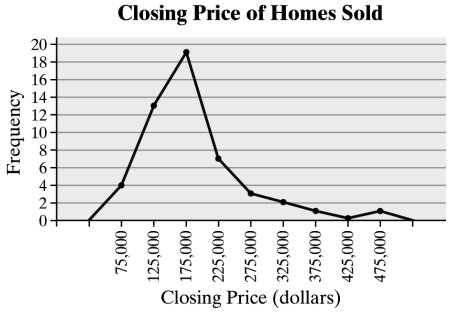
(e) The distribution appears to be roughly bell-shaped.

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

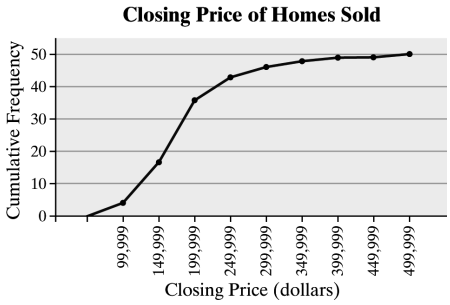
(c) The cumulative relative frequency for the class \$150,000-\$199,000 is 0.72. Therefore, 72% of the homes sold for less than \$200,000.

(d)

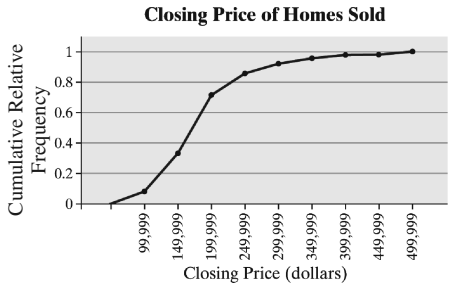


The distribution is skewed right.

(e)



(f)



7. The stem-and-leaf diagram below shows a symmetric (uniform) distribution.

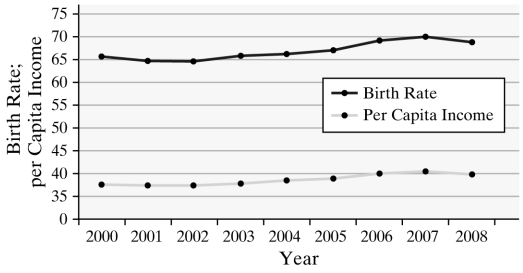
Time Spent on Homework

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.

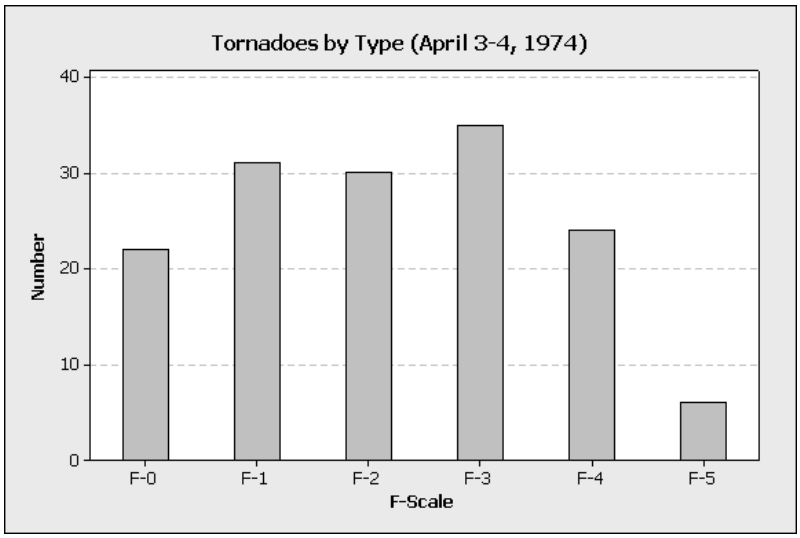
Fertility Rates (births per 1000 women aged 15–44) and per Capita Income (thousands of 2008 dollars)

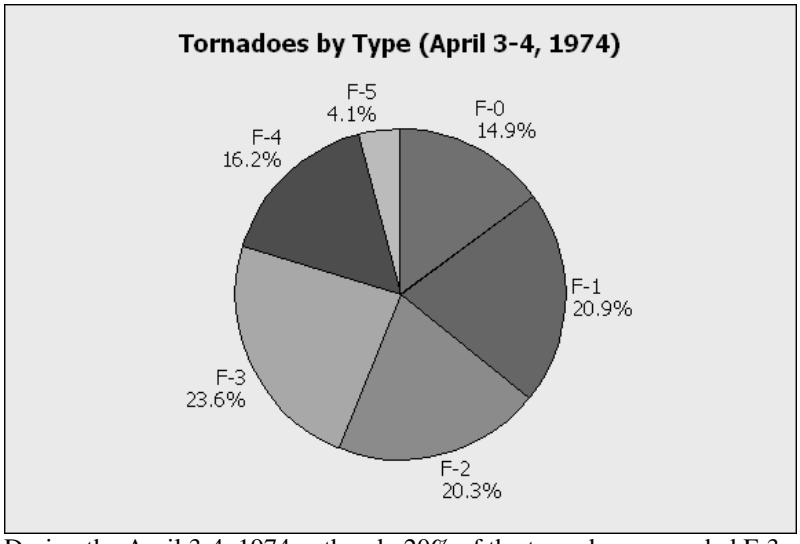


9. Answers may vary. It is difficult to interpret this graph because it is not clear whether the scale is represented by the height of the steps, the width of the steps, or by the graphics above the steps. The graphics are misleading because they must be increased in size both vertically and horizontally 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

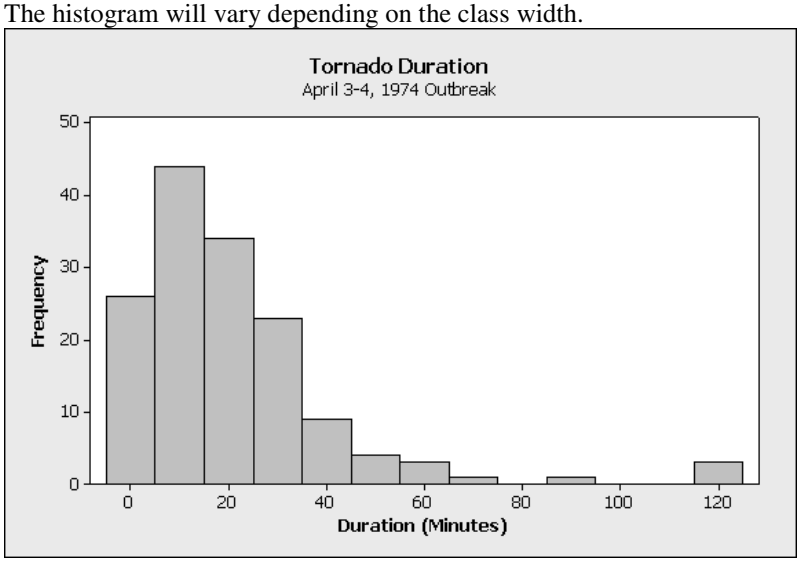
1.



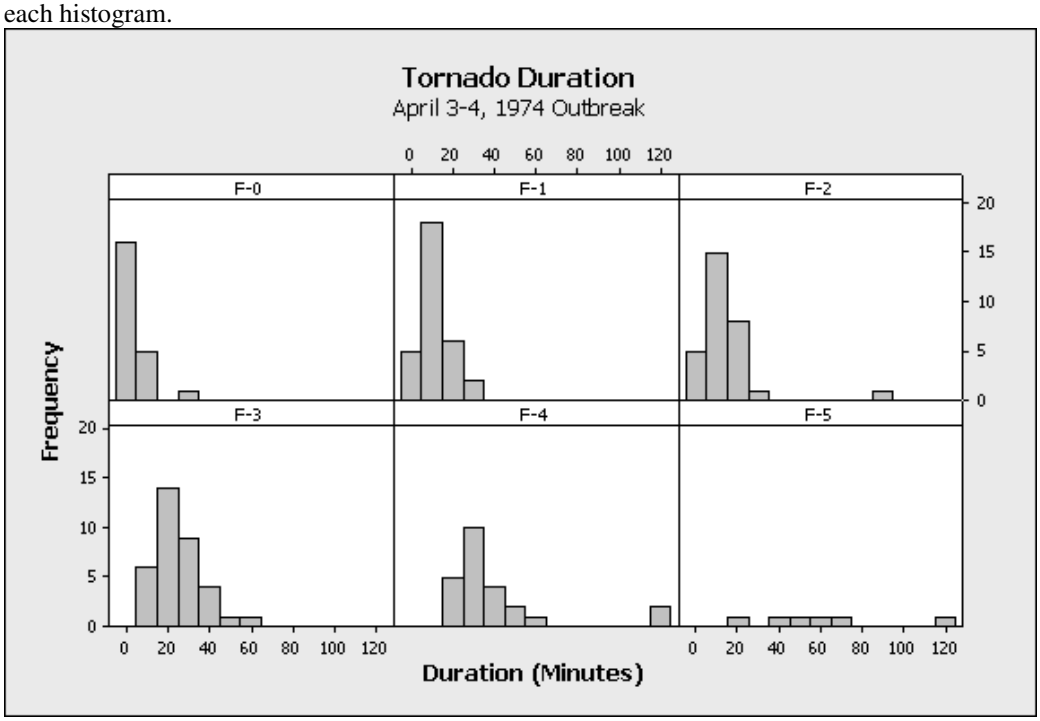


During the April 3-4, 1974 outbreak, 20% of the tornadoes exceeded F-3 on the Fujita Wind Damage Scale. This was much greater than the 1% that typically occurs.

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.



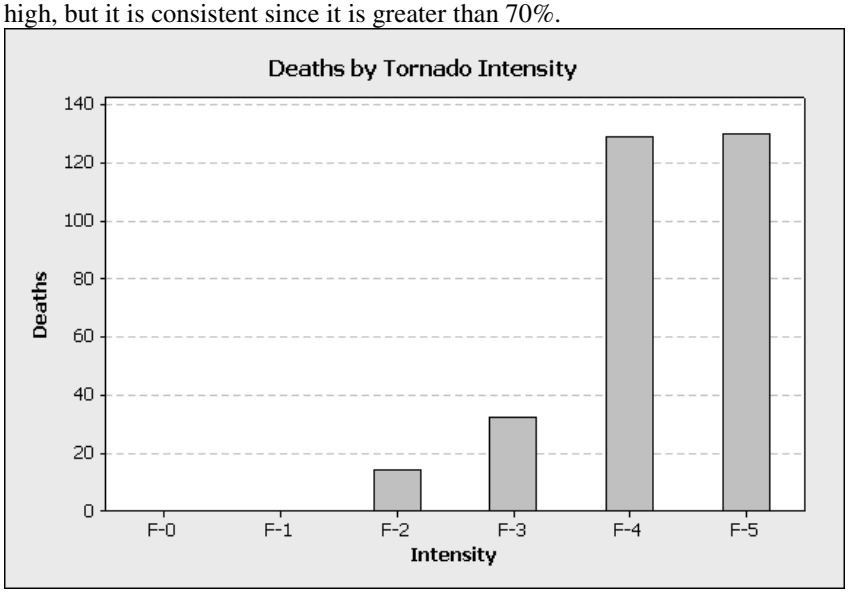
The distributions all appear to be skewed right, though the distribution for F-5 tornadoes is difficult to see due to the low sample size. There is an obvious shift in the distributions. As the strength of the tornado increases, the duration of the tornado increases.

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

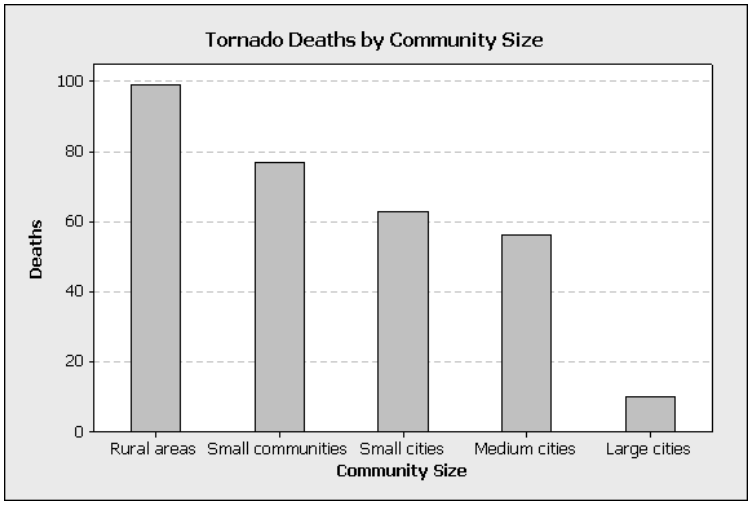
$$\frac{259}{305} \approx 0.8492$$

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.