Contact me in order to access the whole complete document. Email: solution9159@gmail.com
WhatsApp: https://wa.me/message/2H3BV2L5TTSUF1 Telegram: https://t.me/solutionmanual

Contents

1: Describing Data with Graphs

Section 1.1

1.1.1 The experimental unit, the individual or object on which a variable is measured, is the student.

1.1.2 The experimental unit on which the number of errors is measured is the exam.

1.1.3 The experimental unit is the patient.

1.1.4 The experimental unit is the azalea plant.

1.1.5 The experimental unit is the car.

1.1.6 "Time to assemble" is a *quantitative* variable because a numerical quantity (1 hour, 1.5 hours, etc.) is measured.

1.1.7 "Number of students" is a *quantitative* variable because a numerical quantity (1, 2, etc.) is measured.

1.1.8 "Rating of a politician" is a *qualitative* variable since a quality (excellent, good, fair, poor) is measured.

1.1.9 "State of residence" is a *qualitative* variable since a quality (CA, MT, AL, etc.) is measured.

1.1.10 "Population" is a *discrete* variable because it can take on only integer values.

1.1.11 "Weight" is a *continuous* variable, taking on any values associated with an interval on the real line.

1.1.12 Number of claims is a *discrete* variable because it can take on only integer values.

1.1.13 "Number of consumers" is integer-valued and hence *discrete*.

1.1.14 "Number of boating accidents" is integer-valued and hence *discrete*.

1.1.15 "Time" is a *continuous* variable.

1.1.16 "Cost of a head of lettuce" is a *discrete* variable since money can be measured only in dollars and cents.

1.1.17 "Number of brothers and sisters" is integer-valued and hence *discrete*.

1.1.18 "Yield in bushels" is a *continuous* variable, taking on any values associated with an interval on the real line.

1.1.19 The statewide database contains a record of all drivers in the state of Michigan. The data collected represents the *population* of interest to the researcher.

1.1.20 The researcher is interested in the opinions of all citizens, not just the 1000 citizens that have been interviewed. The responses of these 1000 citizens represent a *sample*.

1.1.21 The researcher is interested in the weight gain of all animals that might be put on this diet, not just the twenty animals that have been observed. The responses of these twenty animals is a *sample*.

1.1.22 The data from the Internal Revenue Service contains the records of all wage earners in the United States. The data collected represents the *population* of interest to the researcher.

1.1.23 **a** The experimental unit, the item or object on which variables are measured, is the vehicle.

b Type (qualitative); make (qualitative); carpool or not? (qualitative); one-way commute distance (quantitative continuous); age of vehicle (quantitative continuous)

c Since five variables have been measured, this is *multivariate data.*

1.1.24 **a** The set of ages at death represents a population, because there have only been 38 different presidents in the United States history.

b The variable being measured is the continuous variable "age".

c "Age" is a quantitative variable.

1.1.25 **a** The population of interest consists of voter opinions (for or against the candidate) at the time of the election for all persons voting in the election.

b Note that when a sample is taken (at some time prior or the election), we are not actually sampling from the population of interest. As time passes, voter opinions change. Hence, the population of voter opinions changes with time, and the sample may not be representative of the population of interest.

1.1.26 **a-b** The variable "survival time" is a quantitative continuous variable.

c The population of interest is the population of survival times for all patients having a particular type of cancer and having undergone a particular type of radiotherapy.

d-e Note that there is a problem with sampling in this situation. If we sample from all patients having cancer and radiotherapy, some may still be living and their survival time will not be measurable. Hence, we cannot sample directly from the population of interest, but must arrive at some reasonable alternate population from which to sample.

1.1.27 **a** The variable "reading score" is a quantitative variable, which is probably integer-valued and hence discrete.

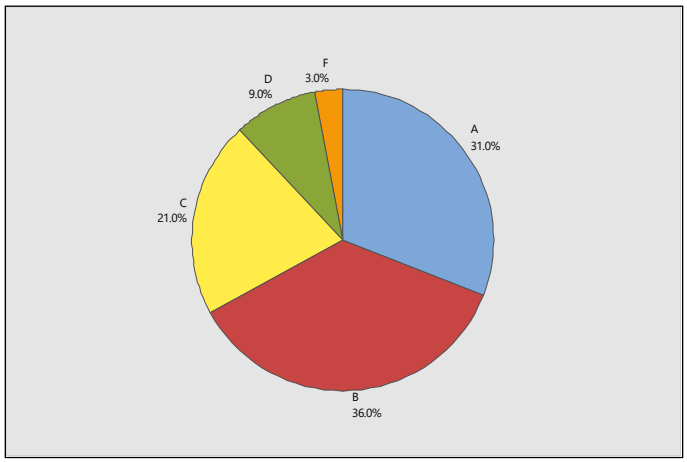
b The individual on which the variable is measured is the student.

c The population is hypothetical – it does not exist in fact – but consists of the reading scores for all students who could possibly be taught by this method.

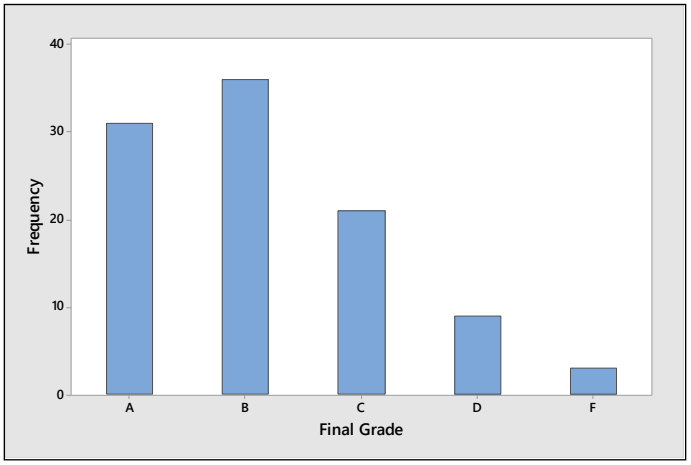
Section 1.2

1.2.1 The pie chart is constructed by partitioning the circle into five parts, according to the total contributed by each part. Since the total number of students is 100, the total number receiving a final grade of A represents $31/100 = 0.31$ or 31% of the total. Thus, this category will be represented by a sector angle of $0.31(360) = 111.6°$. The other sector angles are shown next, along with the pie chart.

Final Grade	Frequency	Fraction of Total	Sector Angle
A	31	.31	111.6
B	36	.36	129.6
C	21	.21	75.6
D	9	.09	32.4
F	3	.03	10.8

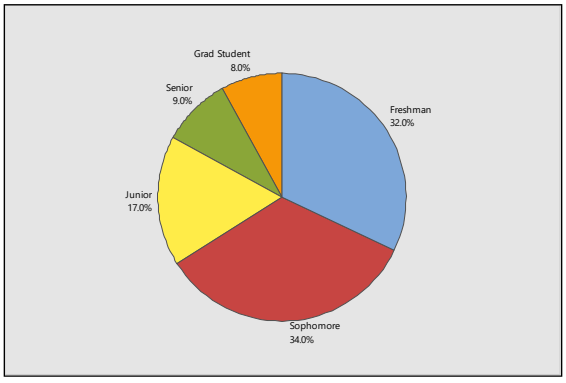


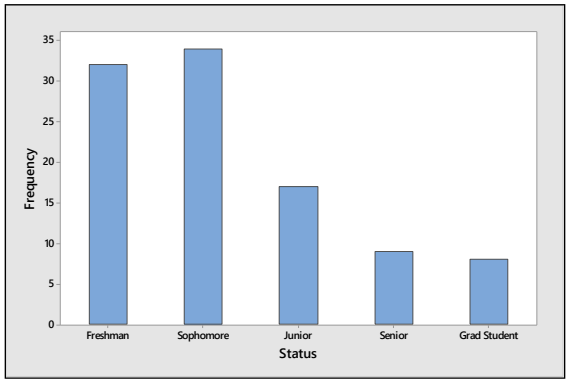
The bar chart represents each category as a bar with height equal to the frequency of occurrence of that category and is shown in the figure that follows.



1.2.2 Construct a statistical table to summarize the data. The pie and bar charts are shown in the figures that follow.

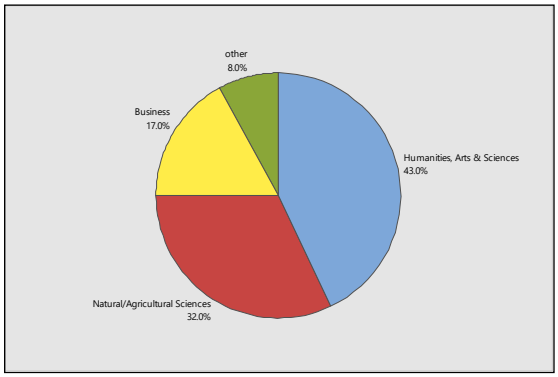
Status	Frequency	Fraction of Total	Sector Angle
Freshman	32	.32	115.2
Sophomore	34	.34	122.4
Junior	17	.17	61.2
Senior	9	.09	32.4
Grad Student	8	.08	28.8

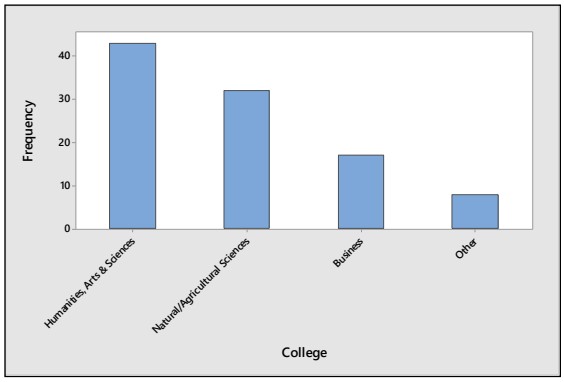




1.2.3 Construct a statistical table to summarize the data. The pie and bar charts are shown in the figures that follow.

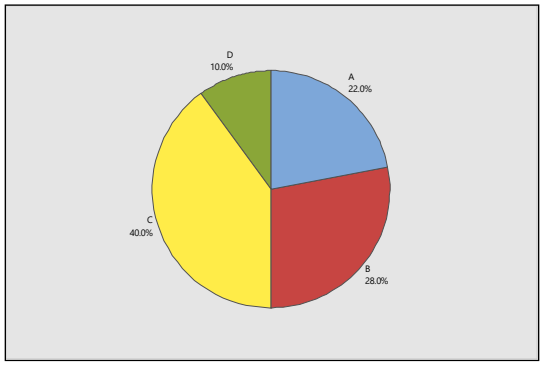
Status	Frequency	Fraction of Total	Sector Angle
Humanities, Arts & Sciences	43	.43	154.8
Natural/Agricultural Sciences	32	.32	115.2
Business	17	.17	61.2
Other	8	.08	28.8

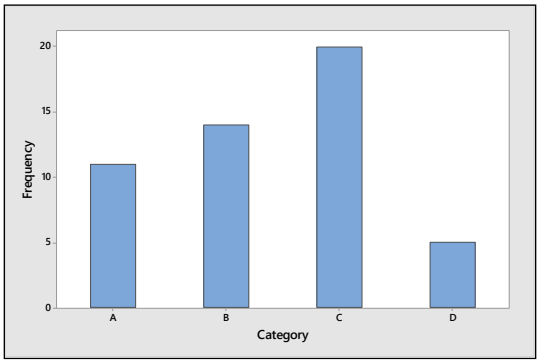




1.2.4 **a** The pie chart is constructed by partitioning the circle into four parts, according to the total contributed by each part. Since the total number of people is 50, the total number in category A represents $11/50 = 0.22$ or 22% of the total. Thus, this category will be represented by a sector angle of $0.22(360) = 79.2^{\circ}$. The other sector angles are shown below. The pie chart is shown in the figure that follows.

Category	Frequency	Fraction of Total	Sector Angle
A	11	.22	79.2
B	14	.28	100.8
C	20	.40	144.0
D	5	.10	36.0





b The bar chart represents each category as a bar with height equal to the frequency of occurrence of that category and is shown in the figure above.

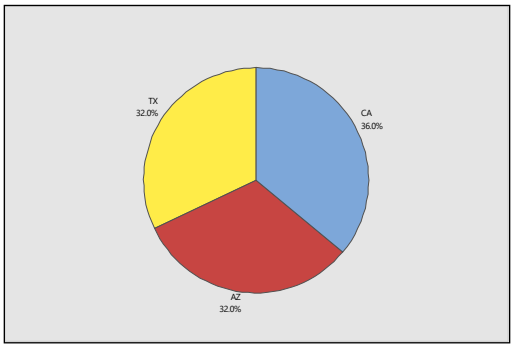
c Yes, the shape will change depending on the order of presentation. The order is unimportant.

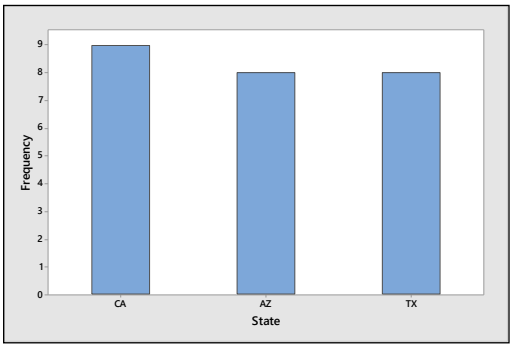
d The proportion of people in categories B, C, or D is found by summing the frequencies in those three categories, and dividing by n = 50. That is, $(14+20+5)/50 = 0.78$.

e Since there are 14 people in category B, there are $50-14=36$ who are not, and the percentage is calculated as $(36/50)100 = 72\%$.

1.2.5 **a-b** Construct a statistical table to summarize the data. The pie and bar charts are shown in the figures that follow.

State	Frequency	Fraction of Total	Sector Angle
CA	9	.36	129.6
AZ	8	.32	115.2
TX	8	.32	115.2



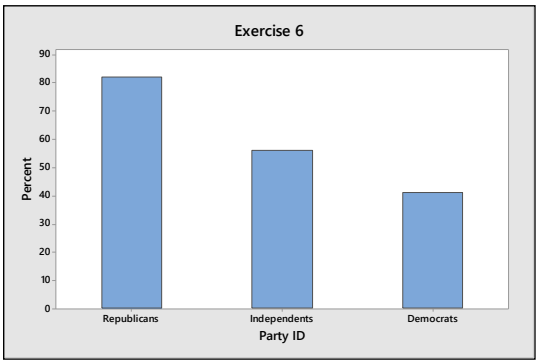


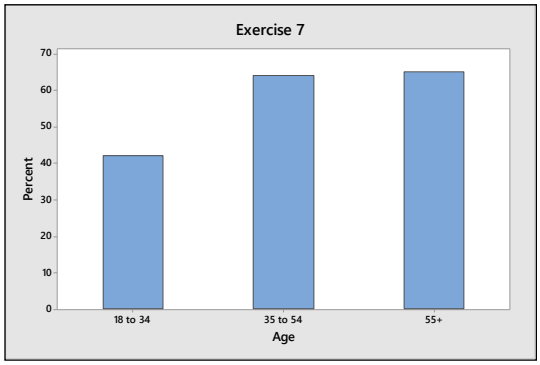
c From the table or the chart, Texas produced $8/25 = 0.32$ of the jeans.

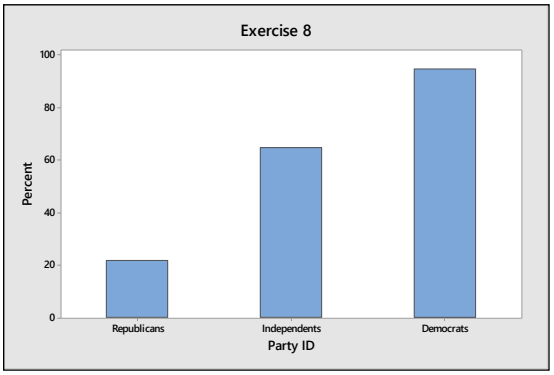
d The highest bar represents California, which produced the most pairs of jeans.

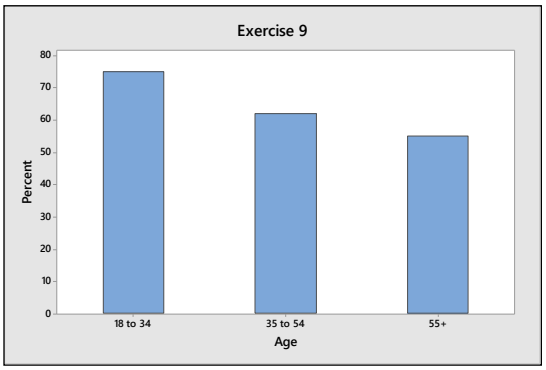
e Since the bars and the sectors are almost equal in size, the three states produced roughly the same number of pairs of jeans.

1.2.6-9 The bar charts represent each category as a bar with height equal to the frequency of occurrence of that category.





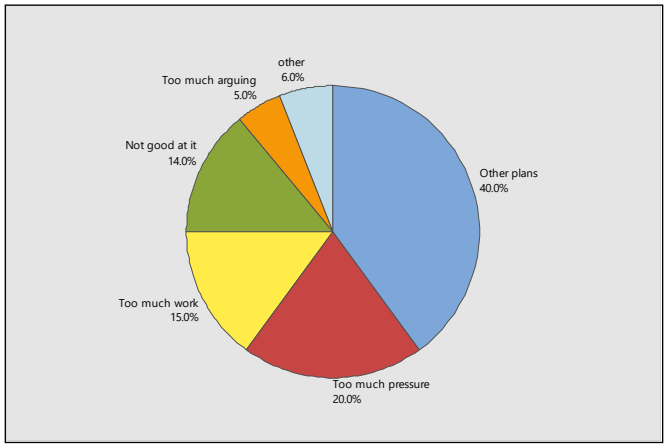




1.2.10 Answers will vary.

1.2.11 **a** The percentages given in the exercise only add to 94%. We should add another category called "Other", which will account for the other 6% of the responses.

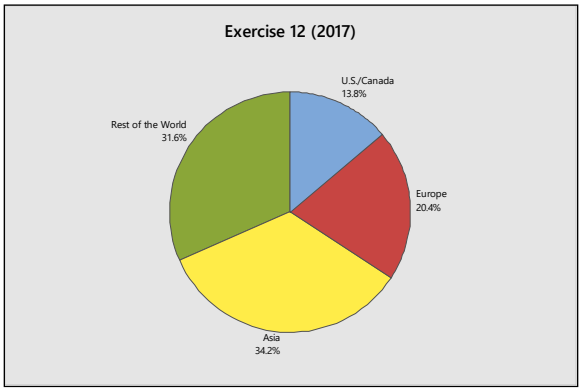
b Either type of chart is appropriate. Since the data is already presented as percentages of the whole group, we choose to use a pie chart, shown in the figure that follows.

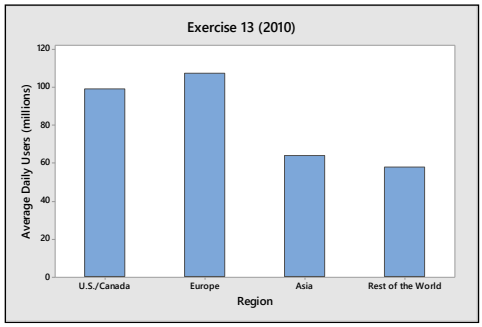


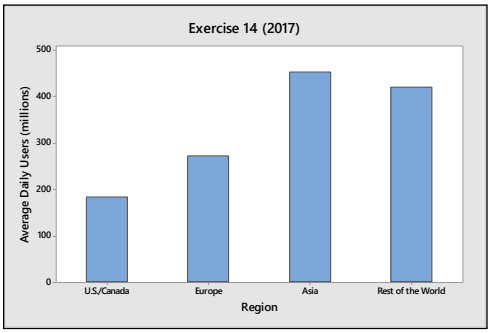
c-d Answers will vary.

1.2.12-14 The percentages falling in each of the four categories in 2017 are shown next (in parentheses), and the pie chart for 2017 and bar charts for 2010 and 2017 follow.

Region	**2010**	**2017**
United States/Canada	99	183 (13.8%)
Europe	107	271 (20.4%)
Asia	64	453 (34.2%)
Rest of the World	58	419 (31.6%)
Total	**328**	**1326 (100%)**





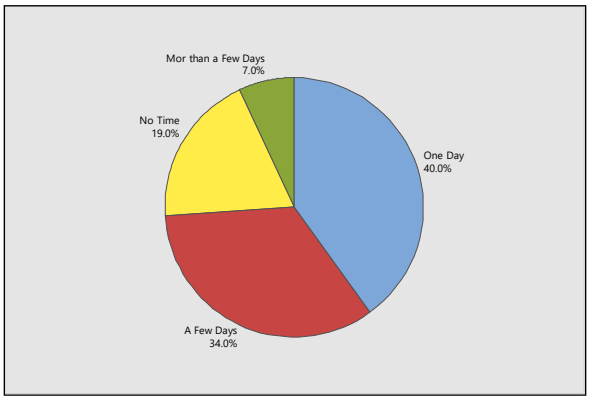


1.2.15 Users in Asia and the rest of the world have increased more rapidly than those in the U.S., Canada or Europe over the seven-year period.

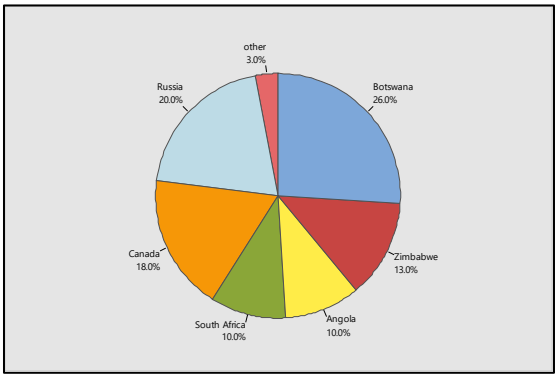
1.2.16 **a** The total percentage of responses given in the table is only $(40+34+19)\% = 93\%$. Hence there are 7% of the opinions not recorded, which should go into a category called "Other" or "More than a few days".

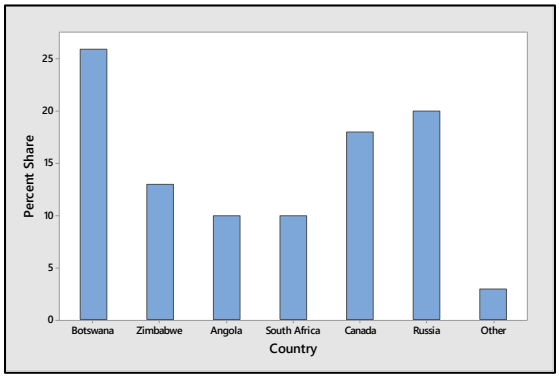
b Yes. The bars are very close to the correct proportions.

c Similar to previous exercises. The pie chart is shown next. The bar chart is probably more interesting to look at.

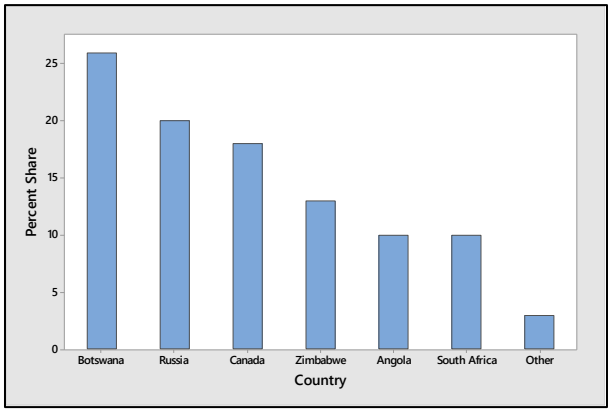


1.2.17-18 Answers will vary from student to student. Since the graph gives a range of values for Zimbabwe's share, we have chosen to use the 13% figure, and have used 3% in the "Other" category. The pie chart and bar charts are shown next.

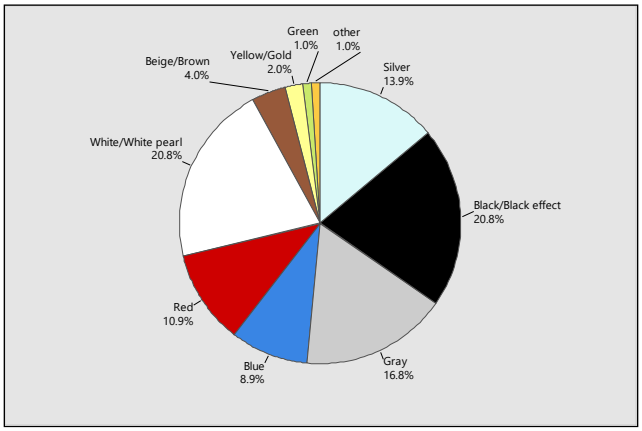




1.2.19-20 The Pareto chart is shown below. The Pareto chart is more effective than the bar chart or the pie chart.

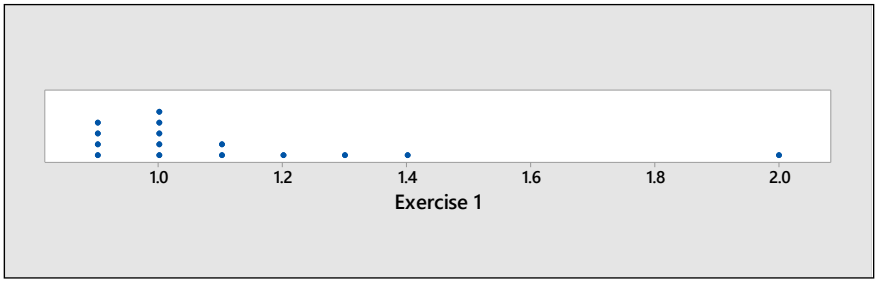


1.2.21 The data should be displayed with either a bar chart or a pie chart. The pie chart is shown next.

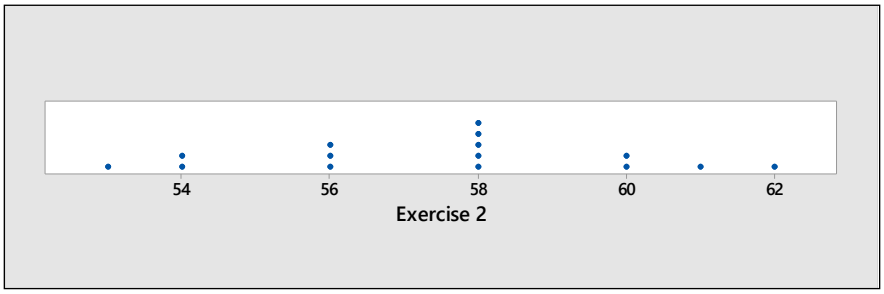


Section 1.3

1.3.1 The dotplot is shown next; the data is skewed right, with one outlier, $x = 2.0$.



1.3.2 The dotplot is shown next; the data is relatively mound-shaped, with no outliers.



1.3.3-5 The most obvious choice of a stem is to use the ones digit. The portion of the observation to the right of the ones digit constitutes the leaf. Observations are classified by row according to stem and also within each stem according to relative magnitude. The stem and leaf display is shown next.

1 6 8

2 1 2 5 5 5 7 8 8 9 9

3 1 1 4 5 5 6 6 6 7 7 7 7 8 9 9 9 leaf digit = 0.1

4 0 0 0 1 2 2 3 4 5 6 7 8 9 9 9 1 2 represents 1.2

5 1 1 6 6 7

6 1 2

3. The stem and leaf display has a mound shaped distribution, with no outliers.

4. From the stem and leaf display, the smallest observation is 1.6 (1 6).

5. The eight and ninth largest observations are both 4.9 (4 9).

1.3.6 The stem is chosen as the ones digit, and the portion of the observation to the right of the ones digit is the leaf.

3 | 2 3 4 5 5 5 6 6 7 9 9 9 9

4 | 0 0 2 2 3 3 3 4 4 5 8 leaf digit = 0.1 1 2 represents 1.2

1.3.7-8 The stems are split, with the leaf digits 0 to 4 belonging to the first part of the stem and the leaf digits 5 to 9 belonging to the second. The stem and leaf display shown below improves the presentation of the data.

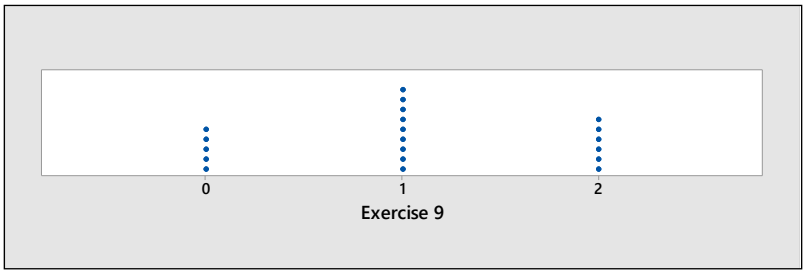
3 | 2 3 4

3 | 5 5 5 6 6 7 9 9 9 9 leaf digit = 0.1 1 2 represents 1.2

4 | 0 0 2 2 3 3 3 4 4

4 | 5 8

1.3.9 The scale is drawn on the horizontal axis and the measurements are represented by dots.



1.3.10 Since there is only one digit in each measurement, the ones digit must be the stem, and the leaf will be a zero digit for each measurement.

0 | 0 0 0 0 0

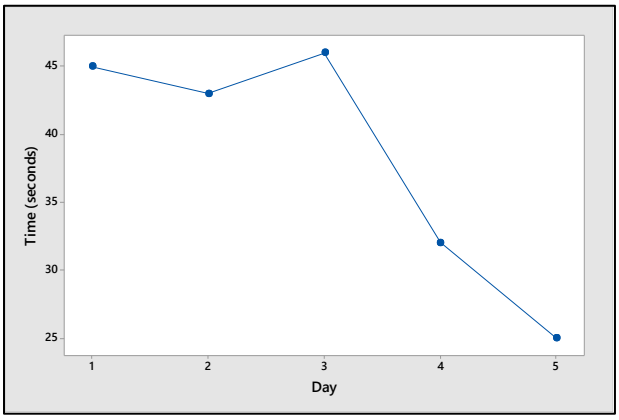
1 | 0 0 0 0 0 0 0 0 0

2 | 0 0 0 0 0 0

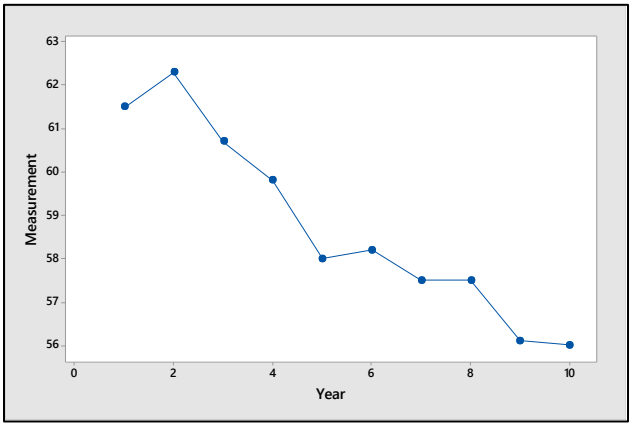
1.3.11 The distribution is relatively mound-shaped, with no outliers.

1.3.12 The two plots convey the same information if the stem and leaf plot is turned 90° and stretched to resemble the dotplot.

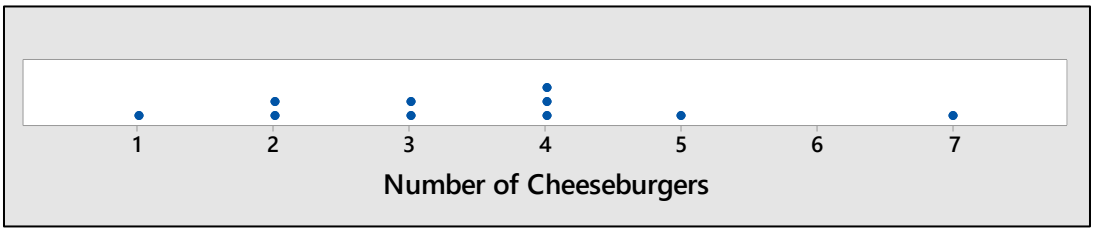
1.3.13 The line chart plots "day" on the horizontal axis and "time" on the vertical axis. The line chart shown next reveals that learning is taking place, since the time decreases each successive day.



1.3.14 The line graph is shown next. Notice the change in y as x increases. The measurements are decreasing over time.



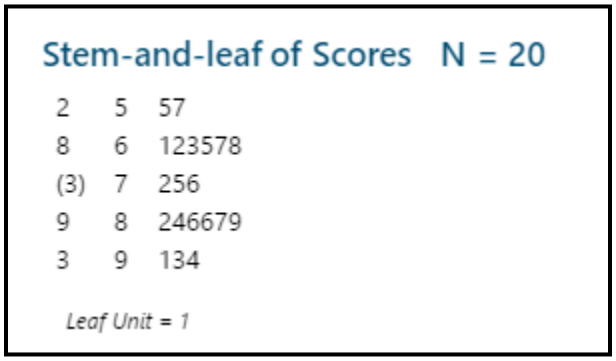
1.3.15 The dotplot is shown next.



a The distribution is somewhat mound-shaped (as much as a small set can be); there are no outliers.

b $2/10 = 0.2$

1.3.16 **a** The test scores are graphed using a stem and leaf plot generated by ***Minitab***.



b-c The distribution is not mound-shaped, but is rather has two peaks centered around the scores 65 and 85. This might indicate that the students are divided into two groups – those who understand the material and do well on exams, and those who do not have a thorough command of the material.

1.3.17 **a** We choose a stem and leaf plot, using the ones and tenths place as the stem, and a zero digit as the leaf. The ***Minitab*** printout is shown next.

Stem-and-leaf of RBC Count N = 15

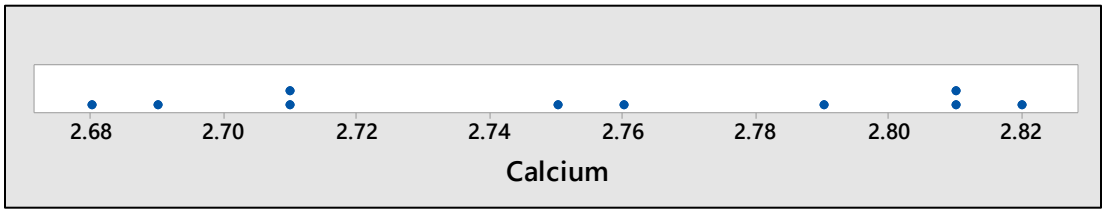
1	49	0
2	50	0
3	51	0
(5)	52	00000
7	53	000
4	54	000
1	55	0

Leaf Unit = 0.01

b The data set is relatively mound-shaped, centered at 5.2.

c The value $x = 5.7$ does not fall within the range of the other cell counts, and would be considered somewhat unusual.

1.3.18 **a-b** The dotplot and the stem and leaf plot are drawn using ***Minitab***.



Stem-and-leaf of Calcium N = 10

2	26	89
(5)	27	11569
3	28	112

Leaf Unit = 0.01

c The measurements all seem to be within the same range of variability. There do not appear to be any outliers.

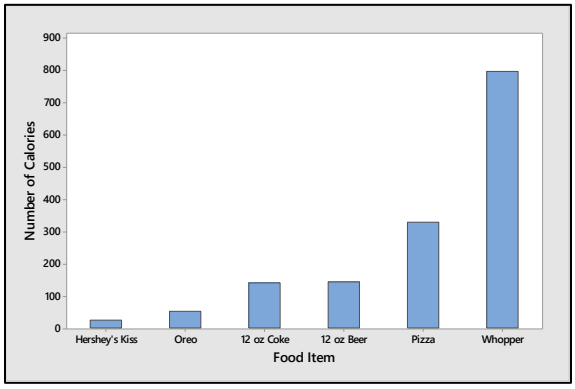
1.3.19 **a** Stem and leaf displays may vary from student to student. The most obvious choice is to use the tens digit as the stem and the ones digit as the leaf.

```
 7 | 8 9
 8 | 0 1 7
 9 | 0 1 2 4 4 5 6 6 6 8 8
10 | 1 7 9
11 | 2
```

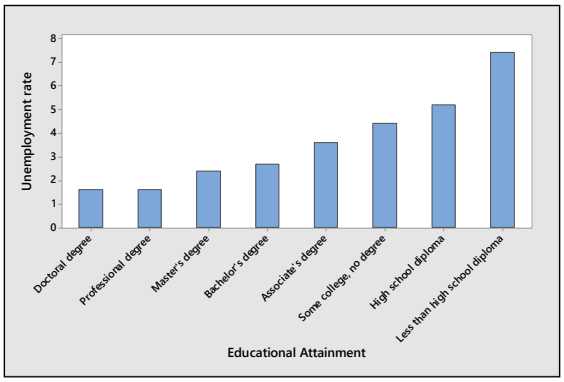
b The display is fairly mound-shaped, with a large peak in the middle.

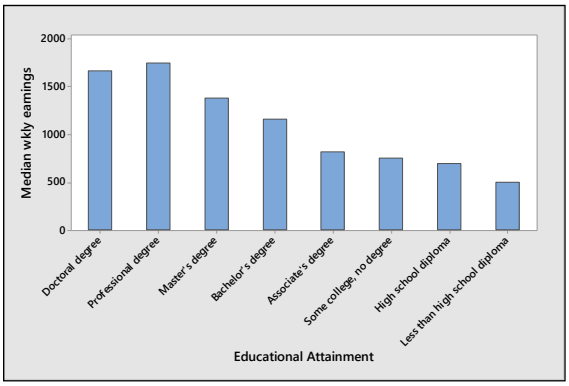
1.3.20 **a** The sizes and volumes of the food items do increase as the number of calories increase, but not in the correct proportion to the actual calories. The differences in calorie content are not accurately portrayed in the graph.

b The bar graph which accurately portrays the number of calories in the six food items is shown next.



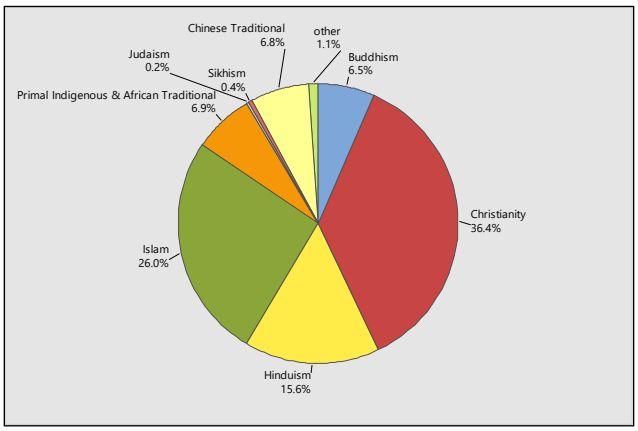
1.3.21 **a-b** The bar charts for the median weekly earnings and unemployment rates for eight different levels of education are shown next.



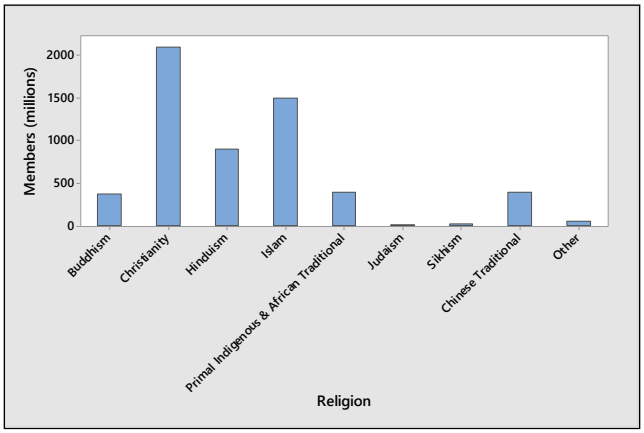


c The unemployment rate drops and the median weekly earnings rise as the level of educational attainment increases.

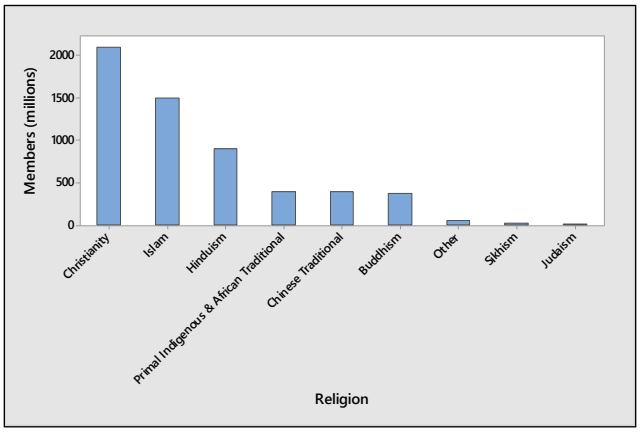
1.3.22 **a** Similar to previous exercises. The pie chart is shown next.



b The bar chart is shown next.



c The Pareto chart is a bar chart with the heights of the bars ordered from large to small. This display is more effective than the pie chart.



1.3.23 **a** The distribution is skewed to the right, with a several unusually large measurements. The five states marked as HI are California, New Jersey, New York and Pennsylvania.

b Three of the four states are quite large in area, which might explain the large number of hazardous waste sites. However, New Jersey is relatively small, and other large states do not have unusually large number of waste sites. The pattern is not clear.

1.3.24 **a**

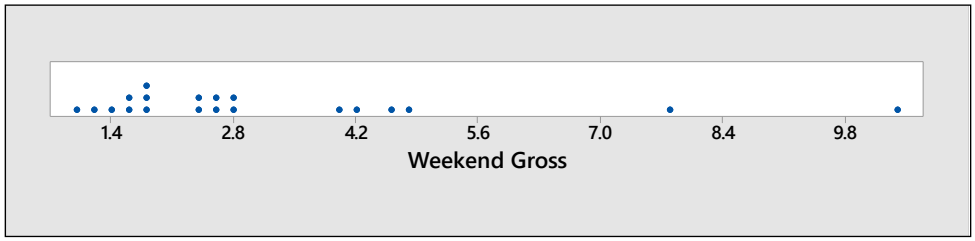
Stem-and-leaf of Weekend Gross N = 20

1	0	9
8	1	2455778
(6)	2	345678
6	3	9
5	4	267
2	5	
2	6	
2	7	7
1	8	
1	9	
1	10	3

Leaf Unit = 0.1

The distribution is skewed to the right, with two outliers.

b The dotplot is shown next. It conveys nearly the same information, but the stem-and-leaf plot may be more informative.



1.3.25 **a** Answers will vary.

b The stem and leaf plot is constructed using the tens place as the stem and the ones place as the leaf. Notice that the distribution is roughly mound-shaped.

Stem-and-leaf of Ages N = 38

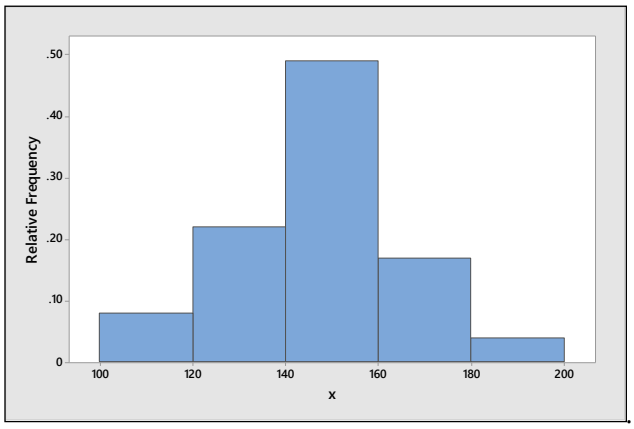
2	4	69
7	5	36778
19	6	003344567778
19	7	0112347889
9	8	01358
4	9	0033

Leaf Unit = 1

c-d Three of the five youngest presidents – Kennedy, Lincoln and Garfield – were assassinated while in office. This would explain the fact that their ages at death were in the lower tail of the distribution.

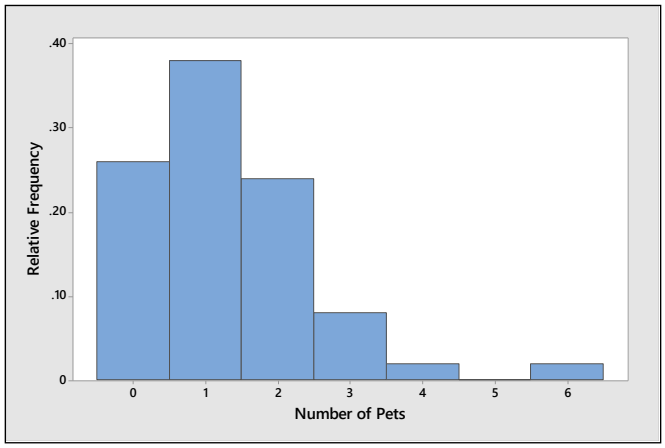
Section 1.4

1.4.1 The relative frequency histogram displays the relative frequency as the height of the bar over the appropriate class interval and is shown next. The distribution is relatively mound-shaped.



1.4.2 Since the variable of interest can only take integer values, the classes can be chosen as the values 0, 1, 2, 3, 4, 5 and 6. The table containing the classes, their corresponding frequencies and their relative frequencies and the relative frequency histogram are shown next. The distribution is skewed to the right.

Number of Household Pets	Frequency	Relative Frequency
0	13	13/50 = .26
1	19	19/50 = .38
2	12	12/50 = .24
3	4	4/50 = .08
4	1	1/50 = .02
5	0	0/50 = .00
6	1	1/50 = .02
Total	50	50/50 = 1.00



1.4.3-8 The proportion of measurements falling in each interval is equal to the sum of the heights of the bars over that interval. Remember that the lower class boundary is included, but not the upper class boundary.

3. .20 + .40 + .15 = .75

4. .05 + .15 + .20 = .40

5. .05

6. .40 + .15 = .55

7. .15

8. .05 + .15 + .20 = .40

1.4.9 Answers will vary. The range of the data is $110-10=90$ and we need to use seven classes. Calculate $90/7=12.86$ which we choose to round up to 15. Convenient class boundaries are created, starting at 10: 10 to < 25, 25 to < 40, …, 100 to < 115.

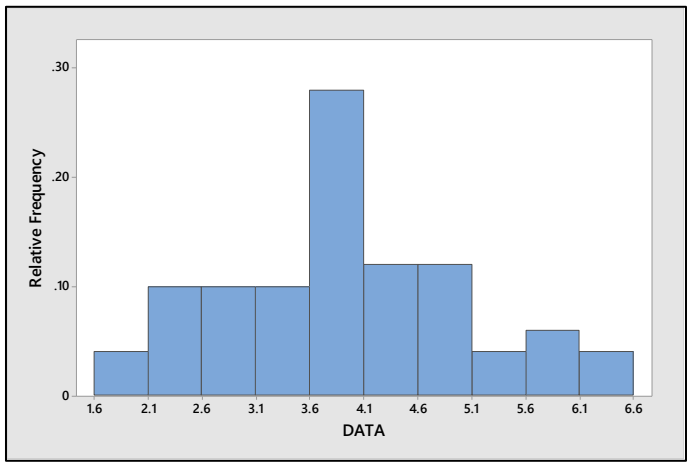
1.4.10 Answers will vary. The range of the data is $76.8-25.5=51.3$ and we need to use six classes. Calculate $51.3/6=8.55$ which we choose to round up to 9. Convenient class boundaries are created, starting at 25: 25 to < 34, 34 to < 43, …, 70 to < 79.

1.4.11 Answers will vary. The range of the data is $1.73-.31=1.42$ and we need to use ten classes. Calculate $1.42/10=.142$ which we choose to round up to .15. Convenient class boundaries are created, starting at .30: .30 to < .45, .45 to < .60, …, 1.65 to < 1.80.

1.4.12 Answers will vary. The range of the data is $192-0=192$ and we need to use eight classes. Calculate $192/8=24$ which we choose to round up to 25. Convenient class boundaries are created, starting at 0: 0 to < 25, 25 to < 50, …, 175 to < 200.

1.4.13-16 The table containing the classes, their corresponding frequencies and their relative frequencies and the relative frequency histogram are shown next.

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	1.6 to < 2.1	11	2	.04
2	2.1 to < 2.6	11111	5	.10
3	2.6 to < 3.1	11111	5	.10
4	3.1 to < 3.6	11111	5	.10
5	3.6 to < 4.1	11111 11111 1111	14	.28
6	4.1 to < 4.6	11111 11	7	.14
7	4.6 to < 5.1	11111	5	.10
8	5.1 to < 5.6	11	2	.04
9	5.6 to < 6.1	111	3	.06
10	6.1 to < 6.6	11	2	.04



13. The distribution is roughly mound-shaped.

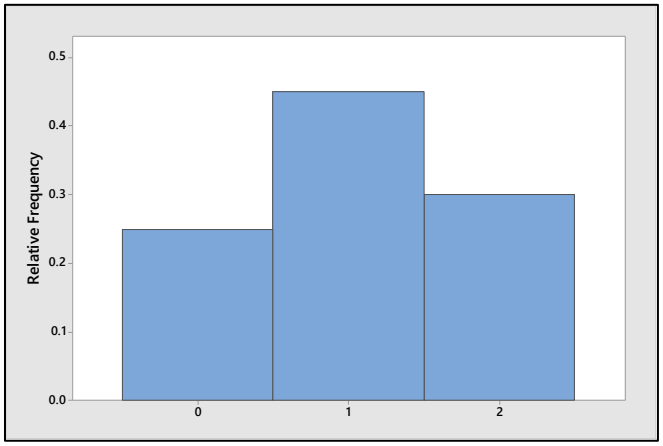
14. The fraction less than 5.1 is that fraction lying in classes 1-7, or $(2+5+\cdots+7+5)/50 = 43/50 = 0.86$.

15. The fraction larger than 3.6 lies in classes 5-10, or $(14+7+\cdots+3+2)/50 = 33/50 = 0.66$.

16. The fraction from 2.6 up to but not including 4.6 lies in classes 3-6, or $(5+5+14+7)/50 = 31/50 = 0.62$.

1.4.17-20 Since the variable of interest can only take the values 0, 1, or 2, the classes can be chosen as the integer values 0, 1, and 2. The table shows the classes, their corresponding frequencies and their relative frequencies. The relative frequency histogram follows the table.

Value	Frequency	Relative Frequency
0	5	.25
1	9	.45
2	6	.30



17. Using the table above, the proportion of measurements greater than 1 is the same as the proportion of "2"s, or 0.30.

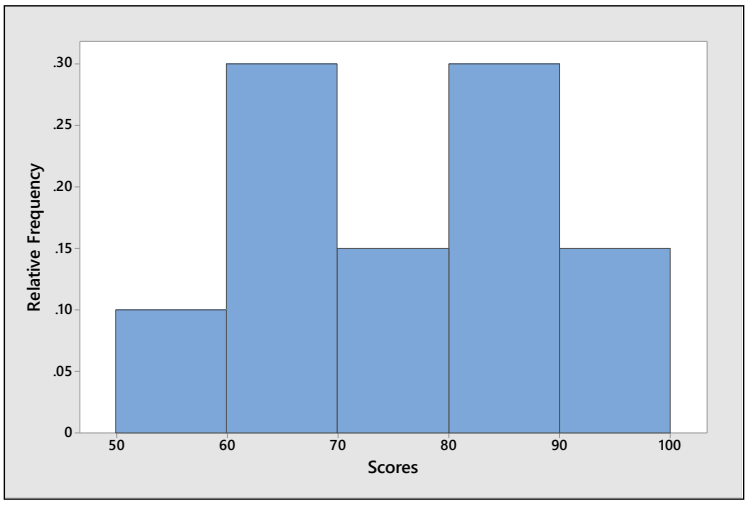
18. The proportion of measurements less than 2 is the same as the proportion of "0"s and "1"s, or $0.25+0.45=.70$.

19. The probability of selecting a "2" in a random selection from these twenty measurements is $6/20=.30$.

20. There are no outliers in this relatively symmetric, mound-shaped distribution.

1.4.21-23 Answers will vary. The range of the data is $94-55=39$ and we choose to use 5 classes. Calculate $39/5=7.8$ which we choose to round up to 10. Convenient class boundaries are created, starting at 50 and the table and relative frequency histogram are created.

Class Boundaries	Frequency	Relative Frequency
50 to < 60	2	.10
60 to < 70	6	.30
70 to < 80	3	.15
80 to < 90	6	.30
90 to < 100	3	.15



21. The distribution has two peaks at about 65 and 85. Depending on the way in which the student constructs the histogram, these peaks may or may not be clearly seen.

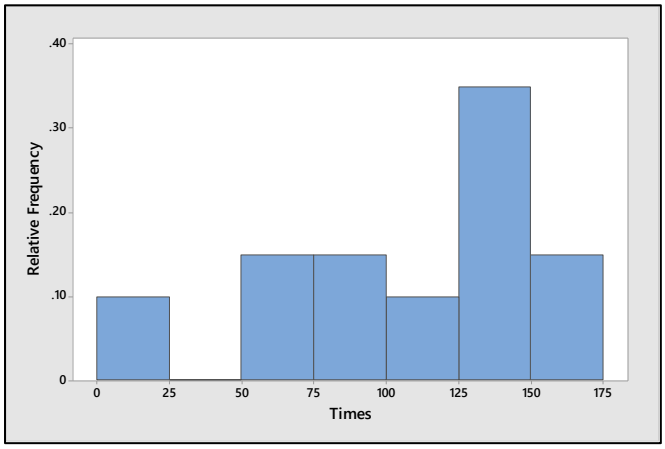
22. The shape is unusual. It might indicate that the students are divided into two groups – those who understand the material and do well on exams, and those who do not have a thorough command of the material.

23. The shapes are roughly the same, but this may not be the case if the student constructs the histogram using different class boundaries.

1.4.24 **a** There are a few extremely small numbers, indicating that the distribution is probably skewed to the left.

b The range of the data $165-8=157$. We choose to use seven class intervals of length 25, with subintervals 0 to < 25, 25 to < 50, 50 to < 75, and so on. The tally and relative frequency histogram are shown next.

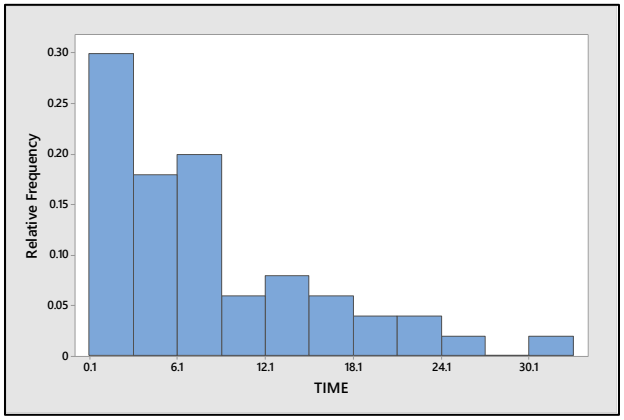
Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	0 to < 25	11	2	2/20
2	25 to < 50		0	0/20
3	50 to < 75	111	3	3/20
4	75 to < 100	111	3	3/20
5	100 to < 125	11	2	2/20
6	125 to < 150	11111 11	7	7/20
7	150 to < 175	111	3	3/20



c The distribution is indeed skewed left with two possible outliers: $x = 8$ and $x = 11$.

1.4.25 **a** The range of the data $32.3-0.2=32.1$. We choose to use eleven class intervals of length 3 ($32.1/11=2.9$, which when rounded to the next largest integer is 3). The subintervals 0.1 to < 3.1, 3.1 to < 6.1, 6.1 to < 9.1, and so on, are convenient and the tally and relative frequency histogram are shown next.

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	0.1 to < 3.1	11111 11111 11111	15	15/50
2	3.1 to < 6.1	11111 1111	9	9/50
3	6.1 to < 9.1	11111 11111	10	10/50
4	9.1 to < 12.1	111	3	3/50
5	12.1 to < 15.1	1111	4	4/50
6	15.1 to < 18.1	111	3	3/50
7	18.1 to < 21.1	11	2	2/50
8	21.1 to < 24.1	11	2	2/50
9	24.1 to < 37.1	1	1	1/50
10	27.1 to < 30.1		0	0/50
11	30.1 to < 33.1	1	1	1/50

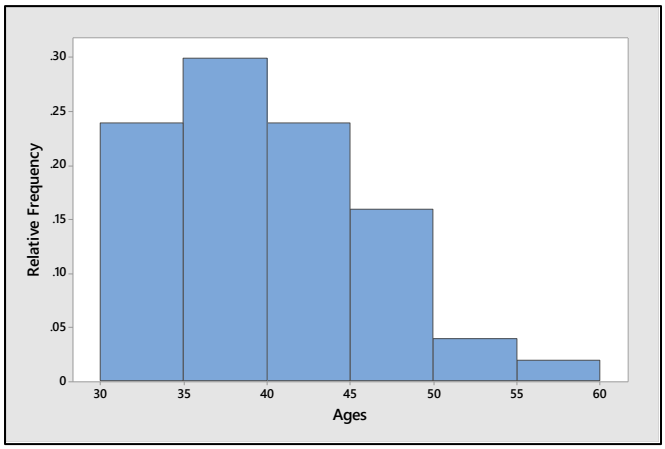


b The data is skewed to the right, with a few unusually large measurements.

c Looking at the data, we see that 36 patients had a disease recurrence within 10 months. Therefore, the fraction of recurrence times less than or equal to 10 is $36/50 = 0.72$.

1.4.26 **a** We use class intervals of length 5, beginning with the subinterval 30 to < 35. The tally and the relative frequency histogram are shown next.

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	30 to < 35	11111 11111 11	12	12/50
2	35 to < 40	11111 11111 11111	15	15/50
3	40 to < 45	11111 11111 11	12	12/50
4	45 to < 50	11111 111	8	8/50
5	50 to < 55	11	2	2/50
6	55 to < 60	1	1	1/50



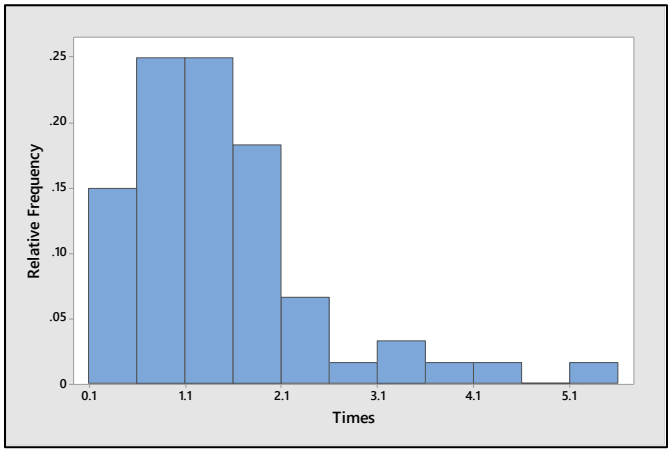
b Use the table or the relative frequency histogram. The proportion of children in the interval 35 to < 45 is (15 + 12)/50 = .54.

c The proportion of children aged less than 50 months is (12 + 15 + 12 + 8)/50 = .94.

1.4.27 **a** The data ranges from .2 to 5.2, or 5.0 units. Since the number of class intervals should be between five and twelve, we choose to use eleven class intervals, with each class interval having length 0.50 ($5.0/11 = .45$, which, rounded to the nearest convenient fraction, is .50). We must now select interval boundaries such that no measurement can fall on a boundary point. The subintervals .1 to < .6, .6 to < 1.1, and so on, are convenient and a tally is constructed.

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	0.1 to < 0.6	11111 11111	10	.167
2	0.6 to < 1.1	11111 11111 11111	15	.250
3	1.1 to < 1.6	11111 11111 11111	15	.250
4	1.6 to < 2.1	11111 11111	10	.167
5	2.1 to < 2.6	1111	4	.067
6	2.6 to < 3.1	1	1	.017
7	3.1 to < 3.6	11	2	.033
8	3.6 to < 4.1	1	1	.017
9	4.1 to < 4.6	1	1	.017
10	4.6 to < 5.1		0	.000
11	5.1 to < 5.6	1	1	.017

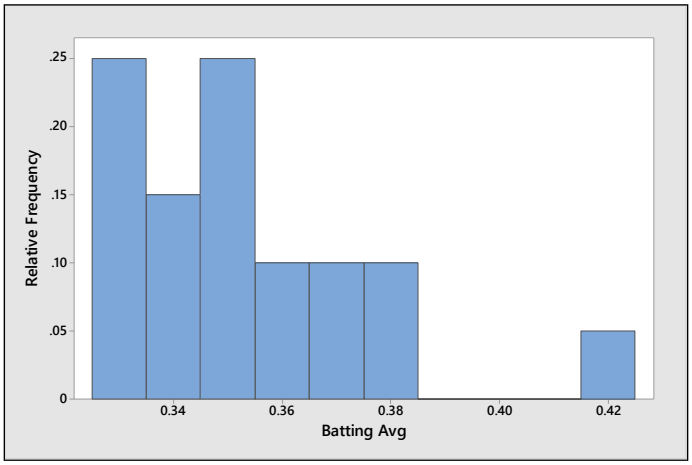
The relative frequency histogram is shown next.



b The distribution is skewed to the right, with several unusually large observations.

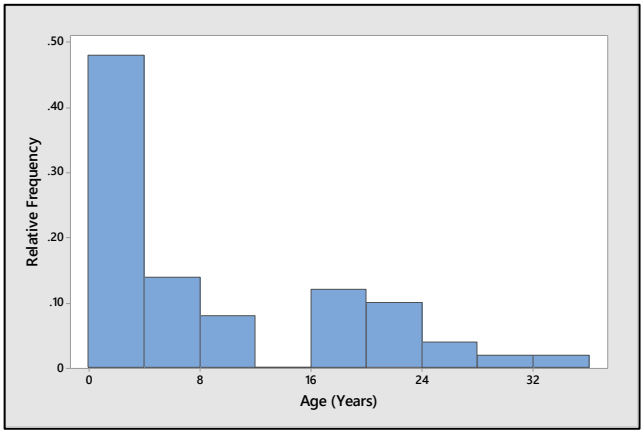
c For some reason, one person had to wait 5.2 minutes. Perhaps the supermarket was understaffed that day, or there may have been an unusually large number of customers in the store.

1.4.28 **a** Histograms will vary from student to student. A typical histogram generated by ***Minitab*** is shown next.

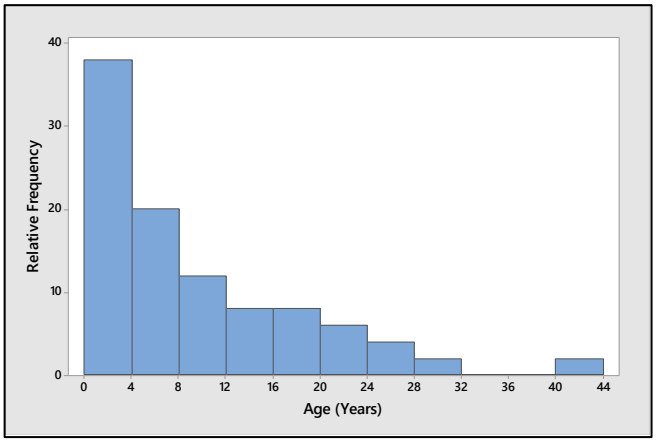


b Since 1 of the 20 players has an average above 0.400, the chance is 1 out of 20 or $1/20 = 0.05$.

1.4.29 **a-b** Answers will vary from student to student. The students should notice that the distribution is skewed to the right with a few pennies being unusually old. A typical histogram is shown next.



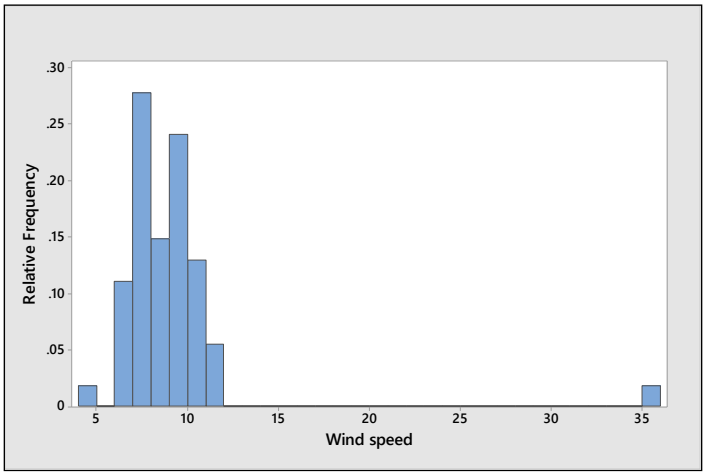
1.4.30 **a** Answers will vary from student to student. A typical histogram is shown next. It looks very similar to the histogram from Exercise 1.4.29.



b There is one outlier, $x = 41$.

1.4.31 **a** Answers will vary from student to student. The relative frequency histogram below was constructed using classes of length 1.0 starting at $x = 4$. The value $x = 35.1$ is not shown in the table but appears on the graph shown next.

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	4.0 to < 5.0	1	1	1/54
2	5.0 to < 6.0	0	0	0/54
3	6.0 to < 7.0	11111 1	6	6/54
4	7.0 to < 8.0	11111 11111 11111	15	15/54
5	8.0 to < 9.0	11111 111	8	8/54
6	9.0 to < 10.0	11111 11111 111	13	13/54
7	10.0 to < 11.0	11111 11	7	7/54
8	11.0 to < 12.0	111	3	3/54



b Since Mt. Washington is a very mountainous area, it is not unusual that the average wind speed would be very high.

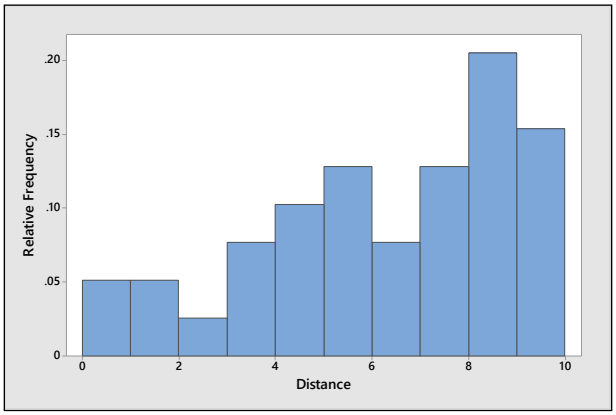
c The value $x = 9.9$ does not lie far from the center of the distribution (excluding $x = 35.1$). It would not be considered unusually high.

1.4.32 **a-b** The data is somewhat mound-shaped, but it appears to have two local peaks – high points from which the frequencies drop off on either side.

c Since these are student heights, the data can be divided into two groups – heights of males and heights of females. Both groups will have an approximate mound-shape, but the average female height will be lower than the average male height. When the two groups are combined into one data set, it causes a "mixture" of two mound-shaped distributions and produces the two peaks seen in the histogram.

1.4.33 **a** The relative frequency histogram below was constructed using classes of length 1.0 starting at $x = 0.0$.

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	0.0 to < 1.0	11	2	2/39
2	1.0 to < 2.0	11	2	2/39
3	2.0 to < 3.0	1	1	1/39
4	3.0 to < 4.0	111	3	3/39
5	4.0 to < 5.0	111	4	4/39
6	5.0 to < 6.0	11111	5	5/39
7	6.0 to < 7.0	111	3	3/39
8	7.0 to < 8.0	11111	5	5/39
9	8.0 to < 9.0	11111 111	8	8/39
10	9.0 to < 10.0	11111 1	6	6/39



a The distribution is skewed to the left, with slightly higher frequency in the first two classes (within two miles of UCR).

b As the distance from UCR increases, each successive area increases in size, thus allowing for more Starbucks stores in that region.

Reviewing What You've Learned

1.R.1 **a** "Ethnic origin" is a *qualitative variable* since a quality (ethnic origin) is measured.

b "Score" is a *quantitative variable* since a numerical quantity (0-100) is measured.

c "Type of establishment" is a *qualitative variable* since a category (Carl's Jr., McDonald's or Burger King) is measured.

d "Mercury concentration" is a *quantitative variable* since a numerical quantity is measured.

1.R.2 To determine whether a distribution is likely to be skewed, look for the likelihood of observing extremely large or extremely small values of the variable of interest.

a The price of an 8-oz can of peas is not likely to contain unusually large or small values.

b Not likely to be skewed.

c If a package is dropped, it is likely that all the shells will be broken. Hence, a few large number of broken shells is possible. The distribution will be skewed.

d If an animal has one tick, he is likely to have more than one. There will be some "0"s with uninfected rabbits, and then a larger number of large values. The distribution will not be symmetric.

1.R.3 **a** The length of time between arrivals at an outpatient clinic is a continuous random variable, since it can be any of the infinite number of positive real values.

b The time required to finish an examination is a continuous random variable as was the random variable described in part **a**.

c Weight is continuous, taking any positive real value.

d Body temperature is continuous, taking any real value.

e Number of people is discrete, taking the values 0, 1, 2, …

1.R.4 **a** Number of properties is discrete, taking the values 0, 1, 2, …

b Depth is continuous, taking any non-negative real value.

c Length of time is continuous, taking any non-negative real value.

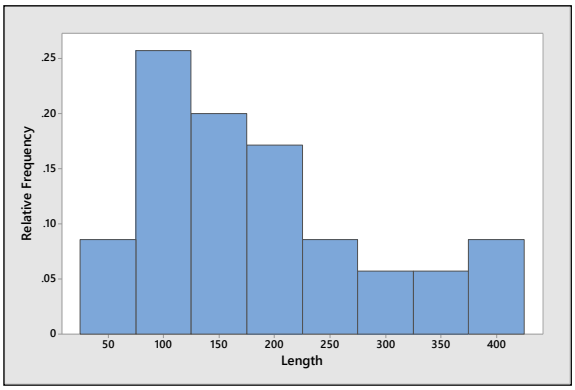
d Number of aircraft is discrete.

1.R.5 **a** **b**

Stem-and-leaf of Length N = 35

7	0	6779999
(11)	1	00122334444
17	1	5799
13	2	0004
9	2	669
6	3	0
5	3	5679
1	4	2

Leaf Unit = 10

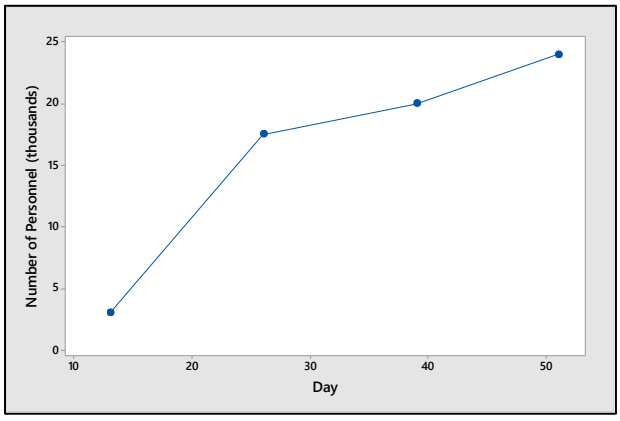


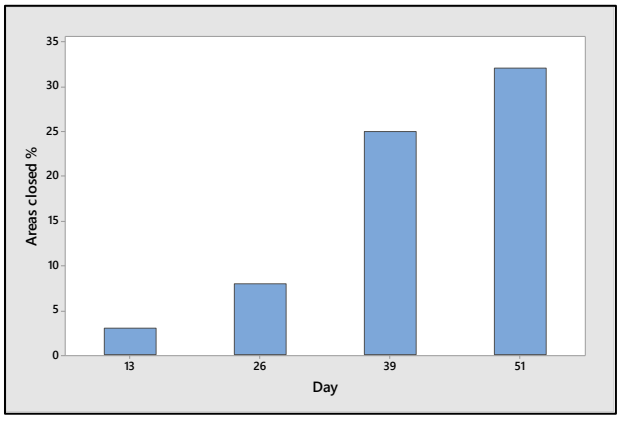
c These data are skewed right.

1.R.6 **a** The five quantitative variables are measured over time two months after the oil spill. Some sort of comparative bar charts (side-by-side or stacked) or a line chart should be used.

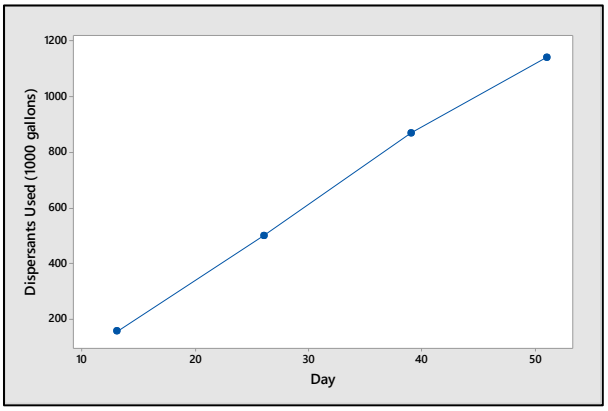
b As the time after the spill increases, the values of all five variables increase.

c-d The line chart for *number of personnel* and the bar chart for *fishing areas closed* are shown next.



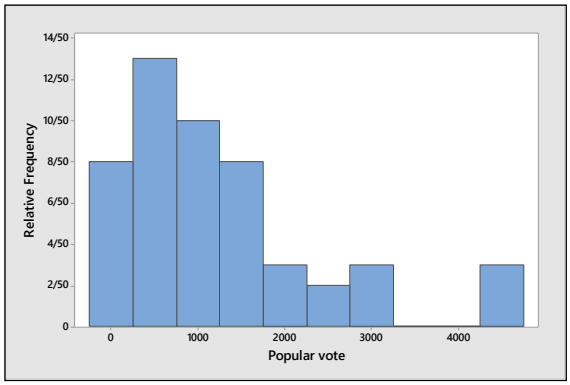


e The line chart for *amount of dispersants* is shown next. There appears to be a straight-line trend.

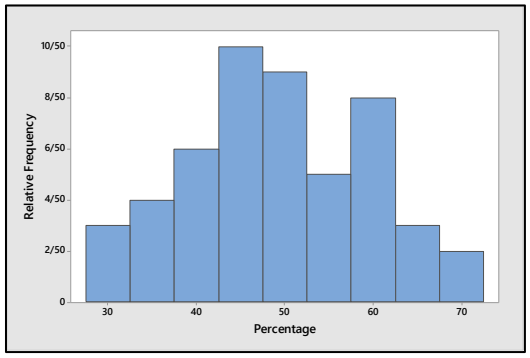


1.R.7 **a** The popular vote within each state should vary depending on the size of the state. Since there are several very large states (in population) in the United States, the distribution should be skewed to the right.

b-c Histograms will vary from student to student but should resemble the histogram generated by ***Minitab*** in the next figure. The distribution is indeed skewed to the right, with three "outliers" – California, Florida and Texas.



1.R.8 **a-b** Once the size of the state is removed by calculating the percentage of the popular vote, the unusually large values in the Exercise 7 data set will disappear, and each state will be measured on an equal basis. Student histograms should resemble the histogram shown next. Notice the relatively mound-shape and the lack of any outliers.



1.R.9 **a-b** Popular vote is skewed to the right while the percentage of popular vote is roughly mound-shaped. While the distribution of popular vote has outliers (California, Florida and Texas), there are no outliers in the distribution of percentage of popular vote. When the stem and leaf plots are turned 90°, the shapes are very similar to the histograms.

c Once the size of the state is removed by calculating the percentage of the popular vote, the unusually large values in the set of "popular votes" will disappear, and each state will be measured on an equal basis. The data then distribute themselves in a mound-shape around the average percentage of the popular vote.

1.R.10 **a** The measurements are obtained by counting the number of beats for 30 seconds, and then multiplying by 2. Thus, the measurements should all be even numbers.

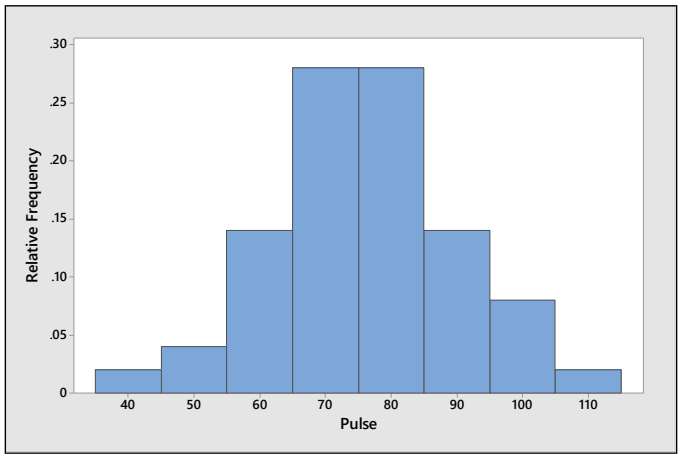
b The stem and leaf plot is shown next.

Stem-and-leaf of Pulse N = 50

1	4	2
1	4	
3	5	24
6	5	688
10	6	0022
15	6	66668
24	7	000222224
25	7	8
25	8	0022444444444
12	8	68888
7	9	00
5	9	66
3	10	04
1	10	
1	11	0

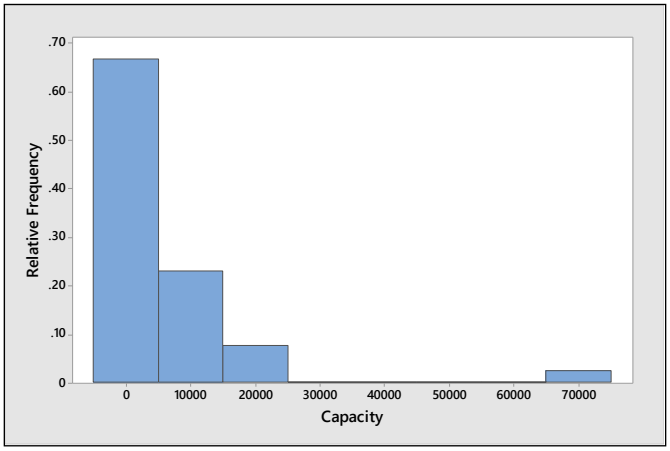
Leaf Unit = 1

c Answers will vary. A typical histogram, generated by ***Minitab***, is shown next.



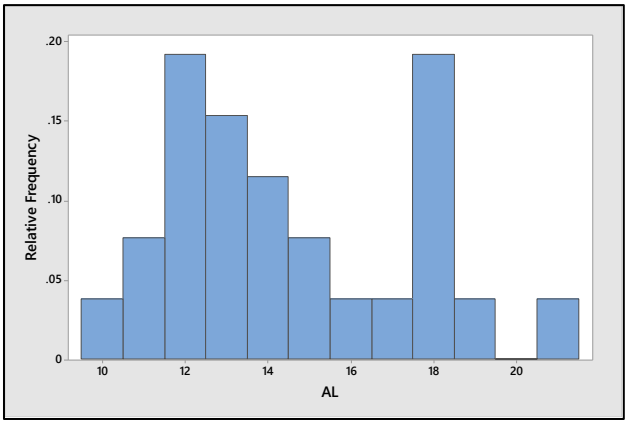
d The distribution of pulse rates is mound-shaped and relatively symmetric around a central location of 75 beats per minute. There are no outliers.

1.R.11 **a-b** Answers will vary from student to student. A typical histogram is shown next—the distribution is skewed to the right, with an extreme outlier (Texas).

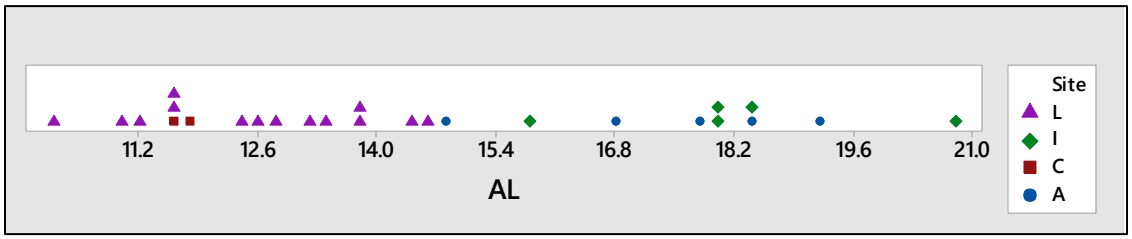


c Answers will vary.

1.R.12 **a-b** Answers will vary. A typical histogram is shown next. Notice the gaps and the bimodal nature of the histogram, probably due to the fact that the samples were collected at different locations.



c The dotplot is shown as follows. The locations are indeed responsible for the unusual gaps and peaks in the relative frequency histogram.



1.R.13 **a-b** The ***Minitab*** stem and leaf plot is shown next. The distribution is slightly skewed to the right.

Stem-and-Leaf Display: Gas Tax

Stem-and-leaf of Gas Tax N = 51

1	1	2
7	1	677889
20	2	0011222333344
(7)	2	6678889
24	3	000011223333444
9	3	6789
5	4	034
2	4	9
1	5	
1	5	8

Leaf Unit = 1

c Pennsylvania (58.20) has an unusually high gas tax.

1.R.14 **a-b** Answers will vary. The ***Minitab*** stem and leaf plot is shown next. The distribution is skewed to the right.

Stem-and-leaf of Megawatts N = 15

(9)	0	556666668
6	1	01344
1	2	2

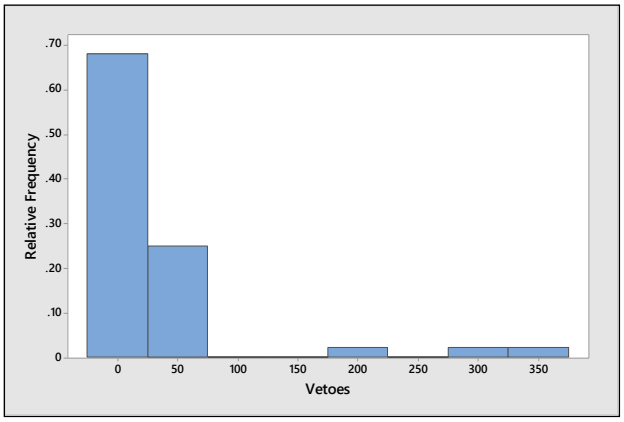
Leaf Unit = 1000

1.R.15 **a-b** The distribution is approximately mound-shaped, with one unusual measurement, in the class with midpoint at 100.8°. Perhaps the person whose temperature was 100.8 has some sort of illness coming on?

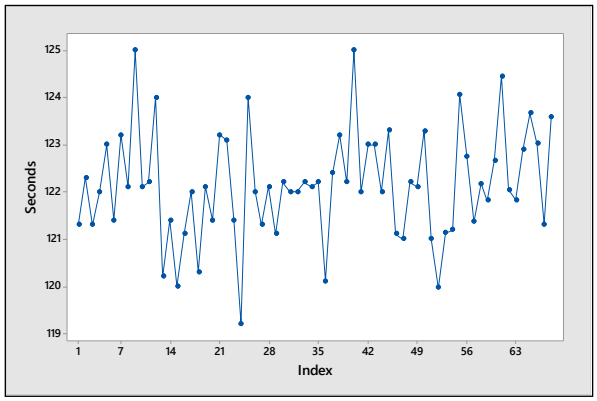
c The value 98.6° is slightly to the right of center.

On Your Own

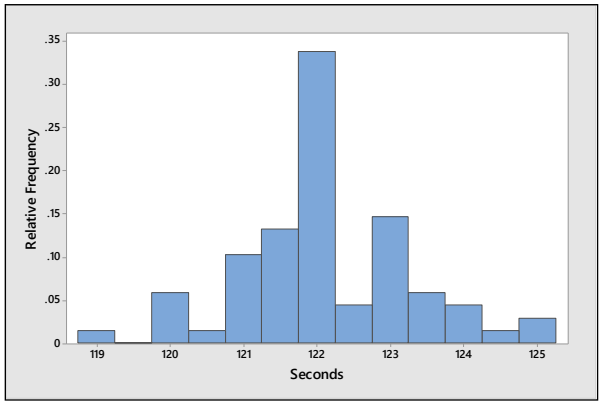
1.R.16 Answers will vary from student to student. The students should notice that the distribution is skewed to the right with a few presidents (Truman, Cleveland, and F.D. Roosevelt) casting an unusually large number of vetoes.



1.R.17 **a** The line chart is shown next. The year in which a horse raced does not appear to have an effect on his winning time.

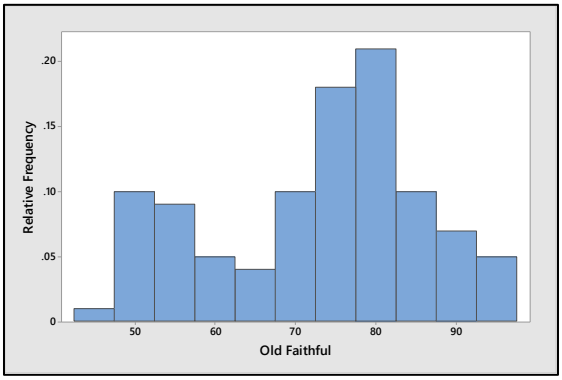


b Since the year of the race is not important in describing the data set, the distribution can be described using a relative frequency histogram. The distribution that follows is roughly mound-shaped with an unusually fast ($x = 119.2$) race times the year that *Secretariat* won the derby.

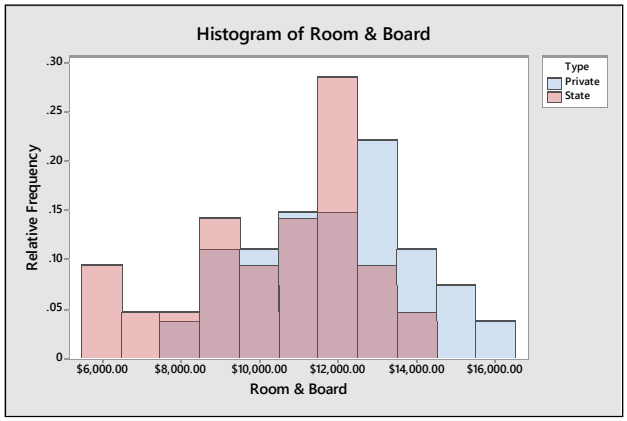


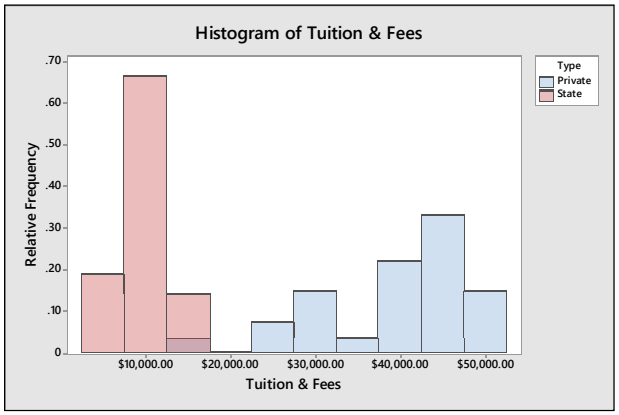
1.R.18 Answers will vary from student to student. Students should notice that both distributions are skewed left. The higher peak with a low bar to its left in the laptop group may indicate that students who would generally receive average scores (65-75) are scoring higher than usual. This may or may not be *caused* by the fact that they used laptop computers.

1.R.19 Answers will vary. A typical relative frequency histogram is shown next. There is an unusual bimodal feature.



1.R.20 Answers will vary. The student should notice that there is a clear difference in tuition and fees between private and state schools. Both distributions are roughly mound-shaped. The distribution of room and board costs are also roughly mound-shaped, but there is less of a difference between private and state schools.





CASE STUDY: How is Your Blood Pressure?

1. The following variables have been measured on the participants in this study: sex (qualitative); age in years (quantitative discrete); diastolic blood pressure (quantitative continuous, but measured to an integer value) and systolic blood pressure (quantitative continuous, but measured to an integer value). For each person, both systolic and diastolic readings are taken, making the data bivariate.

2. The important variables in this study are diastolic and systolic blood pressure, which can be described singly with histograms in various categories (male vs. female or by age categories). Further, the relationship between systolic and diastolic blood pressure can be displayed together using a scatterplot or a bivariate histogram.

3. Answers will vary from student to student, depending on the choice of class boundaries or the software package which is used. The histograms should look fairly mound-shaped.

4. Answers will vary from student to student.

5. In determining how a student's blood pressure compares to those in a comparable sex and age group, female students (ages 15-20) must compare to the population of females, while male students (ages 15-20) must compare to the population of males. The student should use his or her blood pressure and compare it to the scatterplot generated in part 4.