Part V

Complete Solutions

Chapter 1 Getting Started

Section 1.1

1. (a) The variable is the response regarding frequency of eating at fast-food restaurants.
   (b) The variable is qualitative. The categories are the number of times one eats in fast-food restaurants.
   (c) The implied population is responses for all adults in the U.S.

2. (a) The variable is the miles per gallon.
   (b) The variable is quantitative because arithmetic operations can be applied the mpg values.
   (c) The implied population is gasoline mileage for all new 2001 cars.

3. (a) The variable is student fees.
   (b) The variable is quantitative because arithmetic operations can be applied to the fee values.
   (c) The implied population is student fees at all colleges and universities in the U.S.

4. (a) The variable is the shelf life.
   (b) The variable is quantitative because arithmetic operations can be applied to the shelf life values.
   (c) The implied population is the shelf life of all Healthy Crunch granola bars.

5. (a) The variable is the time interval between check arrival and clearance.
   (b) The variable is quantitative because arithmetic operations can be applied to the time intervals.
   (c) The implied population is the time interval between check arrival and clearance for all companies in the five-state region.

6. Form B would be better. Statistical methods can be applied to the ordinal data obtained from Form B, but not to the answers obtained from Form A.

7. (a) *Length of time to complete an exam* is a ratio level of measurement. The data may be arranged in order, differences and ratios are meaningful, and a time of 0 is the starting point for all measurements.
   (b) *Time of first class* is an interval level of measurement. The data may be arranged in order and differences are meaningful.
   (c) *Class categories* is a nominal level of measurement. The data consists of names only.
   (d) *Course evaluation scale* is an ordinal level of measurement. The data may be arranged in order.
   (e) *Score on last exam* is a ratio level of measurement. The data may be arranged in order, differences and ratios are meaningful, and a score of 0 is the starting point for all measurements.
   (f) *Age of student* is a ratio level of measurement. The data may be arranged in order, differences and ratios are meaningful, and an age of 0 is the starting point for all measurements.

8. (a) *Salesperson's performance* is an ordinal level of measurement. The data may be arranged in order.
   (b) *Price of company's stock* is a ratio level of measurement. The data may be arranged in order, differences and ratios are meaningful, and a price of 0 is the starting point for all measurements.
   (c) *Names of new products* is a nominal level of measurement. The data consist of names only.
   (d) *Room temperature* is an interval level of measurement. The data may be arranged in order and differences are meaningful.
   (e) *Gross income* is a ratio level of measurement. The data may be arranged in order, differences and ratios are meaningful, and an income of 0 is the starting point for all measurements.
   (f) *Color of packaging* is a nominal level of measurement. The data consist of names only.

9. (a) *Species of fish* is a nominal level of measurement. Data consist of names only.
   (b) *Cost of rod and reel* is a ratio level of measurement. The data may be arranged in order. differences and ratios are meaningful, and a cost of 0 is the starting point for all measurements.
   (c) *Time of return home* is an interval level of measurement. The data may be arranged in order and differences are meaningful.
   (d) *Guidebook rating* is an ordinal level of measurement. Data may be arranged in order.
   (e) *Number of fish* caught is a ratio level of measurement. The data may be arranged in order, differences and ratios are meaningful. and 0 fish caught is the starting point for all measurements.
   (f) *Temperature of the water* is an interval level of measurement. The data may be arranged in order and differences are meaningful.

Section 1.2

1. Essay

2. Answers vary. Use groups of 3 digits.

3. Answers vary. Use groups of 4 digits.

4. Answers vary. Use groups of 3 digits.

5. (a) Assign a distinct number to each subject. Then use a random number table. Group assignment methods vary.
   (b) Repeat part (a) for 22 subjects.
   (c) Answers vary.

6. Answers vary. Use single digits with odd corresponding to heads and even to tails.

7. (a) Yes, it is appropriate that the same number appears more than once because the outcome of a die roll can repeat. The outcome of the 4th roll is 2.
   (b) No, we do not expect the same sequence because the process is random.

8. Answers vary. Use groups of 3 digits.

9. (a) Reasons may vary. For instance, the first four students may make a special effort to get to class on time.
   (b) Reasons may vary. For instance, four students who come in late might all be nursing students enrolled in an anatomy and physiology class that meets the hour before in a far-away building. They may be more motivated than other students to complete a degree requirement.
   (c) Reasons may vary. For instance, four students sitting in the back row might be less inclined to participate in class discussions.
   (d) Reasons may vary. For instance, the tallest students might all be male.

10. In all cases, assign distinct numbers to the items, and use a random-number table.

11. In all cases, assign distinct numbers to the items, and use a random-number table.

12. Answers vary. Use single digits with even corresponding to true and odd corresponding to false.

13. Answers vary. Use single digits with correct answer placed in corresponding position.

14. (a) This technique is stratified sampling. The population was divided into strata (4 categories of length of hospital stay), then a simple random sample was drawn from each stratum.
    (b) This technique is simple random sampling. Every sample of size n from the population has an equal chance of being selected and every member of the population has an equal chance of being included in the sample.
    (c) This technique is cluster sampling. There are 5 geographic regions and a random sample of hospitals is selected from each region. Then, for each selected hospital, all patients on the discharge list are surveyed to create the patient satisfaction profiles. Within each hospital, the degree of satisfaction varies patient to patient. The sampling units (the hospitals) are clusters of individuals who will be studied.
    (d) This technique is systematic sampling. Every k^{th} element is included in the sample.
    (e) This technique is convenience sampling. This technique uses results or data that are conveniently and readily obtained.

15. (a) This technique is simple random sampling. Every sample of size n from the population has an equal chance of being selected and every member of the population has an equal chance of being included in the sample.
    (b) This technique is cluster sampling. The state, Hawaii, is divided into regions using, say, the first 3 digits of the Zip code. Within each region a random sample of 10 Zip code areas is selected using, say, all 5 digits of the Zip code. Then, within each selected Zip codes, all businesses are surveyed. The sampling units, defined by 5 digit Zip codes, are clusters of businesses, and within each selected Zip code, the benefits package the businesses offer their employees differs business to business.
    (c) This technique is convenience sampling. This technique uses results or data that are conveniently and readily obtained.
    (d) This technique is systematic sampling. Every k^{th} element is included in the sample.
    (e) This technique is stratified sampling. The population was divided into strata (10 business types), then a simple random sample was drawn from each stratum.

Section 1.3

1. (a) This is an observational study because observations and measurements of individuals are conducted in a way that doesn't change the response or the variable being measured.
    (b) This is an experiment because a treatment is deliberately imposed on the individuals in order to observe a possible change in the response or variable being measured.
    (c) This is an experiment because a treatment is deliberately imposed on the individuals in order to observe a possible change in the response or variable being measured.
    (d) This is an observational study because observations and measurements of individuals are conducted in a way that doesn't change the response or the variable being measured.

2. (a) A census was used because data for all the games were used.
    (b) An experiment was used. A treatment is deliberately imposed on the individuals in order to observe change in the response or variable being measured.
    (c) A simulation was used because computer imaging of runners was used.
    (d) Sampling was used because measurements from a representative part of the population were used.

3. (a) Sampling was used because measurements from a representative part of the population were used.
    (b) A simulation was used because computer programs that mimic actual flight were used.
    (c) A census was used because data for all scores are available.
    (d) An experiment was used. A treatment is deliberately imposed on the individuals in order to observe change in the response or variable being measured.

4. (a) No, "over the last few years" could mean the last three years to some and the last five years to others, etc.; answers vary.
   (b) Yes. The response to doubling fines would be affected by whether the responder had ever run a stop sign.
   (c) Answers vary.

5. (a) Use random selection to pick 10 calves to inoculate. Then test all calves to see if there is a difference in resistance to infection between the two groups. There is no placebo being used.
   (b) Use random selection to pick 9 schools to visit. Then survey all the schools to see if there is a difference in views between the two groups. There is no placebo being used.
   (c) Use random selection to pick 40 volunteers for skin patch with drug. Then survey all volunteers to see if a difference exists between the two groups. A placebo for the remaining 35 volunteers in the second group is used.

6. (a) Use random selection to pick 25 cars for high-temperature bond tires. Then examine tires of all the cars to see if a difference exists between the two groups. This is a double-blind experiment because neither the individuals in the study nor the observers know which subjects are receiving the new tires.
   (b) Use random selection to pick 10 bags. Then send all bags through the security check. This is not a double-blind experiment because the agent carrying the bag knows whether or not the bag contains a weapon.
   (c) Use random selection to pick 35 patients for new eye drops. Then measure eye pressure for all patients to see if a difference exists between the two groups. This is a double-blind experiment because neither the patients nor the doctors know which subjects are receiving the new drops.

Chapter 1 Review

1. Answers vary.

2. The implied population is the opinions of all the listeners. The variable is the opinion of a caller. There is probably bias in the selection of the sample because those with the strongest opinions are most likely to call in.

3. Essay

4. Name, social security number, color of hair and eyes, address, phone number, place of birth, and college major are all nominal because the data consist of names or qualities only. Letter grade on test is ordinal because the data may be arranged in order. Year of birth is interval because the data may be arranged in order and differences are meaningful. Height, age, and distance from home to college are ratio because the data may be arranged in order, differences and ratios are meaningful, and 0 is the starting point for all measurements.

5. In the random number table use groups of 2 digits. Select the first six distinct groups of 2 digits that fall in the range from 01 to 42. Choices vary according to the starting place in the random number table.

6. (a) Cluster sampling was used because a random sample of 10 telephone prefixes was selected and all households in the selected prefixes were included in the sample.
   (b) Convenience sampling was used because it uses results or data that are conveniently and readily obtained.
   (c) Systematic sampling was used because every k^{th} element is included in the sample.
   (d) Random sampling was used because every sample of size 30 from the population has an equal chance of being selected and every member of the population has an equal chance of being included.

(e) Stratified sampling was used because the population was divided into strata (three age categories), then a simple random sample was drawn from each stratum.

7. (a) This is an observational study because observations and measurements of individuals are conducted in a way that doesn't change the response or the variable being measured.

(b) This is an experiment because a treatment is deliberately imposed on the individuals in order to observe a possible change in the response or variable being measured.

8. (a) Use random selection to pick half to solicit by mail. Then compute the percentage of donors in each group. Compare the results. No placebo was used.

(b) Use random selection to pick 43 volunteers to be given whitening gel. Evaluate tooth whiteness for all participants. Compare the results. A placebo was used with the remaining 42 in the second group. The experiment could be double-blind if the observers did not know which subjects were receiving the tooth whitening chemicals.

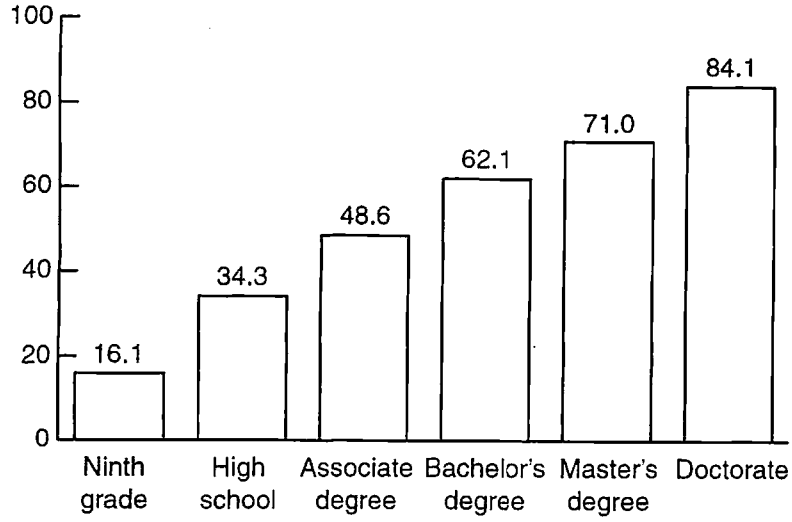
9. This is a good problem for class discussion. Some items such as age and grade point average might be sensitive information. You could ask the class to design a data form that can be filled out anonymously. Other issues to discuss involve the accuracy and honesty of the responses.

10. Students may easily spend several hours at this Web site.

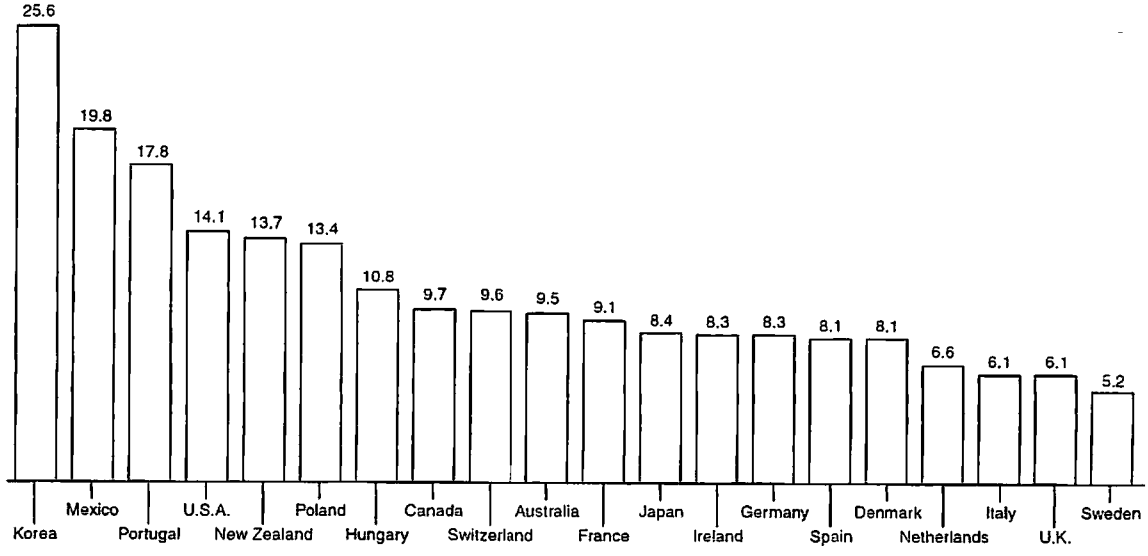
Chapter 2 Organizing Data

Section 2.1

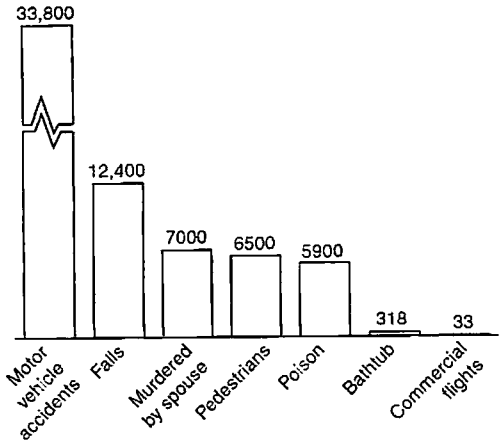
1. Highest Level of Education and Average Annual Household Income (in thousands of dollars)



2. Annual Number of Deaths from Injuries per 100,000 Children (Ages 1 to 14)

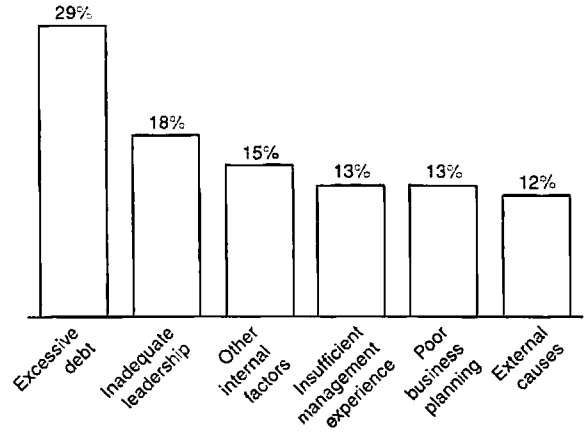


3. Number of People Who Died in a Calendar Year from Listed Causes—Pareto Chart



4. (a) Since 88% of those surveyed cited internal problems, 100% – 88% = 12% cited external factors as the leading cause of business failure. Among the internal causes, 88% – 13% – 13% – 18% – 29% = 15% must have listed various other internal factors for the leading cause of business failure.

Causes for Business Failure—Pareto Chart



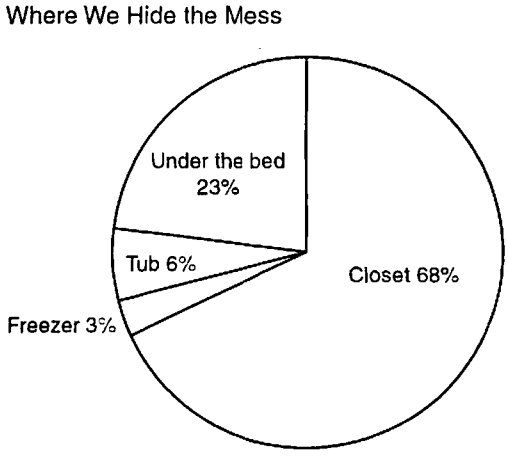
(b) As shown in part (a), 15% of those interviewed cited other internal factors as the leading cause of business failure. Excessive debt was the most commonly cited (internal and overall) cause for business failure.

Cause of Business Failure	Percentage	Frequency
Insufficient management experiences	13%	13% × 1300 = 169
Poor business planning	13%	13% × 1300 = 169
Inadequate leadership	18%	18% × 1300 = 234
Excessive debt	29%	29% × 1300 = 377
Other internal factors	15%	15% × 1300 = 195
External factors	12%	12% × 1300 = 156
Total	100%	1300

5.

Hiding place	Percentage	Number of Degrees
In the closet	68%	$68\% \times 360° \approx 245°$
Under the bed	23%	$23\% \times 360° \approx 83°$
In the bathtub	6%	$6\% \times 360° \approx 22°$
In the freezer	3%	$3\% \times 360° \approx 11°$
Total	100%	361°*

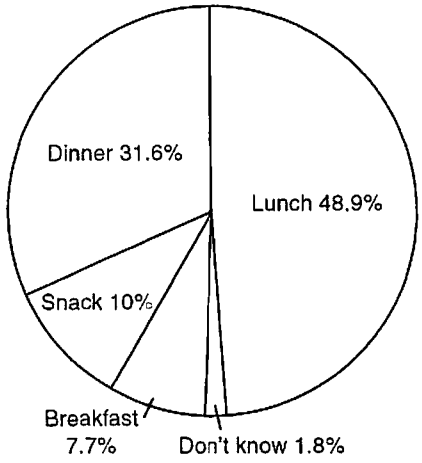
*Total does not add to 360° due to rounding.



6.

Meal	Percentage	Number of Degrees
Lunch	48.9%	$48.9\% \times 360° \approx 176°$
Breakfast	7.7%	$7.7\% \times 360° \approx 28°$
Dinner	31.6%	$31.6\% \times 360° \approx 114°$
Snack	10.0%	$10.0\% \times 360° = 36°$
Don't know	1.8%	$1.8\% \times 360° \approx 6°$
Total	100.0%	360°

Meals We Are Most Likely to Eat in a Fast-Food Restaurant

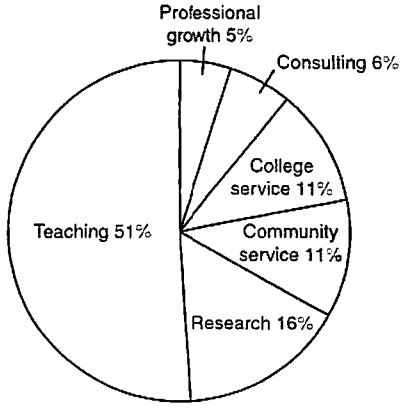


7.

Professional Activity	Percentage	Number of Degrees
Teaching	51%	$51\% \times 360^\circ \approx 184^\circ$
Research	16%	$16\% \times 360^\circ \approx 58^\circ$
Professional growth	5%	$5\% \times 360^\circ = 18^\circ$
Community service	11%	$11\% \times 360^\circ \approx 40^\circ$
Service to the college	11%	$11\% \times 360^\circ \approx 40^\circ$
Consulting outside the college	6%	$6\% \times 360^\circ \approx 22^\circ$
Total	100%	362°*

* Total does not add to 360° due to rounding.

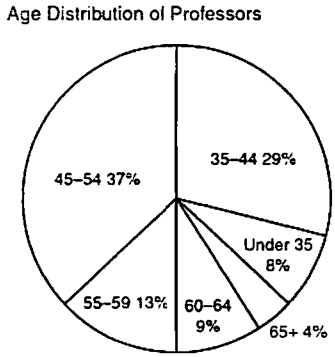
How College Professors Spend Time



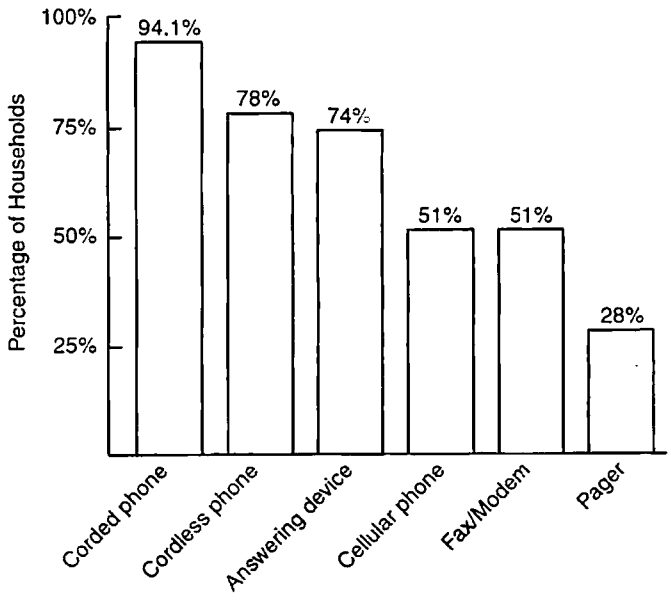
8.

Age	Percentage	Number of Degrees
Under 35 years	8%	8% × 360° ≈ 29°
35–44 years	29%	29% × 360° ≈ 104°
45–54 years	37%	37% × 360° ≈ 133°
55–59 years	13%	13% × 360° ≈ 47°
60–64 years	9%	9% × 360° ≈ 32°
65 years and over	4%	4% × 360° ≈ 14°
Total	100%	359°*

*Total does not add to 360° due to rounding.

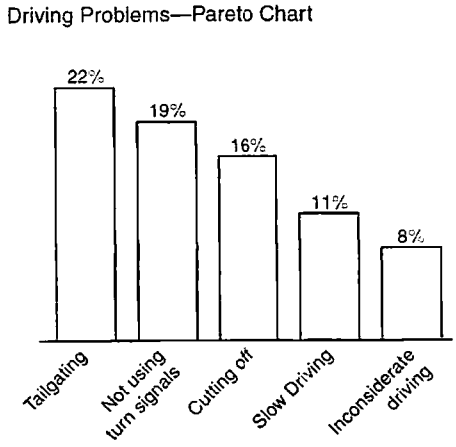


9. Percentage of Households with Telephone Gadgets



No. Since household can report having more than one telephone gadget, the percentages will not add to 100%.

10. The following Pareto Chart shows the percentage of drivers for each stated complaint.

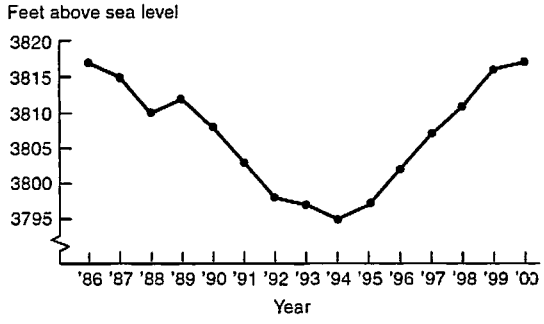


By subtraction, 100% – 22% – 19% – 16% – 11% – 8% = 24% of the respondents citied other bad habits.

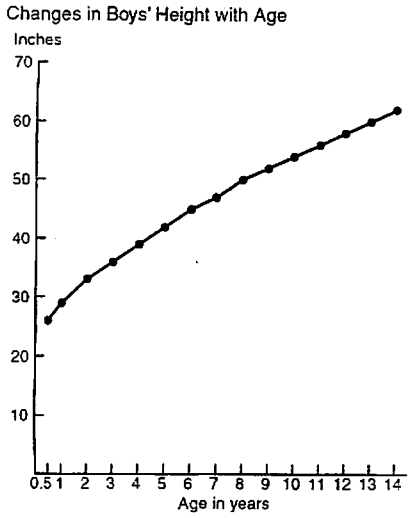
Bad Habit	Percentage	Frequency
Tailgating	22%	22% × 500 = 110
Not using turn signals	19%	19% × 500 = 95
Cutting off other drivers	16%	16% × 500 = 80
Driving too slowly	11%	11% × 500 = 55
Being inconsiderate	8%	8% × 500 = 40
Other	24%	24% × 500 = 120
Total	100%	500

As reported, the percentages add to 76%, not the 100% needed for a circle graph. However, if there was only one response per person, knowing that the company surveyed 500 drivers tells us that 120 drivers, or 24%, had other bad driving complaints. Using this fact, a circle graph could be used.

11. Elevation of Pyramid Lake Surface—Time Plot



12.



13. Both stocks ended down for the one-year period. Coca-Cola ranged from a high of about \$64 to a low of about \$42.50. McDonald's ranged from a high of about \$37 to a low of about \$25.McDonald's attained its high during the first week and was never as high again. From the high of the first week to the high of the last week shown, we can calculate the approximate percentage change in price by finding the difference in the high values and dividing that by the 6/9/00 high. Thus, Coca-Cola declined $\frac{54-47}{54} \approx 13\%$ while McDonald's declined $\frac{36-29}{36} \approx 19\%$.

Append: Volatility is shown by the moving average lines: the smoother the line, the less volatile the stock price. The 200-day moving average will always be smoother than the 50-day moving average, and it is less sensitive than the 50-day moving average to abrupt changes. Coke's 200-day moving average does not yet register the steep drop in price during the February to April period. Both moving average lines are smoother for McDonald's indicating its price is less volatile.

Both McDonald's and Coca-Cola's week-to-week-patterns are similar, rising and dropping over approximately the same periods. This, coupled with the DJIA's increase of 2.3% over the same year, may reflect some factor affecting the fast food/soft drink sector that does not impact the market as a whole. For example, a rise in gasoline prices would depress the fast food and soft drink sectors as well as the overall market, by increasing the cost of transporting food and reducing disposable income overall. Impacted sectors would become less profitable, and their stock values would decrease.

Section 2.2

1. (a) largest data value = 360

smallest data value = 236

number of classes specified = 5

class width $= \frac{360-236}{5} = 24.8$, increased to next whole number, 25

(b) The lower class limit of the first class in the smallest value, 236.

The lower class limit of the next class is the previous class's lower class limit plus the class width; for the second class, this is 236 + 25 = 261.

The upper class limit is one value less than lower class limit of the next class: for the first class, the upper class limit is 261 – 1 = 260.

The class boundaries are the halfway points between (i.e., the average of) the (adjacent) upper class limit of one class and the lower class limit of the next class. The lower class boundary of the first class is the lower class limit minus one-half unit. The upper class boundary for the last class is the upper class limit plus one-half unit. For the first class, the class boundaries are $236-\frac{1}{2}=235.5$ and $\frac{260+261}{2}=260.5$. For the last class, the class boundaries are $\frac{335+336}{2}=335.5$ and $360+\frac{1}{2}=360.5$.

The class mark or midpoint is the average of the class limits for that class. For the first class, the midpoint is $\frac{236+260}{2}=248$.

The class frequency is the number of data values that belong to that class; call this value f.

The relative frequency of a class is the class frequency, f, divided by the total number of data values, i.e., the overall sample size, n.

For the first class, $f = 4$, $n = 57$, and the relative frequency is $f/n = \frac{4}{57} \approx 0.07$.

The cumulative frequency of a class is the sum of the frequencies for all previous classes, plus the frequency of that class. For the first and second classes, the class cumulative frequencies are 4 and 4 + 9 = 13, respectively.

Class Limits	Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
236–260	235.5–260.5	248	4	0.07	4
261–285	260.5–285.5	273	9	0.16	13
286–310	285.5–310.5	298	25	0.44	38
311–335	310.5–335.5	323	16	0.28	54
336–360	335.5–360.5	348	3	0.05	57

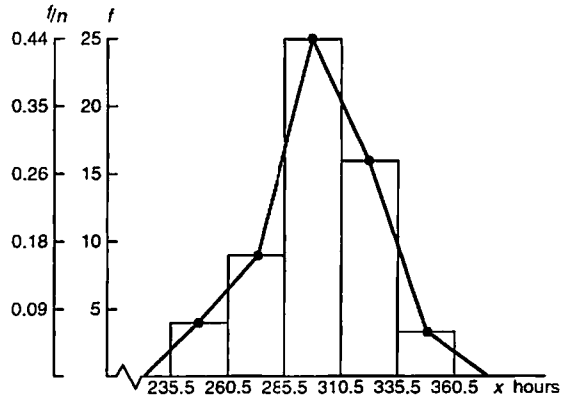
(c) The histogram plots the class frequencies on the y-axis and the class boundaries on the x-axis. Since adjacent classes share boundary values, the bars touch each other. [Alternatively, the bars may be centered over the class marks (midpoints).]

(d) A frequency polygon connects the midpoints of each class (shown as a dot in the middle of the top of the histogram bar) with line segments. Place a dot on the x-axis one class width below the midpoint of the first class, and place another dot on the x-axis one class width above the last class's midpoint. Connect these dots to the adjacent midpoint dots with line segments.

(e) The relative frequency histogram is exactly the same shape as the frequency histogram, but the vertical scale is relative frequency, f/n, instead of actual frequency, f.

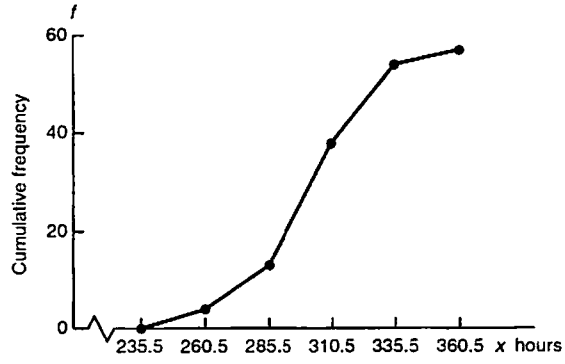
The following figure shows the histogram, frequency polygon, and relative-frequency histogram for (c), (d), and (e) above, overlaying one another. (Note that two vertical scales are shown.)

Hours to Complete the Iditarod—Histogram, Frequency Polygon, Relative-Frequency Histogram



(f) To create the ogive, place a dot on the x-axis at the lower class boundary of the first class and then, for each class, place a dot above the upper class boundary value at the height of the cumulative frequency for the class. Connect the dots with line segments.

Hours to Complete Iditarod—Ogive



2. (a) largest data value = 65

smallest data value = 20

number of classes specified = 5

class width $= \dfrac{65-20}{5} = 9$, increased to next whole number, 10

(b) The lower class limit of the first class in the smallest value, 20.

The lower class limit of the next class is the previous class's lower class limit plus the class width: for the second class, this is 20 + 10 = 30.

The upper class limit is one value less than lower class limit of the next class; for the first class, the upper class limit is 30 – 1 = 29.

The class boundaries are the halfway points between (i.e., the average of) the (adjacent) upper class limit of one class and the lower class limit of the next class. The lower class boundary of the first class is the lower class limit minus one-half unit. The upper class boundary for the last class is the upper class limit plus one-half unit. For the first class, the class boundaries are $20-\frac{1}{2}=19.5$ and $\frac{29+30}{2}=29.5$. For the last class, the class boundaries are $\frac{59+60}{2}=59.5$ and $69+\frac{1}{2}=69.5$.

The class mark or midpoint is the average of the class limits for that class. For the first class, the midpoint is $\frac{20+29}{2}=24.5$.

The class frequency is the number of data values that belong to that class; call this value f.

The relative frequency of a class is the class frequency, f, divided by the total number of data values, i.e., the overall sample size, n.

For the first class, $f = 3$. $n = 35$, and the relative frequency is $f/n = 3/35 \approx 0.0857$.

The cumulative frequency of a class is the sum of the frequencies for all previous classes, plus the frequency of that class. For the first and second classes, the class cumulative frequencies are 3 and 3 + 6 = 9, respectively.

Percent Difficult Ski Terrain

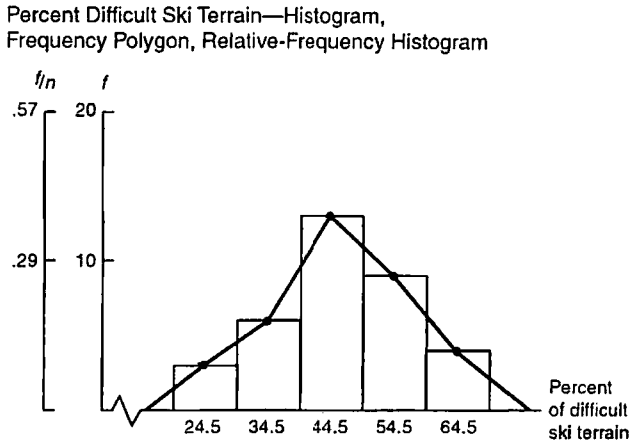
Class Limits	Class Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
20–29	19.5–29.5	24.5	3	0.0857	3
30–39	29.5–39.5	34.5	6	0.1714	9
40–49	39.5–49.5	44.5	13	0.3714	22
50–59	49.5–59.5	54.5	9	0.2571	31
60–69	59.5–69.5	64.5	4	0.1143	35

(c) The histogram plots the class frequencies on the y-axis and the class boundaries on the x-axis. Since adjacent classes share boundary values, the bars touch each other. [Alternatively, the bars may be centered over the class marks (midpoints).]

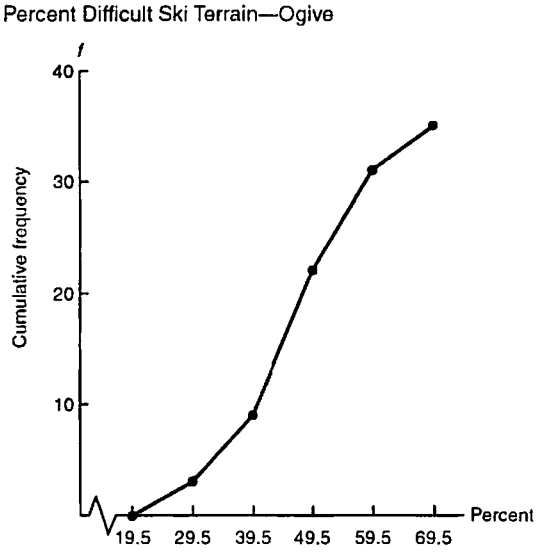
(d) A frequency polygon connects the midpoints of each class (shown as a dot in the middle of the top of the histogram bar) with line segments. Place a dot on the x-axis one class width below the midpoint of the first class, and place another dot on the x-axis one class width above the last class's midpoint. Connect these dots to the adjacent midpoint dots with line segments.

(e) The relative frequency histogram is exactly the same shape as the frequency histogram, but the vertical scale is relative frequency, f/n, instead of actual frequency, f.

The following figure shows the histogram, frequency polygon, and relative-frequency histogram for (c), (d), and (e) above, overlaying one another. (Note that two vertical scales are shown.)



(f) To create the ogive, place a dot on the x-axis at the lower class boundary of the first class and then, for each class, place a dot above the upper class boundary value at the height of the cumulative frequency for the class. Connect the dots with line segments.



3. (a) largest data value = 53

smallest data value = 5

number of classes specified = 7

class width $= \dfrac{53-5}{7} \approx 6.86$, increased to next whole number, 7

(b) The lower class limit of the first class in the smallest value, 5.

The lower class limit of the next class is the previous class's lower class limit plus the class width; for the second class, this is 5 + 7 = 12.

The upper class limit is one value less than lower class limit of the next class; for the first class, the upper class limit is 12 – 1 = 11.

The class boundaries are the halfway points between (i.e., the average of) the (adjacent) upper class limit of one class and the lower class limit of the next class. The lower class boundary of the first class is the lower class limit minus one-half unit. The upper class boundary for the last class is the upper class limit plus one-half unit. For the first class, the class boundaries are $5-\frac{1}{2}=4.5$ and $\frac{11+12}{2}=11.5$. For the last class, the class boundaries are $\frac{46+47}{2}=46.5$ and $53+\frac{1}{2}=53.5$.

The class mark or midpoint is the average of the class limits for that class. For the first class, the midpoint is $\frac{5+11}{2}=8$.

The class frequency is the number of data values that belong to that class; call this value f.

The relative frequency of a class is the class frequency, f, divided by the total number of data values, i.e., the overall sample size, n.

For the first class, $f = 4$, $n = 50$, and the relative frequency is $f/n = 4/50 = 0.08$.

The cumulative frequency of a class is the sum of the frequencies for all previous classes, plus the frequency of that class. For the first and second classes, the class cumulative frequencies are 4 and 4 + 7 = 11, respectively.

Class Limits	Class Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
5–11	4.5–11.5	8	4	0.08	4
12–18	11.5–18.5	15	7	0.14	11
19–25	18.5–25.5	22	12	0.24	22
26–32	25.5–32.5	29	12	0.24	35
33–39	32.5–39.5	36	12	0.24	47
40–46	39.5–46.5	43	2	0.04	49
47–53	46.5–53.5	50	1	0.02	50

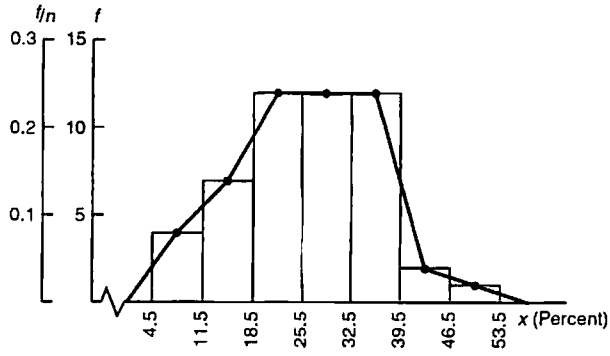
(c) The histogram plots the class frequencies on the y-axis and the class boundaries on the x-axis. Since adjacent classes share boundary values, the bars touch each other. [Alternatively, the bars may be centered over the class marks (midpoints).]

(d) A frequency polygon connects the midpoints of each class (shown as a dot in the middle of the top of the histogram bar) with line segments. Place a dot on the x-axis one class width below the midpoint of the first class, and place another dot on the x-axis one class width above the last class's midpoint. Connect these dots to the adjacent midpoint dots with line segments.

(e) The relative frequency histogram is exactly the same shape as the frequency histogram, but the vertical scale is relative frequency, f/n, instead of actual frequency, f.

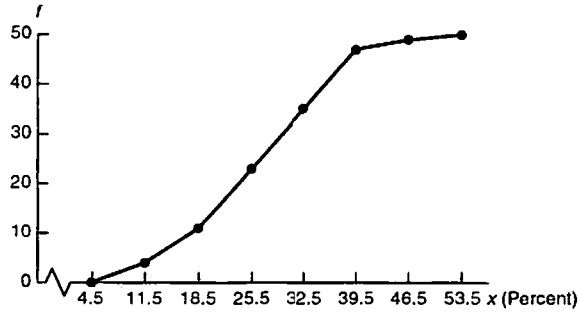
The following figure shows the histogram, frequency polygon, and relative-frequency histogram for (c), (d), and (e) above, overlaying one another. (Note that two vertical scales are shown.)

Percentage of Children in Neighorhood—Histogram, Frequency Polygon, Relative-Frequency Histogram



(f) To create the ogive, place a dot on the x-axis at the lower class boundary of the first class and then, for each class, place a dot above the upper class boundary value at the height of the cumulative frequency for the class. Connect the dots with line segments.

Percentage of Children in Neighborhood—Ogive



4. (a) largest data value = 75

smallest data value = 5

number of classes specified = 5

class width $= \dfrac{75-5}{5} = 14$, increased to next whole number, 15

(b) The lower class limit of the first class in the smallest value, 5.

The lower class limit of the next class is the previous class's lower class limit plus the class width; for the second class, this is 5 + 15 = 20.

The upper class limit is one value less than lower class limit of the next class; for the first class, the upper class limit is 20 – 1 = 19.

The class boundaries are the halfway points between (i.e., the average of) the (adjacent) upper class limit of one class and the lower class limit of the next class. The lower class boundary of the first class is the lower class limit minus one-half unit. The upper class boundary for the last class is the upper class limit plus one-half unit. For the first class, the class boundaries are $5-\frac{1}{2}=4.5$ and $\frac{19+20}{2}=19.5$. For the last class, the class boundaries are $\frac{64+65}{2}=64.5$ and $79+\frac{1}{2}=79.5$.

The class mark or midpoint is the average of the class limits for that class. For the first class, the midpoint is $\frac{5+19}{2}=12$.

The class frequency is the number of data values that belong to that class; call this value f.

The relative frequency of a class is the class frequency, f, divided by the total number of data values, i.e., the overall sample size, n.

For the first class, $f = 21$, $n = 63$, and the relative frequency is $f/n = 21/63 \approx 0.3333$.

The cumulative frequency of a class is the sum of the frequencies for all previous classes, plus the frequency of that class. For the first and second classes, the class cumulative frequencies are 21 and 21 + 35 = 56, respectively.

Fast-Food Franchise Fees (in thousands)

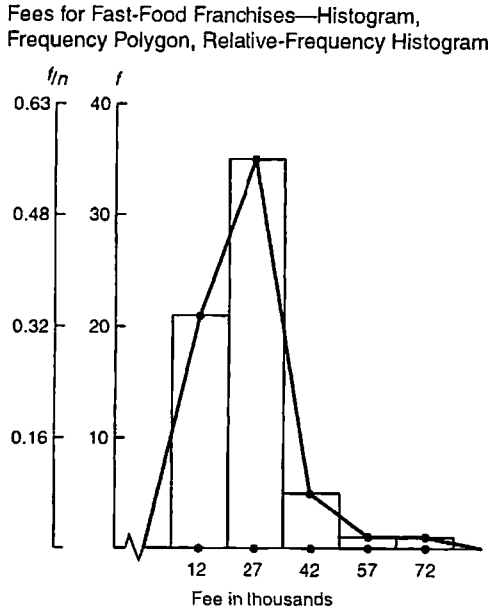
Class Limits	Class Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
5–19	4.5–19.5	12	21	0.3333	21
20–34	19.5–34.5	27	35	0.5556	56
35–49	34.5–49.5	42	5	0.0794	61
50–64	49.5–64.5	57	1	0.0159	62
65–79	64.5–79.5	72	1	0.0159	63

(c) The histogram plots the class frequencies on the y-axis and the class boundaries on the x-axis. Since adjacent classes share boundary values, the bars touch each other. [Alternatively, the bars may be centered over the class marks (midpoints).]

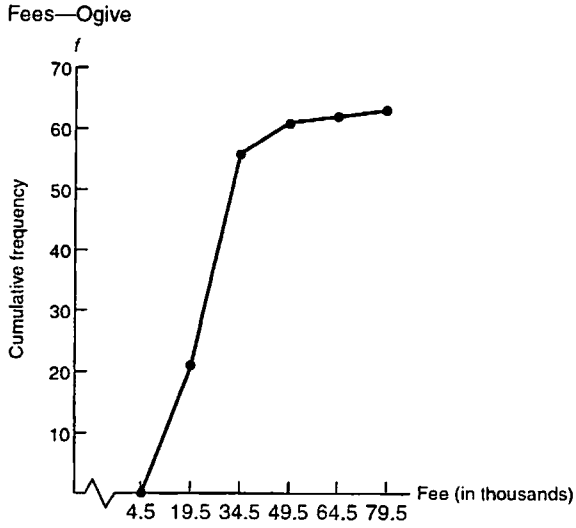
(d) A frequency polygon connects the midpoints of each class (shown as a dot in the middle of the top of the histogram bar) with line segments. Place a dot on the x-axis one class width below the midpoint of the first class, and place another dot on the x-axis one class width above the last class's midpoint. Connect these dots to the adjacent midpoint dots with line segments.

(e) The relative frequency histogram is exactly the same shape as the frequency histogram, but the vertical scale is relative frequency, f/n, instead of actual frequency, f.

The following figure shows the histogram, frequency polygon, and relative-frequency histogram for (c), (d), and (e) above, overlaying one another. (Note that two vertical scales are shown.)



(f) To create the ogive, place a dot on the x-axis at the lower class boundary of the first class and then, for each class, place a dot above the upper class boundary value at the height of the cumulative frequency for the class. Connect the dots with line segments.



5. (a) largest data value = 102

smallest data value = 18

number of classes specified = 5

class width $= \dfrac{102-18}{5} = 16.8$, increased to next whole number, 17

(b) The lower class limit of the first class in the smallest value, 18.

The lower class limit of the next class is the previous class's lower class limit plus the class width; for the second class, this is 18 + 17 = 35.

The upper class limit is one value less than lower class limit of the next class; for the first class, the upper class limit is 35 – 1 = 34.

The class boundaries are the halfway points between (i.e., the average of) the (adjacent) upper class limit of one class and the lower class limit of the next class. The lower class boundary of the first class is the lower class limit minus one-half unit. The upper class boundary for the last class is the upper class limit plus one-half unit. For the first class, the class boundaries are $18-\frac{1}{2}=17.5$ and $\frac{34+35}{2}=34.5$. For the last class, the class boundaries are $\frac{85+86}{2}=85.5$ and $102+\frac{1}{2}=102.5$.

The class mark or midpoint is the average of the class limits for that class. For the first class, the midpoint is $\frac{18+34}{2}=26$.

The class frequency is the number of data values that belong to that class; call this value f.

The relative frequency of a class is the class frequency, f, divided by the total number of data values, i.e., the overall sample size, n.

For the first class, $f = 1$, $n = 35$, and the relative frequency is $f/n = 1/35 \approx 0.03$.

The cumulative frequency of a class is the sum of the frequencies for all previous classes, plus the frequency of that class. For the first and second classes, the class cumulative frequencies are 1 and 1 + 2 = 3, respectively.

Number of Room Calls per Night

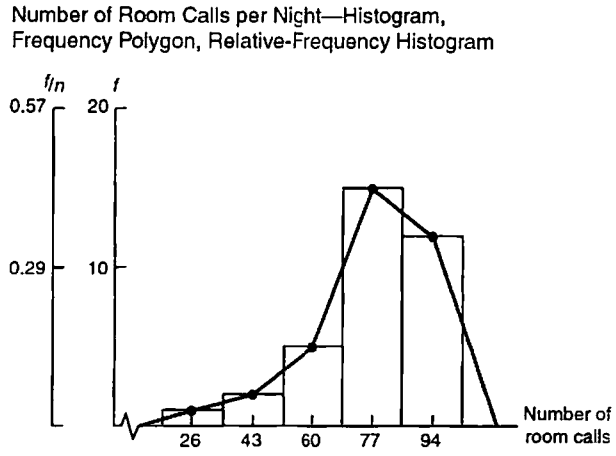
Class Limits	Class Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
18–34	17.5–34.5	26	1	0.03	1
35–51	34.5–51.5	43	2	0.06	3
52–68	51.5–68.5	60	5	0.14	8
69–85	68.5–85.5	77	15	0.43	23
86–102	85.5–102.5	94	12	0.34	35

(c) The histogram plots the class frequencies on the y-axis and the class boundaries on the x-axis. Since adjacent classes share boundary values, the bars touch each other. [Alternatively, the bars may be centered over the class marks (midpoints).]

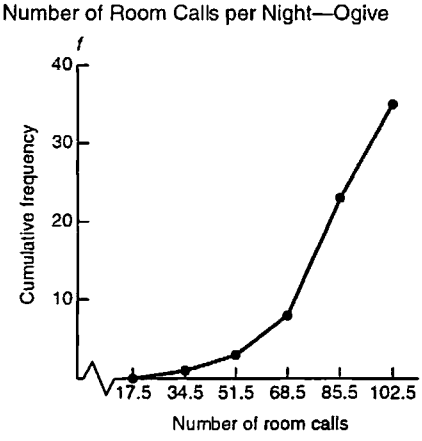
(d) A frequency polygon connects the midpoints of each class (shown as a dot in the middle of the top of the histogram bar) with line segments. Place a dot on the x-axis one class width below the midpoint of the first class, and place another dot on the x-axis one class width above the last class's midpoint. Connect these dots to the adjacent midpoint dots with line segments.

(e) The relative frequency histogram is exactly the same shape as the frequency histogram. but the vertical scale is relative frequency, f/n, instead of actual frequency, f.

The following figure shows the histogram, frequency polygon, and relative-frequency histogram for (c), (d), and (e) above, overlaying one another. (Note that two vertical scales are shown.)



(f) To create the ogive, place a dot on the x-axis at the lower class boundary of the first class and then, for each class, place a dot above the upper class boundary value at the height of the cumulative frequency for the class. Connect the dots with line segments.



6. (a) largest data value = 43

smallest data value = 0

number of classes specified = 8

class width $= \dfrac{43-0}{8} = 5.375$, increased to next whole number, 6

(b) The lower class limit of the first class in the smallest value, 0.

The lower class limit of the next class is the previous class's lower class limit plus the class width; for the second class, this is 0 + 6 = 6.

The upper class limit is one value less than lower class limit of the next class; for the first class, the upper class limit is 6 – 1 = 5.

The class boundaries are the halfway points between (i.e., the average of) the (adjacent) upper class limit of one class and the lower class limit of the next class. The lower class boundary of the first class is the lower class limit minus one-half unit. The upper class boundary for the last class is the upper class limit plus one-half unit. For the first class, the class boundaries are $0-\frac{1}{2}=-0.5$ and $\frac{5+6}{2}=5.5$.

For the last class, the class boundaries are $\frac{41+42}{2}=41.5$ and $47+\frac{1}{2}=47.5$.

The class mark or midpoint is the average of the class limits for that class. For the first class, the midpoint is $\frac{0+5}{2}=2.5$.

The class frequency is the number of data values that belong to that class: call this value f.

The relative frequency of a class is the class frequency, f, divided by the total number of data values, i.e., the overall sample size, n.

For the first class, $f = 13$, $n = 55$, and the relative frequency is $f/n = 13/55 \approx 0.24$.

The cumulative frequency of a class is the sum of the frequencies for all previous classes, plus the frequency of that class. For the first and second classes, the class cumulative frequencies are 13 and 13 + 15 = 28, respectively.

Words of Three Syllables or More

Class Limits	Class Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
0–5	0.5–5.5	2.5	13	0.24	13
6–11	5.5–11.5	8.5	15	0.27	28
12–17	11.5–17.5	14.5	11	0.20	39
18–23	17.5–23.5	20.5	3	0.05	42
24–29	23.5–29.5	26.5	6	0.11	48
30–35	29.5–35.5	32.5	4	0.07	52
36–41	35.5–41.5	38.5	2	0.04	54
42–47	41.5–47.5	44.5	1	0.02	55

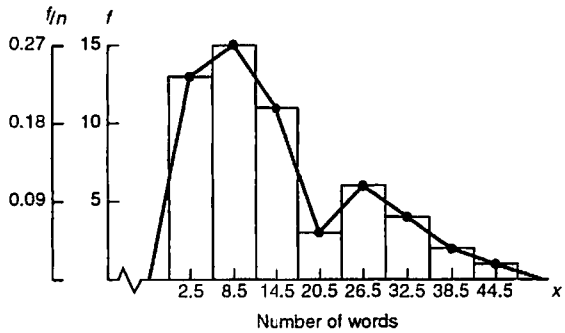
(c) The histogram plots the class frequencies on the y-axis and the class boundaries on the x-axis. Since adjacent classes share boundary values, the bars touch each other. [Alternatively, the bars may be centered over the class marks (midpoints).]

(d) A frequency polygon connects the midpoints of each class (shown as a dot in the middle of the top of the histogram bar) with line segments. Place a dot on the x-axis one class width below the midpoint of the first class, and place another dot on the x-axis one class width above the last class's midpoint. Connect these dots to the adjacent midpoint dots with line segments.

(e) The relative frequency histogram is exactly the same shape as the frequency histogram, but the vertical scale is relative frequency, f/n, instead of actual frequency, f.

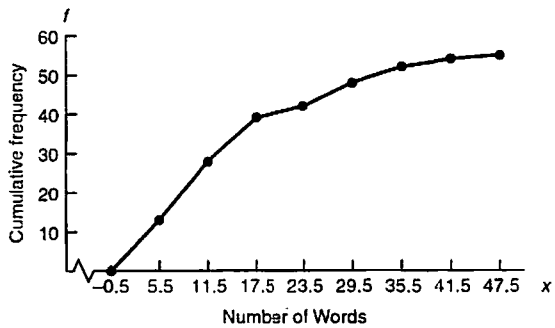
The following figure shows the histogram, frequency polygon, and relative-frequency histogram for (c), (d), and (e) above, overlaying one another. (Note that two vertical scales are shown.)

Words of Three Syllables or More—Histogram, Frequency Polygon, Relative-Frequency Histogram



(f) To create the ogive, place a dot on the x-axis at the lower class boundary of the first class and then, for each class, place a dot above the upper class boundary value at the height of the cumulative frequency for the class. Connect the dots with line segments.

Words of Three Syllables or More—Ogive

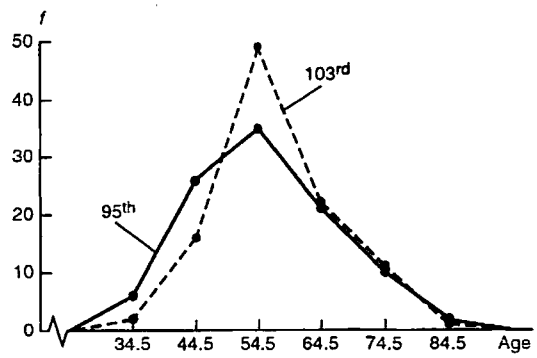


7. (a) The class midpoint is the average of the class limits for that class.

Class Midpoints: 34.5; 44.5; 54.5; 64.5; 74.5; 84.5.

(b) The frequency polygon connects to the x-axis one class width below the smallest midpoint and one class width above the largest midpoint; here the class width is 10, so we have points on the x-axis at $34.5 - 10 = 24.5$ and $84.5 + 10 = 94.5$.

Age of Senators—Frequency Polygons



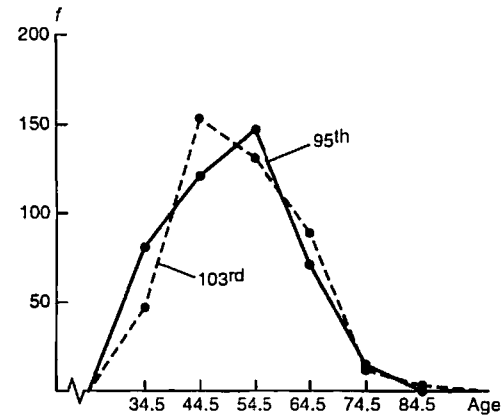
(c) The two polygons have the same general shape, but the dashed polygon is shifted slightly to the right (older ages). so. in general. the members of the 103rd Congress are older.

8. (a) The class midpoint is the average of the class limits for that class.

Class Midpoints: 34.5; 44.5; 54.5; 64.5; 74.5; 84.5.

(b) The frequency polygon connects to the x-axis one class width below the smallest midpoint and one class width above the largest midpoint; here the class width is 10, so we have points on the x-axis at $34.5 - 10 = 24.5$ and $84.5 + 10 = 94.5$.

Ages of Representatives—Frequency Polygons



(c) The age distribution shapes are similar. The 95th Congress members of the House have more people in their 30s and (to a lesser extent) 50s, but fewer in their 40s and (to a lesser extent) 60s. There is essentially no difference in frequencies for members over 70.

9. (a)

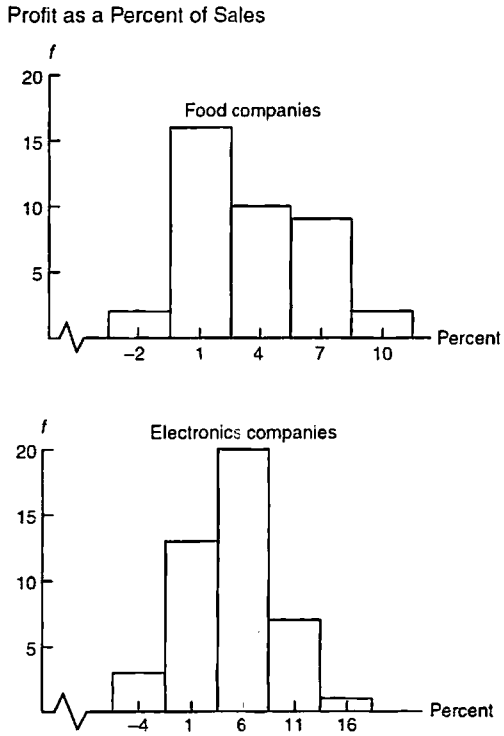
	Largest value	Smallest value	Class width
Food Companies	11	−3	$\frac{11-(-3)}{5} = 2.8$; use 3
Electronic Companies	16	−6	$\frac{16-(-6)}{5} = 4.4$; use 5

Profit as Percent of Sales—Food Companies

Class	Frequency	Midpoint
−3 to −1	2	−2
0–2	16	1
3–5	10	4
6–8	9	7
9–11	2	10

Profit as Percent of Sales—Electronic Companies

Class	Frequency	Midpoint
−6 to −2	3	−4
−1 to 3	13	1
4–8	20	6
9–13	7	11
14–18	1	16



(b) Because the classes and class widths are different for the two company types, it is difficult to compare profits as a percentage of sales. We can notice that for the electronic companies the 16 profits as a percentage of sales extends as high as 18, while for the food companies the highest profit as a percentage of sales is 11. On the other hand, some of the electronic companies also have greater losses than the food companies. Had we made the class limits the same for both company types and overlaid the histograms, it would be easier to compare the data.

10. (a)

	Largest value	Smallest value	Class width
Miami Dolphins	295	175	$\frac{295-175}{6} = 20$: use 21
San Diego Charges	310	119	$\frac{310-119}{6} \approx 31.8$: use 32

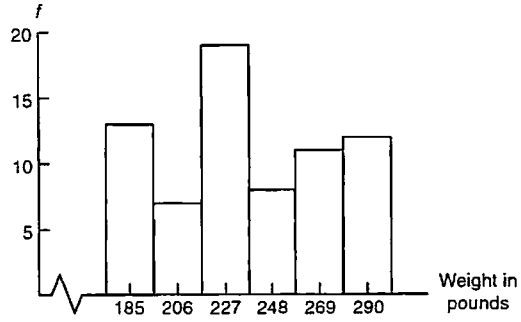
Weights of Football Players:
Miami Dolphins

Class	Midpoint	Frequency
175–195	185	13
196–216	206	7
217–237	227	19
238–258	248	8
259–279	269	11
280–300	290	12

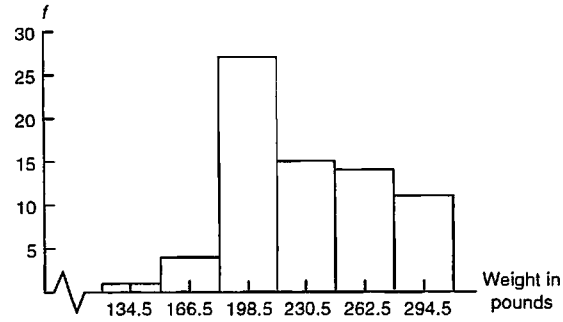
San Diego Chargers

Class	Midpoint	Frequency
119–150	134.5	1
151–182	166.5	4
183–214	198.5	27
215–246	230.5	15
247–278	262.5	14
279–310	294.5	11

Weights of Football Players—Miami Dolphins



Weights of Football Players—San Diego Chargers



(b) Because the class widths are different, it is difficult to compare the histograms. However, San Diego has 4 players who are smaller than the smallest Miami player, and 4 players who are larger than the largest player.

It would be easier to compare the teams' weights if the histograms had common classes and were overlaid.

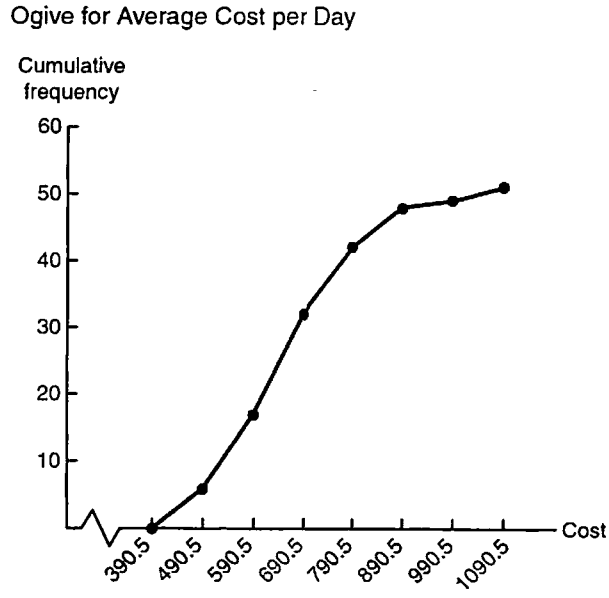
11. (a) Since ogives show the cumulative frequency at the upper class boundary, and begin at the point with (x, y) coordinates (lower class boundary of the first class, 0), the numbers on the x-axis are class boundaries. Recall that class boundary values are not values the data can attain. Thus the point marked 85 over the x-value 7.15 means that 85 winning times were less than or equal to 7.15 and, since 7.15 is not a possible data value, 85 winning times were less than 7.15 (i.e., less than 2 minutes 7.15 seconds). Eighty-five of 101 times are less than 7.15, or $\frac{85}{101} \approx 84.2\%$.

(b) Subtract the cumulative frequency at or below 5.15 seconds from the cumulative frequency at or below 11.15 seconds (which includes all the values at or below 5.15 seconds) to get the number of winning times between 5.15 and 11.15 seconds/over two minutes): $100 - 75 = 25$, or $\frac{25}{101} \approx 24.8\%$.

12. (a) Since the values at the edges of the bars on the histogram are shown, these are class boundaries.

Class	Frequency	Cumulative Frequency
390.5–490.5	6	6
490.5–590.5	11	17
590.5–690.5	15	32
690.5–790.5	10	42
790.5–890.5	6	48
890.5–990.5	1	49
990.5–1090.5	2	51

Begin the ogive at (390.5, 0) (i.e., at the lower class boundary of the first class) and plot the values (upper class boundary, cumulative frequency).



(b) From the ogive point (690.5, 32), (or the table above) we have that 32 of the "51 states" (states plus D.C.) have an average cost per day per patient less than $690.50.

13. (a) Uniform is rectangular, symmetric looks like mirror images on each side of the middle, bimodal has two modes (peaks), and skewed distributions have long tails on one side, and are skewed in the direction of the tail ("skew, few"). (Note that uniform distributions are also symmetric, but "uniform" is more descriptive.)

(a) skewed left; (b) uniform, (c) symmetric, (d) bimodal, (e) skewed right.

(b) Answers vary. Students would probably like (a) since there are many high scores and few low scores. Students would probably dislike (e) since there are few high scores but lots of lows scores. (b) is designed to give approximately the same number of As, Bs, etc. (d) has more Bs and Ds, say. (c) is the way many tests are designed: As and Fs for the exceptionally high and low scores with most students receiving Cs.

14. **(a)** Uniform is rectangular, symmetric looks like mirror images on each side of the middle, bimodal has two modes (peaks), and skewed distributions have long tails on one side, and are skewed in the direction of the tail ("skew, few"). (Note that uniform distributions are also symmetric, but "uniform" is more descriptive.)

(a) uniform, (b) skewed right, (c) bimodal, (d) bimodal, (e) symmetric. [Note that (c) has a major and a minor mode. "Tails" in a distribution's shape "tail off," i.e., get thinner, and do not have "bumps" in them as (c) does.]

(b) Answers vary. Ads should target the largest number of potential buyers, so ads should be aimed at the income levels with the greatest concentration (frequency) of households.

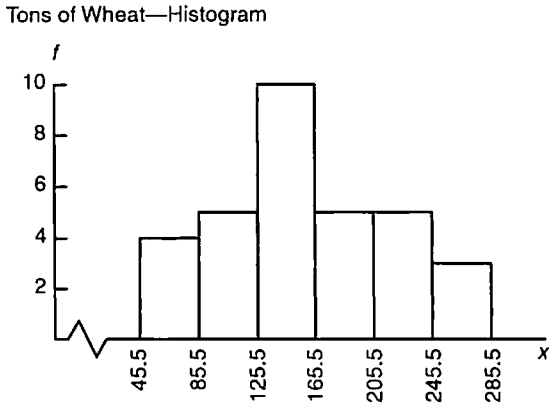
(c) Answers vary. Since warranty/registration cards are returned voluntarily, the income data are most likely not representative of the buying public in general, and probably are not even representative of those buying the specific product. Also, people tend to inflate their income levels on most forms, except those sent to the IRS.

15. **(a)** $2.71 \times 100 = 271$, $1.62 \times 100 = 162$,..., $0.70 \times 100 = 70$.

(b) largest value = 282, smallest value = 46

$$\text{class width} = \frac{282 - 46}{6} \approx 39.3; \text{ use } 40$$

Class Limits	Class Boundaries	Midpoint	Frequency
46–85	45.5–85.5	65.5	4
86–125	85.5–125.5	105.5	5
126–165	125.5–165.5	145.5	10
166–205	165.5–205.5	185.5	5
206–245	205.5–245.5	225.5	5
246–285	245.5–285.5	265.5	3



(c) class width is $\frac{40}{100} = 0.40$

Class Limits	Class Boundaries	Midpoint	Frequency
0.46–0.85	0.455–0.855	0. 655	4
0.86–1.25	0.855–1.255	1.055	5
1.26–1.65	1.255–1.655	1.455	10
1.66–2.05	1.655–2.055	1.855	5
2.06–2.45	2.055–2.455	2.255	5
2.46–2.85	2.455–2.855	2.655	3

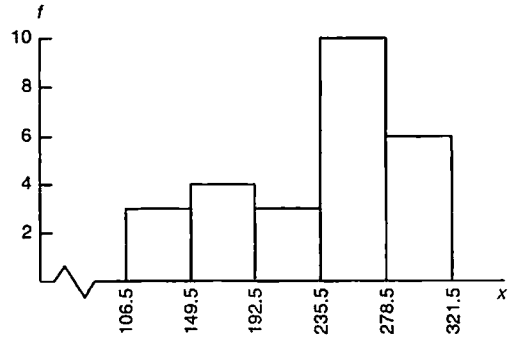
16. (a) $0.194\times1000 = 194, 0.258\times1000 = 258, \ldots, 0.200\times1000 = 200.$

(b) largest value = 317, smallest value = 107

class width = $\frac{317-107}{5} = 42$, use 43

Class Limits	Class Boundaries	Midpoint	Frequency
107–149	106.5–149.5	128	3
150–192	149.5–192.5	171	4
193–235	192.5–235.5	214	3
236–278	235.5–278.5	257	10
279–321	278.5–321.5	300	6

Baseball Batting Averages—Histogram



(c) class width $= \frac{43}{1000} = 0.043$

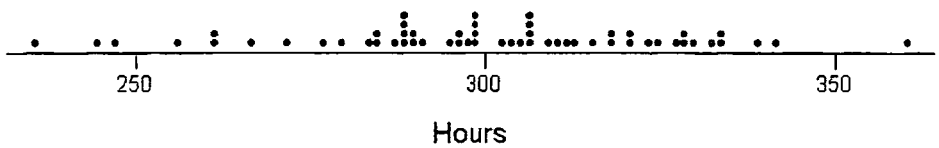
Class Limits	Class Boundaries	Midpoint	Frequency
0.107–0.149	0.1065–0.1495	0.128	3
0.150–0.192	0.1495–0.1925	0.171	4
0.193–0.235	0.1925–0.2355	0.214	3
0.236–0.278	0.2355–0.2785	0.257	10
0.279–0.321	0.2785–0.3215	0.300	6

17. **(a)** There is one dot below 600, so 1 state has 600 or fewer licensed drivers per 1000 residents.

(b) 5 values are close to 800; $\frac{5}{51} \approx 0.0980 \approx 9.8\%$

(c) 9 values below 650
37 values between 650 and 750
5 values above 750
From either the counts or the dotplot, the interval from 650 to 750 licensed drivers per 1000 residents has the most "states."

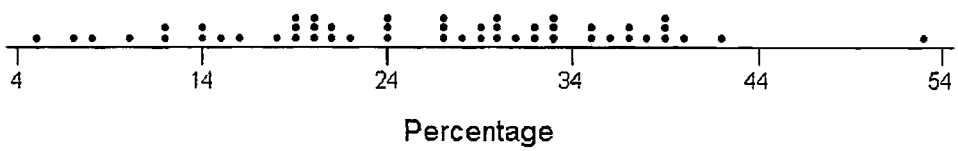
18. The dotplot shows some of the characteristics of the histogram such as more dot density from, say 280 to 340, corresponding roughly to the histogram bars of heights 25 and 16.
However, they are somewhat difficult to compare since the dotplot can be thought of as a histogram with one value, the class mark, i.e., the data value, per class.
Because the definitions of the classes and, therefore, the class widths, differ, it is difficult to compare the two figures.



19. The dotplot shows some of the characteristics of the histogram, such as the concentration of most of the data from, say, 20 to 40; this corresponds roughly to the 3 histogram bars of height 12. There are more data (dots) below 20 than above 40, which corresponds to the histogram bars of heights 4 and 7, and the bars of heights 2 and 1, respectively.

However, they are somewhat difficult to compare since the dotplot can be thought of as a histogram with one value, the class mark, i.e., the data value, per class.

Because the definitions of the classes and, therefore, the class widths, differ, it is difficult to compare the two figures.



Section 2.3

1. (a) The smallest value is 47 and the largest is 97, so we need stems 4, 5, 6, 7, 8, and 9. Use the tens digit as the stem and the ones digit as the leaf.

Longevity of Cowboys

4	7 = 47 years
4	7
5	2 7 8 8
6	1 6 6 8 8
7	0 2 2 3 3 5 6 7
8	4 4 4 5 6 6 7 9
9	0 1 1 2 3 7

(b) Yes, certainly these cowboys lived long lives, as evidenced by the high frequency of leaves for stems 7, 8, and 9 (i.e., 70-, 80-, and 90-year olds).

2. The largest value is 91 (percent of wetlands lost) and the smallest value is 9 (percent), which is coded as 09. We need stems 0 to 9. Use the tens digit as the stem and the ones digit as the leaf. The percentages are concentrated from 20 to 50 percent. The distribution is asymmetrical but not skewed because of the "bump" in the 80s. If we smoothed the shape, we might consider this bimodal. There is a gap showing none of the lower 48 states has lost from 10 to 19% of its wetlands.

Percent of Wetlands Lost

4	0 = 40%
0	9
1	
2	0 3 4 7 7 8
3	0 1 3 5 5 5 6 7 8 8 9
4	2 2 6 6 6 8 9 9
5	0 0 0 2 2 4 6 6 9 9
6	0 7
7	2 3 4
8	1 5 7 7 9
9	0 1

3. The longest average length of stay is 11.1 days in North Dakota and the shortest is 5.2 days in Utah. We need stems from 5 to 11. Use the digit(s) to the left of the decimal point as the stem, and the digit to the right as the leaf.

Average Length of Hospital Stay

5	2 = 5.2 days
5	2 3 5 5 6 7
6	0 2 4 6 6 7 7 8 8 8 8 9 9
7	0 0 0 0 0 0 1 1 1 2 2 2 3 3 3 3 4 4 5 5 6 6 8
8	4 5 7
9	4 6 9
10	0 3
11	1

The distribution is skewed right.

4. Number of Hospitals per State

0	8 = 8 hospitals		
0	8	15	
1	1 2 5 6 9	16	2
2	1 7 7	17	5
3	5 7 8	18	
4	1 2 7	19	3
5	1 2 3 9	20	9
6	1 6 8	21	
7	1	22	7
8	8	23	1 6
9	0 2 6 8		
10	1 2 7	42	1
11	3 3 7 9	43	
12	2 3 9	44	0
13	3 3 6		
14	8		

Texas and California have the highest number of hospitals, 421 and 440, respectively. Both states have large populations and large areas. The four largest states by area are Alaska, Texas, California, and Montana; however, both Alaska and Montana have small populations, but the population tends to cluster at their largest cities, thus reducing the number of hospitals needed.

5. (a) The longest time during 1961–1980 is 23 minutes (i.e., 2:23) and the shortest time is 9 minutes (2:09). We need stems 0, 1, and 2, which we'll write as 0*, 0•, 1*, 1•, 2*, and 2•. (We can eliminate 0* since no time was 2:04 or less and 2• because no winning time was 2:25 or more. We'll use the tens digit as the stem and the ones digit as the leaf, placing leaves 0, 1, 2, 3, and 4 on the "* stem" and leaves 5, 6, 7, 8, and 9 on the "• stem."

Minutes Beyond 2 Hours (1961-1980)

0	9 = 9 minutes past 2 hours
0•	9 9
1*	0 0 2 3 3
1•	5 5 6 6 7 8 8 9
2*	0 2 3 3

(b) The longest time during the period 1981–2000 was 14 (2:14), and the shortest was 7 (2:07), so we'll need stems 0• and 1* only.

Minutes Beyond 2 Hours (1981-2000)

0	7 = 7 minutes past 2 hours
0•	7 7 7 8 8 8 8 9 9 9 9 9 9 9 9
1*	0 0 1 1 4

(c) In more recent times, the winning times have been closer to 2 hours, with all 20 times between 7 and 14 minutes over two hours. In the earlier period, more than half the times (12 or 20) were more than 2 hours and 14 minutes.

6. (a) The largest (worst) score in the first round was 75; the smallest (best) score was 65. We need stems 6• and both 7* and 7• ; leaves 0 to 4 go on the "* stem" and leaves 5–9 belong on the "• stem."

First Round Scores

6	5 = score of 65
6•	5 6 7 7
7*	0 1 1 1 1 1 1 1 1 1 1 2 2 2 3 3 3 3 4 4 4
7•	5 5 5 5 5 5 5

(b) the largest score in the fourth round was 74 and the smallest was 68. Here we need stems 6• and 7*, we don't need 7• because no scores were over 74.

Fourth Round Scores

6	8 = score of 68
6•	8 9 9 9 9 9
7*	0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 4 4 4

(c) Scores are lower in the fourth round. In the first round both the low and high scores were more extreme than in the fourth round.

7. The largest value is 1.808 arc seconds per century; the smallest is 0.008. These values would be coded as 18|08 and 00|08, respectively. We need stems 00 to 18.

Angular Momentum of Stars

00	14 = 0.014 arc sec/century
00	08 14 38 42 50 57
01	73
02	16 19 51
03	51 69
04	30
05	
06	23 67
07	59 88
08	88
09	
10	24 57
11	69 69
12	60 60
13	
14	38
15	
16	16 60
17	
18	08

There are no large gaps, but 4 small gaps. Interestingly, gaps at 05, 09, and 13 are 4 stem units apart, as in the gap from 13 to 17, except that the gap at 15 falls between 13 and 17. This might indicate a cycle. The *midrange** is the average of the largest and smallest values; here, $\frac{0.008+1.808}{2} = 0.908$. If this value had occurred in the data, it would be shown as 09|08. In a sense, this value locates the "middle" of the data. We can see that more of the data (18/28, or approximately 64%) occurs below the midrange than above it (where $10/28 \approx 36\%$ of the data are located).

*The midrange is often calculated in Exploratory Data Analysis (EDA), which is discussed in the next chapter. It is also used in nonparametrics, which is the topic of Chapter 12.

8. The largest value in the data is 67.0×10^{-26} watts per square meter per hertz and the smallest value is 9.0 (i.e., 09.0). We need stems ranging from 0 to 6, and we will use the ones digit and the number after the decimal point to create 2 digit leaves.

Radio Brightness

0	90 = 09.0 units
0	90 90 94 95 95 95 98
1	05 10 15 15 15 25 25 35 36 37 65 65 65
2	00 00 80
3	
4	40 40 40 40
5	
6	70

The measurement 67.0 is unusually bright. Values which are extremely large or extremely small, relative to the rest of the data, are called *outliers*. These are discussed in Chapter 3.

9. The largest value in the data is 29.8 mg. Of tar per cigarette smoked, and the smallest value is 1.0. We will need stems from 1 to 29, and we will use the numbers to the right of the decimal point as the leaves.

Milligrams of Tar per Cigarette

1	0 = 1.0 mg tar
1	0
2	
3	
4	1 5
5	
6	
7	3 8
8	0 6 8
9	0
10	
11	4
12	0 4 8
13	7
14	1 5 9
15	0 1 2 8
16	0 6
17	0
29	8

10. The largest value in the data set is 23.5 mg Carbon monoxide per cigarette smoked. and the smallest is 1.5. We need stems from 1 to 23. and we'll use the numbers to the right of the decimal point as leaves.

Milligrams of Carbon Monoxide

1	5 = 1.5 mg CO
1	5
2	
3	
4	9
5	4
6	
7	
8	5
9	0 5
10	0 2 2 6
11	
12	3 6
13	0 6 9
14	4 9
15	0 4 9
16	3 6
17	5
18	5
23	5

11. The largest value in the data set is 2.03 mg nicotine per cigarette smoked. The smallest value is 0.13. We will need stems 0*, 0•, 1*, 1•, and 2*. Leaves 0 to 4 belong on the * stems and leaves 5 to 9 belong on the • stems. We will use the number to the left of the decimal point as the stem and the first number to the right of the decimal point as the leaf. The number 2 places to the right of the decimal point (the hundredths digit) will be truncated (chopped off; not rounded off).

Milligrams of Nicotine per Cigarette

0	1 = 0.1 milligram
0*	1 4 4
0•	5 6 6 6 7 7 7 8 8 9 9 9
1*	0 0 0 0 0 0 0 1 2
1•	
2*	0

12. (a) For Site I, read the values in Figure 2-27 from the center (stem) to the left to find the least depth is 25 cm and the greatest depth is 110 cm. For Site II, read the values from the center (stem) to the right to find the least depth is 20 cm and the greatest depth is 125 cm.

(b) The Site I depth distribution is, smoothed out, fairly symmetrical around approximately 70 cm. Site II, however, is fairly uniform in shape except that it has a huge gap with no artifacts from about 70 to 100 cm.

(c) It would appear that Site II was probably unoccupied during the time period associated with 70 cm to 100 cm.

13. (a) Average salaries in California range from \$49,000 to \$126,000. Salaries in New York range from \$45,000 to \$120,000.

(b) New York has a greater number of average salaries in the \$60,000 than California, but California has more average salaries than New York in the \$70,000 range.

(c) The California data appear to be similar in shape to the New York data, but California's distribution has been shifted up approximately \$10,000. It is also heavier in the upper tail and shows no gap in average salaries, unlike New York which has no salaries in the \$110,000 range. California has higher average salaries.

Chapter 2 Review

1. Figure 2-1 (a) (in the text) is essentially a bar graph with a "horizontal" axis showing years and a "vertical" axis showing miles per gallon. However, in depicting the data as a highway and showing it in perspective, the ability to correctly compare bar heights visually has been lost. For example, determining what would appear to be the bar heights by measuring from the white line on the road to the edge of the road along a line drawn from the year to its mpg value, we get the bar height for 1983 to be approximately 7/8 inch and the bar height for 1985 to be approximately 1 3/8 inches (i.e., 11/8 inches). Taking the ratio of the given bar heights, we see that the bar for 1985 should be $\frac{27.5}{26} \approx 1.06$ times the length of the 1983 bar. However, the measurements show a ratio of $\frac{\frac{11}{8}}{\frac{7}{8}} = \frac{11}{7} \approx 1.60$, i.e., the 1985 bar is (visually) 1.6 times the length of the 1983 bar. Also, the years are evenly spaced numerically, but the figure shows the more recent years to be more widely spaced due to the use of perspective.

   Figure 2-1(b) is a time plot, showing the years on the x-axis and miles per gallon on the y-axis. Everything is to scale and not distorted visually by the use of perspective. It is easy to see the mpg standards for each year, and you can also see how fuel economy standards for new cars have changed over the eight years shown (i.e., a steep increase in the early years and a leveling off in the later years).

2. (a) By reading the y-coordinate of the dot associated with the year, we estimate the 1980 prison population at approximately 140 prisoners per 100,000, and the 1997 population at approximately 440 prisoners per 100,000 people

(b) The number of inmates per 100,000 increased.

(c) The population 266,574,000 is $2{,}665.74 \times 100{,}000$, and 444 per 100,000 is $\frac{444}{100{,}000}$.

So $\frac{444}{100{,}000} \times (2{,}665.74 \times 100{,}000) \approx 1{,}183{,}589$ prisoners.

The projected 2020 population is 323,724,000, or $3{,}237.24 \times 100{,}000$.

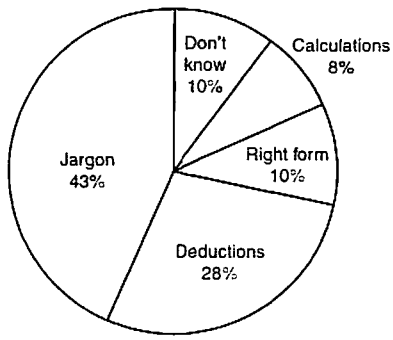
So $\frac{444}{100{,}000} \times (3{,}237.24 \times 100{,}000) \approx 1{,}437{,}335$ prisoners.

3.

Most Difficult Task	Percentage	Degrees
IRS jargon	43%	$0.43 \times 360° \approx 155°$
Deductions	28%	$0.28 \times 360° \approx 101°$
Right form	10%	$0.10 \times 360° = 36°$
Calculations	8%	$0.08 \times 360° \approx 29°$
Don't know	10%	$0.10 \times 360° = 36°$

Note: Degrees do not total 360° due to rounding.

Problems with Tax Returns



4. (a) Since the ages are two digit numbers, use the tens digit as the stem and the ones digit as the leaf.

Age of DUI Arrests

1	6 = 16 years
1	6 8
2	0 1 1 2 2 2 3 4 4 5 6 6 6 7 7 7 9
3	0 0 1 1 2 3 4 4 5 5 6 7 8 9
4	0 0 1 3 5 6 7 7 9 9
5	1 3 5 6 8
6	3 4

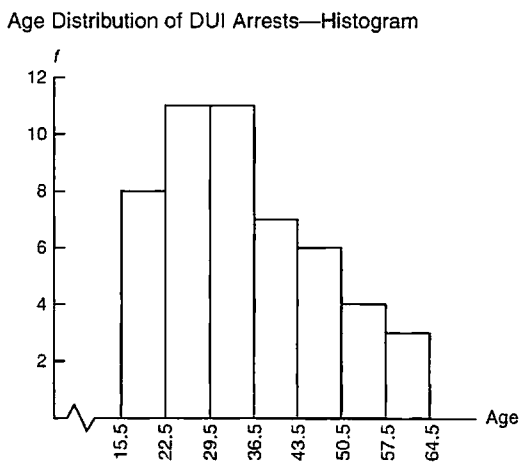
(b) The largest age is 64 and the smallest is 16. so the class with for 7 classes is $\frac{64-16}{7} \approx 6.86$; use 7. The lower class limit for the first class is 16; the lower class limit for the second class is 16 + 7 = 23. The total number of data points is 50, so calculate the relative frequency by dividing the class frequency by 50.

Age Distribution of DUI Arrests

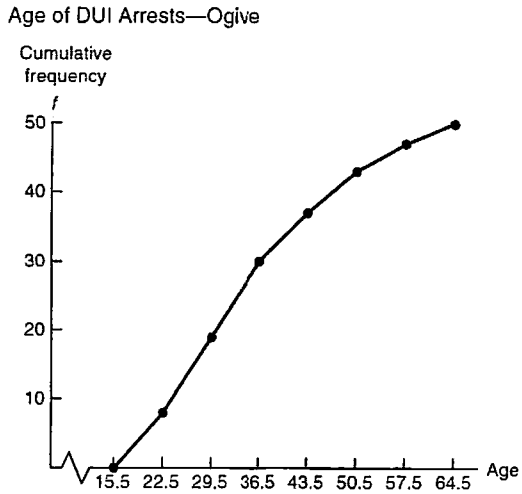
Class Limits	Class Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
16–22	15.5–22.5	19	8	0.16	8
23–29	22.5–29.5	26	11	0.22	19
30–36	29.5–36.5	33	11	0.22	30
37–43	36.5–43.5	40	7	0.14	37
44–50	43.5–50.5	47	6	0.12	43
51–57	50.5–57.5	54	4	0.08	47
58–64	57.5–64.5	61	3	0.06	50

The class boundaries are the average of the upper class limit of the next class. The midpoint is the average of the class limits for that class.

(c) The class boundaries are shown in (b).



(d) The ogive plots the cumulative frequency up to the upper class boundary value.



By reading the y-axis value for the dot over the upper boundary 29.5, we see that 19 of 50, or $\frac{19}{50} = 38\%$ of the drivers were 29 years or younger when arrested.

5. (a) The largest value is 96 mg of glucose per 100 ml of blood, and the smallest value is 59. For 7 classes we need a class width of $\frac{96-59}{7} \approx 5.3$; use 6. The lower class limit of the first class is 59, and the lower class limit of the second class is 59 + 6 = 65.

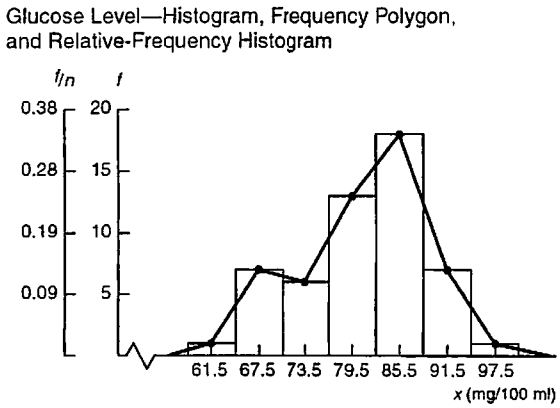
The class boundaries are the average of the upper class limit of one class and the lower class limit of the next higher class. The midpoint is the average of the class limits for that class. There are 53 data values total so the relative frequency is the class frequency divided by 53.

Class Limits	Class Boundaries	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
59–64	58.5–64.5	61.5	1	0.02	1
65–70	64.5–70.5	67.5	7	0.13	8
71–76	70.5–76.5	73.5	6	0.11	14
77–82	76.5–82.5	79.5	13	0.25	27
83–88	82.5–88.5	85.5	18	0.34	45
89–94	88.5–94.5	91.5	7	0.13	52
95–100	94.5–100.5	97.5	1	0.02	53

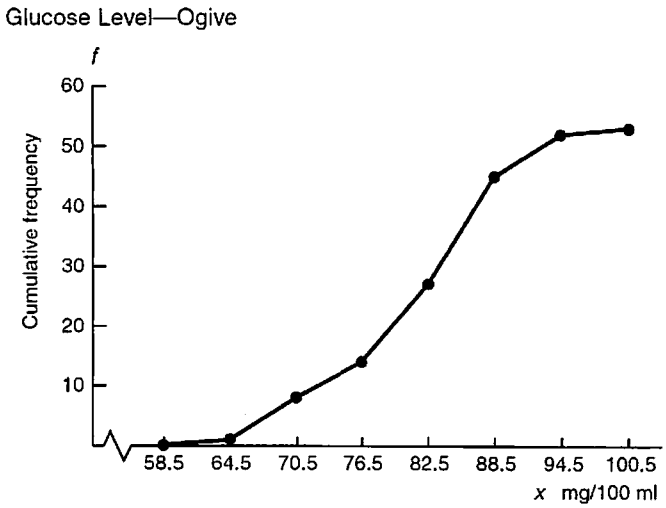
(b) The histogram shows the bars centered over the midpoints of each class.

(c) The frequency polygon begins on the x-axis at the point one class width below the first class midpoint: $61.5 - 6 = 55.5$. It connects this point and the other midpoints with line segments. It ends on the x-axis one class width above the last class midpoint: $97.5 + 6 = 103.5$.

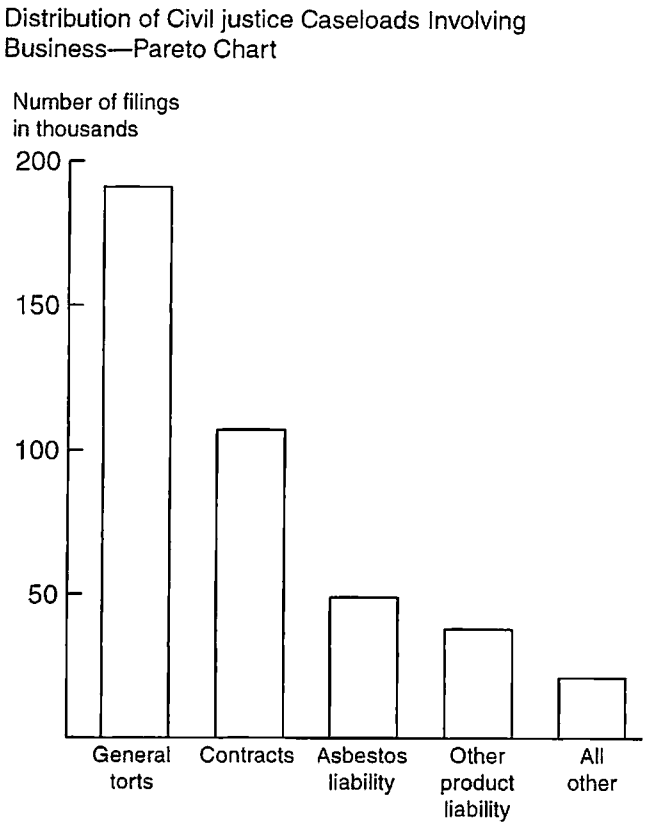
(d) The frequency histogram and the relative frequency histogram are the same except in the latter, the vertical scale is relative frequency, not frequency.



(e) The ogive begins on the x-axis at the lower class boundary and connects dots placed at (x, y) coordinates (upper class boundary, cumulative frequency).



6. (a) A pareto chart is similar to a bar chart, except the bars are in decreasing order by frequency.



The general torts (personal injury) lawsuits occur with the greatest frequency.

(b) The total number of filings shown is 406 (thousand).

Case Type	Percentage	Degrees
Contracts	$107/406 \approx 26\%$	$0.26 \times 360° \approx 94°$
General torts	$191/406 \approx 47\%$	$0.47 \times 360° \approx 169°$
Asbestos liability	$49/406 \approx 12\%$	$0.12 \times 360° \approx 43°$
Other product liability	$38/406 \approx 9\%$	$0.09 \times 360° \approx 32°$
All other	$21/406 \approx 5\%$	$0.05 \times 360° = 18°$

Note: Percentages do not add to 100% due to rounding. Similarly, the degrees do not add to 360° due to rounding.

Distribution of Civil justice Caseloads Involving Business—Pie Chart

7. **(a)** To determine the decade which contained the most samples, count both rows (if shown) of leaves; recall leaves 0–4 belong on the first line and 5–9 belong on the second line when two lines per stem are used. The greatest number of leaves is found on stem 124, i.e., the 1240s (the 40s decade in the 1200s), with 40 samples.

(b) The number of samples with tree ring dates 1200 A.D. to 1239 A.D. is 28 + 3 + 19 + 25 = 75.

(c) The dates of the longest interval with no sample values are 1204 through 1211 A.D. This might mean that for these eight years, the pueblo was unoccupied (thus no new or repaired structures) or that the population remained stable (no new structures needed) or that, say, weather conditions were favorable these years so existing structures didn't need repair. If relatively few new structures were built or repaired during this period, their tree rings might have been missed during sample selection.

8. **(a)** It has a long tail on the left, so it is skewed left.

(b) The class width is the difference between any two adjacent midpoints. Here, for example, the class width is 4 – 3.5 = 0.5 grade points. The average of any two adjacent midpoints is the boundary value between the two midpoints classes*. So, for midpoints 1 and 1.5, the boundary value is $1+\frac{1.5}{26}=1.25$.

The difference between any two adjacent boundary values is also the class width, so the other class boundary values within the histogram are 1.25 + 0.5 = 1.75, 1.75 + 0.5 = 2.25, 2.25 + 0.5 = 2.75, 2.75 + 0.5 = 3.25, and 3.25 + 0.5 = 3.75; 3.75 is the lower class boundary for the class, so its upper class boundary is 3.75 + 0.5 = 4.25. Similarly, the upper class boundary of the first class was 1.25, so its lower class boundary is 1.25 – 0.5 = 0.75. The class boundaries are, therefore, 0.75, 1.25, 1.75, 2.25, 2.75, 3.25, 3.75 and 4.25 (from left to right).

*Recall that the average of a and b is $\frac{a+b}{2}$ which is also the value halfway between a and b.

(c) The relative frequencies are f/n, so if we multiply this decimal value by 100, we have the relative frequency expressed as a percent. The relative frequencies, expressed as percents, are 1%, 1%, 2%, 8%, 17%, 27%, and 44%, from left to right. The GPA of 3.25 is a boundary value, so to find the percentage of college graduates who had high school GPAs less than 3.25 is the sum of the relative frequency percentages for bars at or below 3.25: 1% + 1% + 2% + 8% + 17% = 29%. A high school GPA of 3.75 is the next boundary value above 3.25, so if we take the percentage of students with GPAs less 3.25 (29%), and add the percentage of students with GPAs between 3.25 and 3.75 (27%), we find 29% + 27% = 56% of college graduates had high school GPAs of less than 3.75. (Recall that, technically, boundary values are not values the data can take on. They are values between the upper class limit of one class and the lower class limit of the next class, and the class limits specify the largest and smallest data values, respectively, that can be put in those classes. Traditionally, the boundary values are specified to one more decimal place than the data, and that is the case here: the data are reported to one decimal place, but the boundaries are reported to two decimal places.)

Class Midpoints	Class Boundaries	Relative Frequency	Relative Frequency	Cumulative Relative Frequency (%)
1	0.75–1.25	0.01	1%	1%
1.5	1.25–1.75	0.01	1%	2%
2	1.75–2.25	0.02	2%	4%
2.5	2.25–2.75	0.08	8%	12%
3	2.75–3.25	0.17	17%	29%
3.5	3.25–3.75	0.27	27%	56%
4	3.75–4.25	0.44	44%	100%

9. (a) The age group that is most frequently in the hospital has the highest frequency and, therefore, the highest relative frequency: the age group with boundaries 64.5 and 84.5, enclosing ages 65 to 84.

(b)

Class Limits	Class Boundaries	Relative Frequency	Relative Frequency	Cumulative Relative Frequency (%)
5–24	4.5–24.5	0.16	16%	16%
25–44	24.5–44.5	0.28	28%	44%
45–64	44.5–64.5	0.21	21%	65%
65–84	64.5–84.5	0.35	35%	100%

The percentage of patients older than 44, i.e., from 45 to 84, is 21% + 35% = 56%.

(c) The percentage of patients 44 or younger is 16% + 28% = 44%.

10.(a) The largest value is 93 years of age, and the smallest value is 34 years of age (probably Bill Gates of Microsoft). We will need stems from 3 to 9. Use the tens digit as the stem and the ones digit as the leaf.

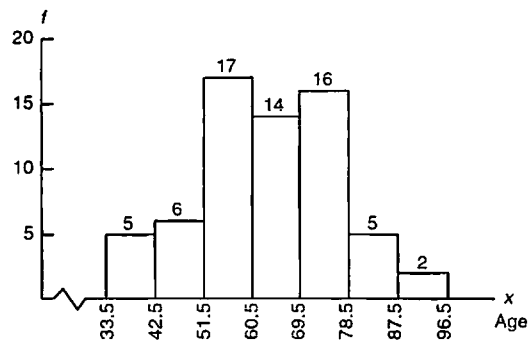
Ages of Wealthy

3	4 = 34 years old
3	4
4	0 0 0 1 3 7 8 8 8
5	0 2 2 2 2 3 3 3 3 4 6 6 7 7 8 9
6	0 0 1 3 4 5 5 6 6 6 6 6 7 7 8 8
7	0 0 0 1 1 2 3 3 3 4 5 6 6 7 7 7 9
8	2 2 3 3 8
9	3

(b) The class width for 7 class is $\frac{93-34}{7} \approx 8.4$: use 9. The first class' lower limit is 34 and the second class' lower limit is 34 + 9 = 43. The boundary value between them is $\frac{34+43}{2} = 38.5$.

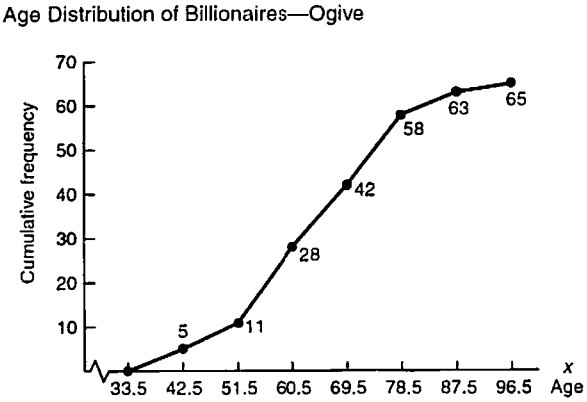
Class Limits	Class Boundaries	Frequency	Cumulative Frequency
34-42	33.5-42.5	5	5
43-51	42.5-51.5	6	11
52-60	51.5-60.5	17	28
61-69	60.5-69.5	14	42
70-78	69.5-78.5	16	58
79-87	78.5-87.5	5	63
88-96	87.5-96.5	2	65

Age Distribution of Billionaires—Histogram



Smoothed, the histogram would be fairly symmetrical.

(c) The ogive connects dots placed over the upper boundary values at the height of the cumulative frequency at those values. It begins with a dot on the x-axis at the lower class boundary of the first class.



The number of multi billionaires 51 years old or younger in the cumulative frequency at boundary value 51.5 (which is 11). The percentage of such persons is $11/65 \approx 17\%$ (where 65 is the total number of ages given in the data).