

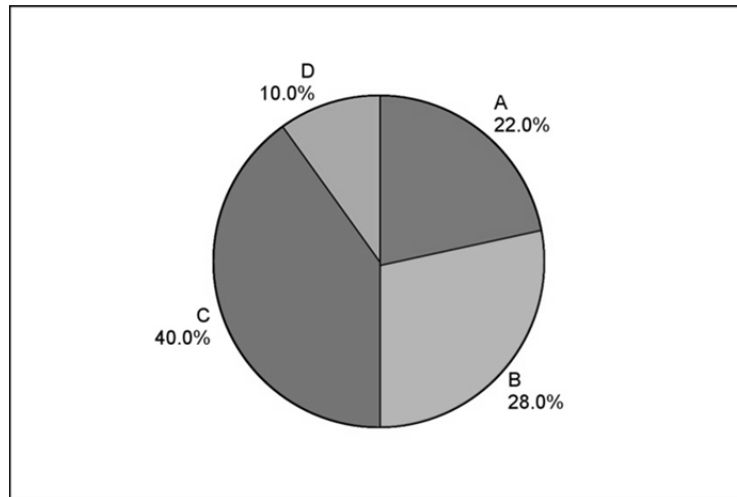
Chapter 1: Describing Data with Graphs

- 1.1**
- a** The experimental unit, the individual or object on which a variable is measured, is the student.
 - b** The experimental unit on which the number of errors is measured is the exam.
 - c** The experimental unit is the patient.
 - d** The experimental unit is the azalea plant.
 - e** The experimental unit is the car.
- 1.2**
- a** “Time to assemble” is a *quantitative* variable because a numerical quantity (1 hour, 1.5 hours, etc.) is measured.
 - b** “Number of students” is a *quantitative* variable because a numerical quantity (1, 2, etc.) is measured.
 - c** “Rating of a politician” is a *qualitative* variable since a quality (excellent, good, fair, poor) is measured.
 - d** “Province or territory of residence” is a *qualitative* variable since a quality (ON, AB, BC, etc.) is measured.
- 1.3**
- a** “Population” is a *discrete* variable because it can take on only integer values.
 - b** “Weight” is a *continuous* variable, taking on any values associated with an interval on the real line.
 - c** “Time” is a *continuous* variable.
 - d** “Number of consumers” is integer-valued and hence *discrete*.
- 1.4**
- a** “Number of boating accidents” is integer-valued and hence *discrete*.
 - b** “Time” is a *continuous* variable.
 - c** “Choice of colour” is a *qualitative* variable since a quality (white, cream, black, etc.) is measured.
 - d** “Number of brothers and sisters” is integer-valued and hence *discrete*.
 - e** “Yield in kilograms” is a *continuous* variable, taking on any values associated with an interval on the real line.
- 1.5**
- a** The experimental unit, the item or object on which variables are measured, is the vehicle.
 - b** Type (qualitative); make (qualitative); carpool (qualitative); one-way commute distance (quantitative continuous); age of vehicle (quantitative continuous)
 - c** Since five variables have been measured, the data is *multivariate*.
- 1.6**
- a** The set of ages at death represents a population, because there have only been 15 different prime ministers in Canadian history.
 - b** The variable being measured is the continuous variable “age.”
 - c** “Age” is a quantitative variable.
- 1.7**
- 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. 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.
- 1.8**
- a–b** The variable “survival times” 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.

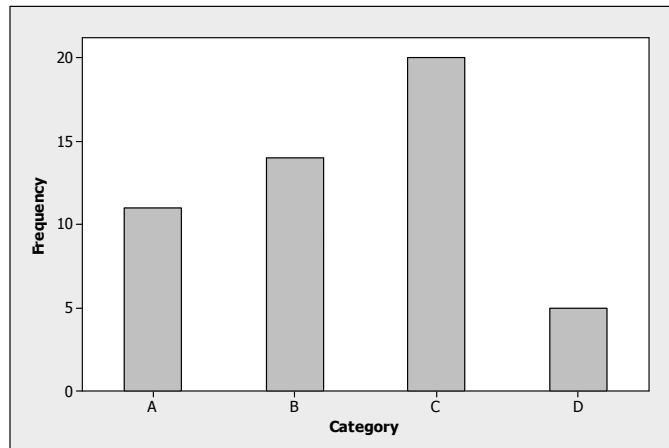
- 1.9**
- a** The variable “reading score” is a quantitative variable, which is usually integer-valued and hence discrete.
 - b** The individual on which the variable is measured is the student.
 - c** The population is hypothetical—it does not exist in fact—but consists of the reading scores for all students who could possibly be taught by this method.

- 1.10**
- a–b** The variable “category” is a qualitative variable measured for each of 50 people who constitute the experimental units.
 - c** 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 below.

Category	Frequency	Fraction of Total	Sector Angle
A	11	0.22	79.2
B	14	0.28	100.8
C	20	0.40	144.0
D	5	0.10	36.0



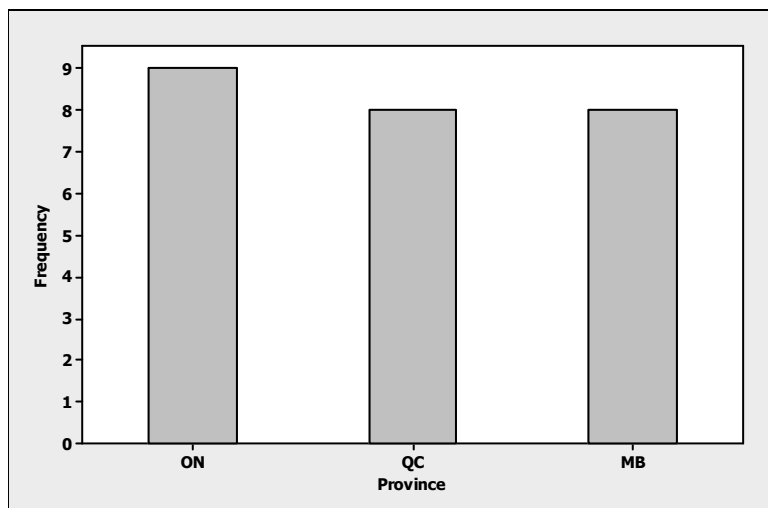
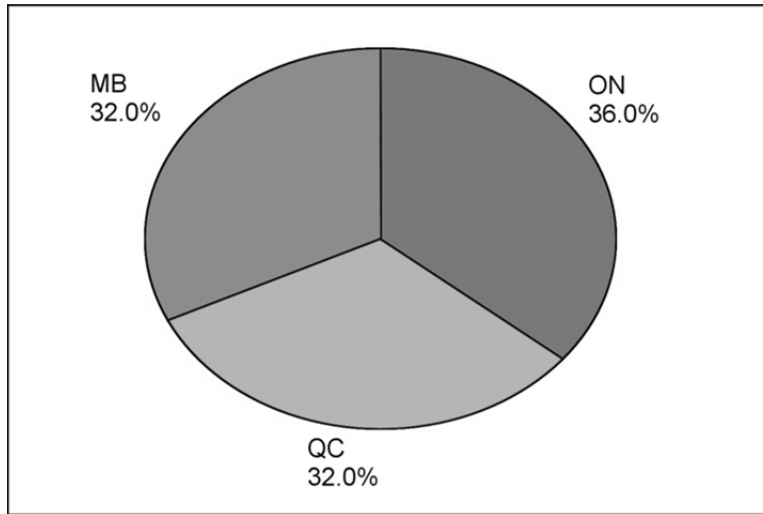
- d** The bar chart represents each category as a bar with height equal to the frequency of occurrence of that category and is shown in the figure below.



- e Yes, the shape will change depending on the order of presentation. The order is unimportant.
- f 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, $(14 + 20 + 5)/50 = 0.78$.
- g Since there are 14 people in category B, there are $50 - 14 = 36$ who are not, and the percentage is calculated as $(36/50)100 = 72\%$.

- 1.11 a–b The experimental unit is the pair of jeans, on which the qualitative variable “province” is measured.
 c–d First, construct a statistical table to summarize the data. The pie and bar charts are shown in the figures below.

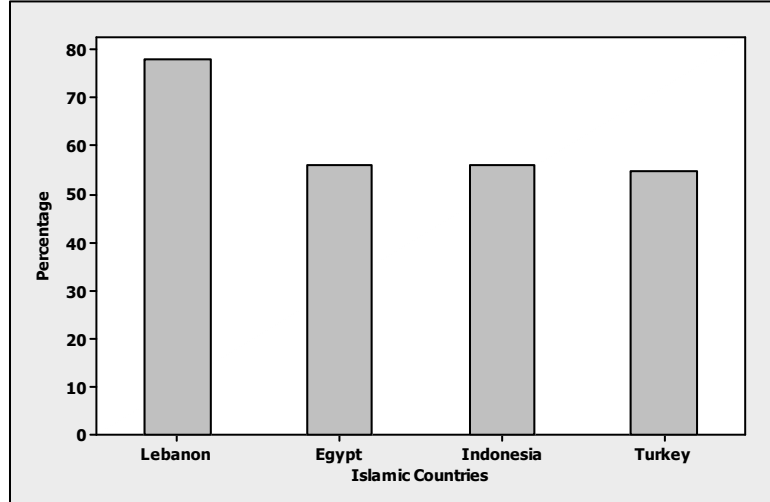
Province	Frequency	Fraction of Total	Sector Angle
ON	9	0.36	129.6
QC	8	0.32	115.2
MB	8	0.32	115.2



- e From the table or the chart, Quebec produced $8/25 = 0.32$ of the jeans.
- f The highest bar represents Ontario, which produced the most pairs of jeans.
- g Since the bars and the sectors are almost equal in size, the three provinces produced roughly the same number of pairs of jeans.

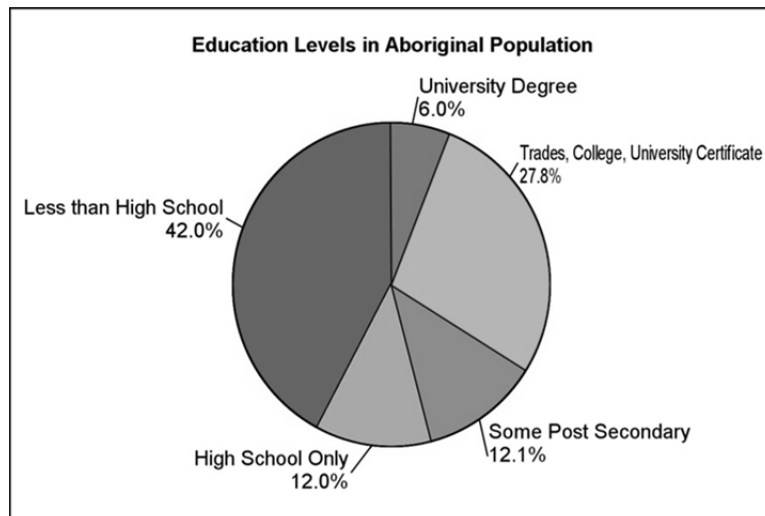
- 1.12**
- a** The population of interest consists of voter opinions (political or religious) on the conflict between Islam and the West.
 - b** The population from which the pollsters have sampled is the population of all adults from 27 countries (no further details available).
 - c** The percentages given in the exercise only add to 85%. We should add another category called "Other," which will account for the other 15% of the responses.
 - d** Answers will vary.

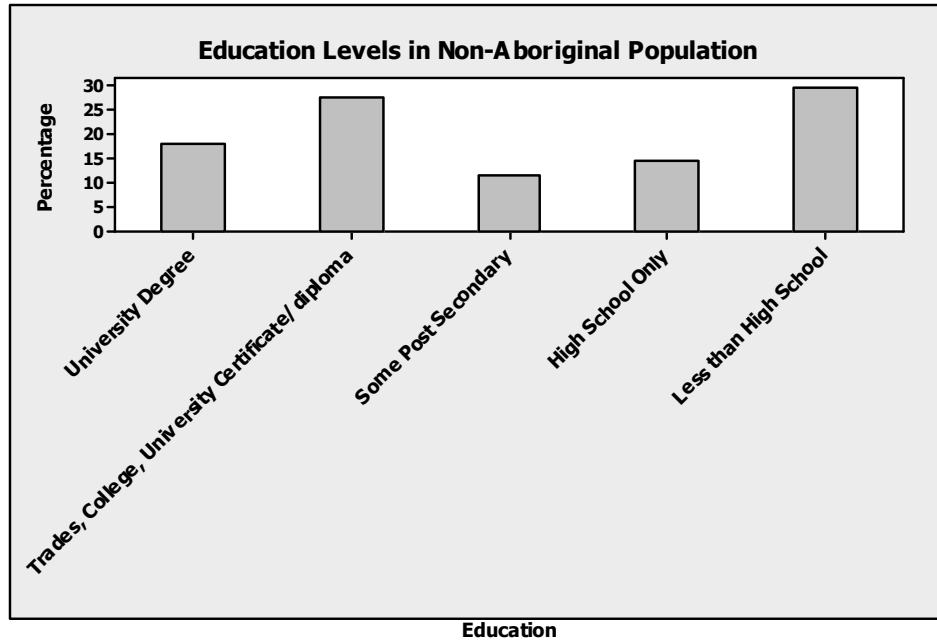
- 1.13**
- a** No, a few more Islamic countries (Iraq, Pakistan, Afghanistan, Syria, etc.) can be added in the table.
 - b** A bar chart is appropriate.
 - c**



- d** Answers will vary.

- 1.14**
- a–b** The variable being measured is a qualitative variable, which would be described as “educational attainment.”
 - c** The numbers represent the percentages of Aboriginal and non-Aboriginal population who fall in each of the five educational attainment categories.
 - d–e** The percentages falling in each of the five categories have already been calculated, and the pie chart (Aboriginal) and bar chart (non-Aboriginal) are shown in the figures below and on the next page.

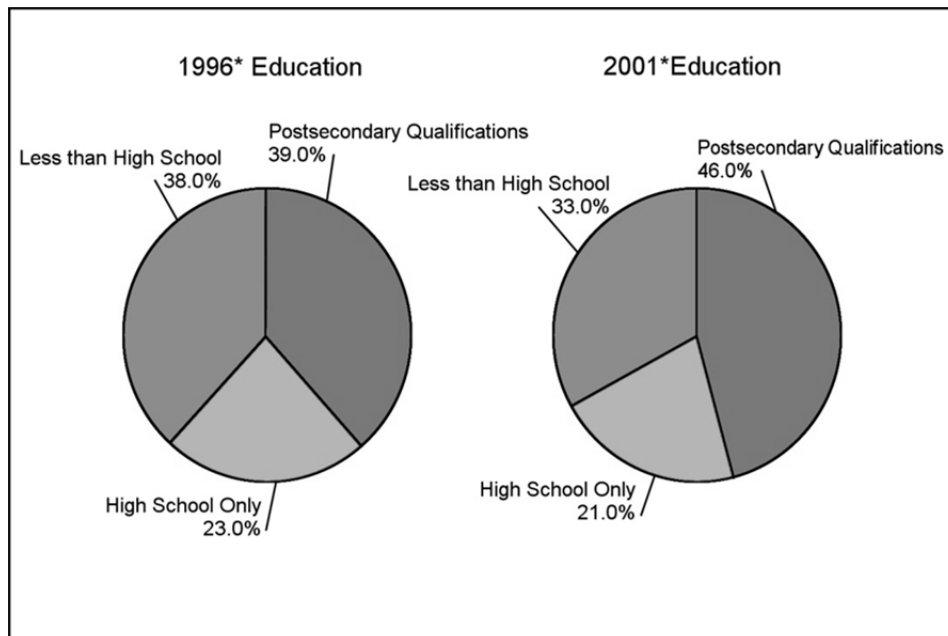




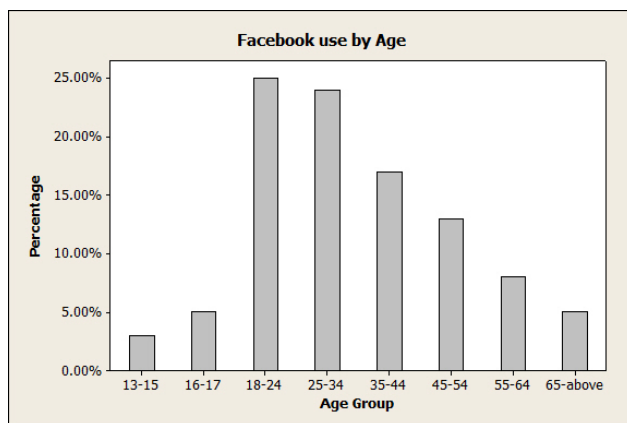
f There is a significant gap—only 6% of the members of Aboriginal population have a university degree; whereas more than three times (17.7%) of the non-Aboriginal population have university degree.

1.15

- a Yes. The total percentage of education level in each bar graph is 100.
- b Yes. There is a significant increase (from 39% to 46%) in the post-secondary education attainment over the years.
- c The pie chart is shown below. The bar chart is probably visually more interesting.



- 1.16**
- a** The variable being measured is “age of Facebook users.”
 - b** Although “age of Facebook users” is quantitative variable, the variable is recorded in age group categories, and hence it is a qualitative variable.
 - c** The percentages represent percentage of the Facebook users in different age groups.
 - d** The pie chart is constructed correctly since the percentages do add up to 100%.
 - e** The bar chart is shown below.

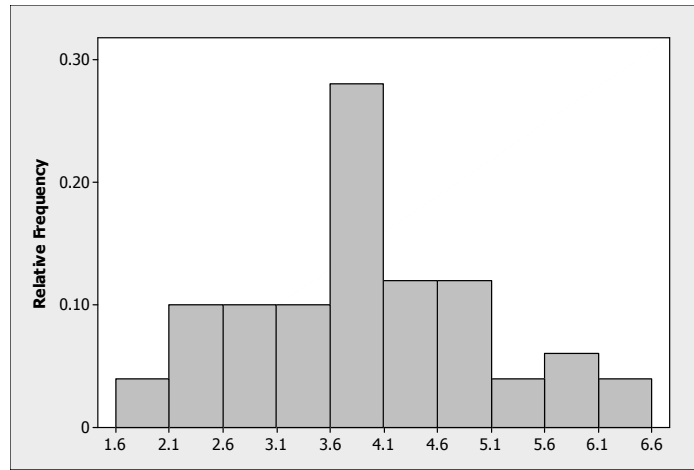


- f** The bar chart is easier to follow; the pie chart is visually more interesting.
 - g** Gender, type of device being used (laptop, desktop, tablet, phone), and Internet connection speed (high speed) may be some other interesting variables that one might want to explore.
- 1.17** The most obvious 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 plot is shown below.

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	

- a** The stem and leaf plot has a mound-shaped distribution.
 - b** From the stem and leaf plot, the smallest observation is 1.6 (1 6).
 - c** The eight and ninth largest observations are both 4.9 (4 9).
- 1.18**
- a** For $n = 5$, use between 8 and 10 classes.
 - b**

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	1.6 to < 2.1	11	2	0.04
2	2.1 to < 2.6	11111	5	0.10
3	2.6 to < 3.1	11111	5	0.10
4	3.1 to < 3.6	11111	5	0.10
5	3.6 to < 4.1	11111 11111 1111	14	0.28
6	4.1 to < 4.6	11111 11	7	0.14
7	4.6 to < 5.1	11111	5	0.10
8	5.1 to < 5.6	11	2	0.04
9	5.6 to < 6.1	111	3	0.06
10	6.1 to < 6.6	11	2	0.04



- c** From **b**, the fraction less than 5.1 is that fraction lying in classes 1–7, or $(2 + 5 + \dots + 7 + 5)/50 = 43/50 = 0.86$
- d** From **b**, the fraction larger than 3.6 lies in classes 5–10, or $(14 + 7 + \dots + 3 + 2)/50 = 33/50 = 0.66$
- e** The stem and leaf plot has a more peaked mound-shaped distribution than the relative frequency histogram because of the smaller number of groups.

- 1.19 a** As in Exercise 1.17, 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

- b** 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 plot shown below improves the presentation of the data.

```

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

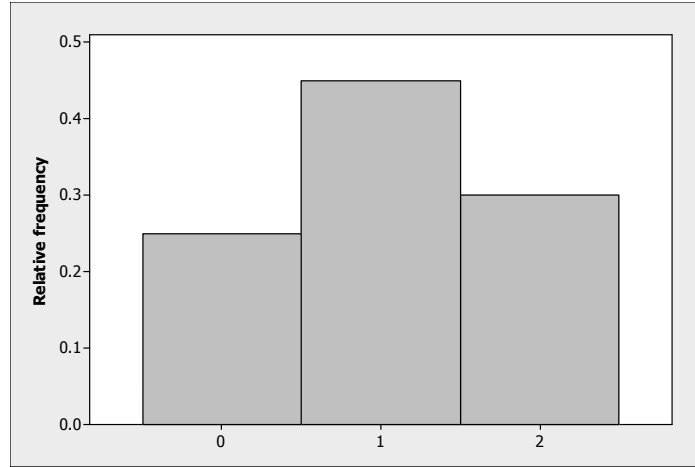
```

leaf digit = 0.1
1 2 represents 1.2

- 1.20 a** 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 below shows the classes, their corresponding frequencies, and their relative frequencies.

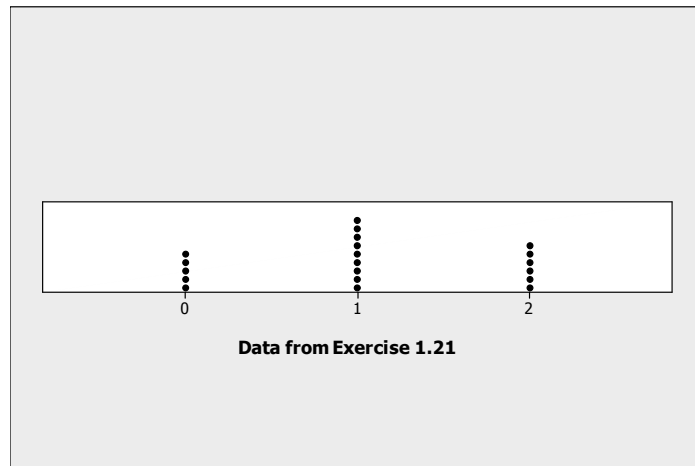
Value	Frequency	Relative Frequency
0	5	.25
1	9	.45
2	6	.30

The relative frequency histogram is shown on the next page.



- b** Using the table in part **a**, the proportion of measurements greater than 1 is the same as the proportion of “2,” or 0.30.
- c** The proportion of measurements less than 2 is the same as the proportion of “0” and “1,” or $0.25 + 0.45 = 0.70$.
- d** The probability of selecting a “2” in a random selection from these twenty measurements is $6/20 = 0.30$.
- e** There are no outliers in this relatively symmetric, mound-shaped distribution.

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

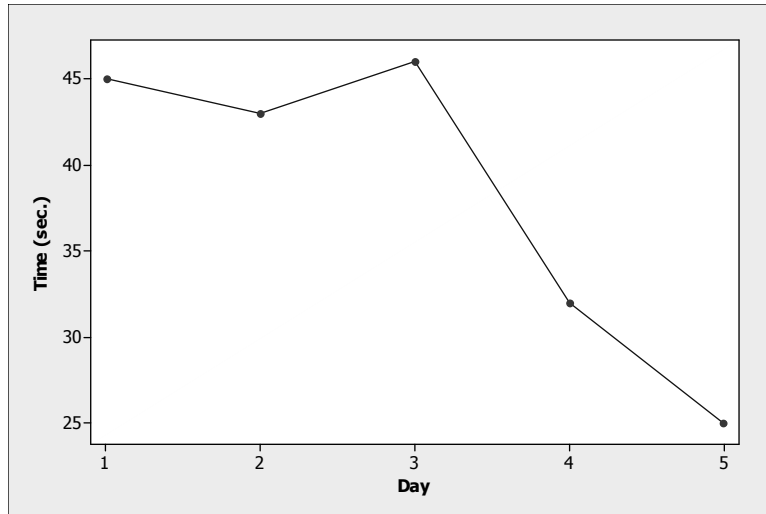


- b** Since there is only one digit in each measurement, the ones digit must be the stem, and the leaf will be a zero digit for each measurement.
- c**

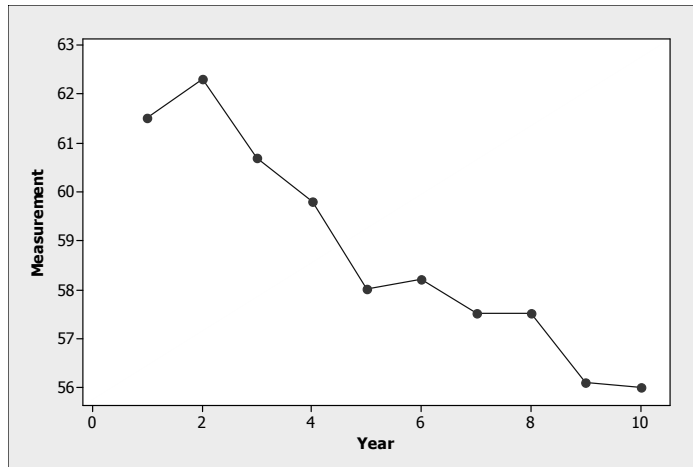
```

0 | 0 0 0 0 0
1 | 0 0 0 0 0 0 0 0 0
2 | 0 0 0 0 0 0
                    
```
- d** The two plots convey the same information if the stem and leaf plot is turned 90° and stretched to resemble the dotplot.

- 1.22 The line chart plots “day” on the horizontal axis and “time” on the vertical axis. The line chart shown below reveals that learning is taking place, since the time decreases each successive day.



- 1.23 a–b The line graph is shown below. Notice the change in y as x increases. The measurements are decreasing over time.



- 1.24 a The test scores are graphed using a stem and leaf plot generated by *MINITAB*.

Stem and Leaf Plot: Scores

Stem and leaf of Scores N = 20
Leaf Unit = 1.0

```

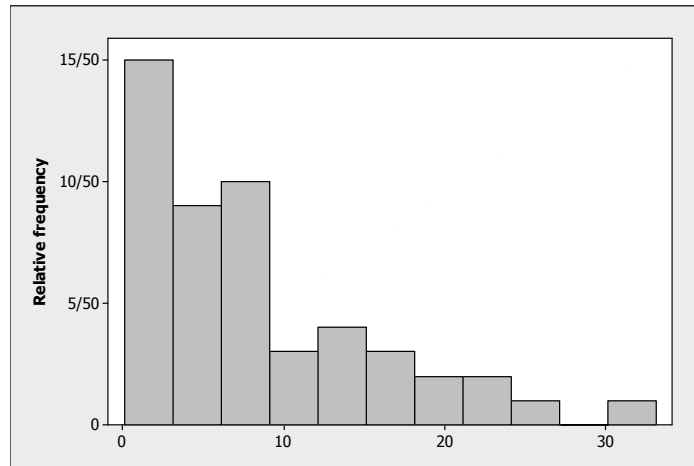
2  5  57
5  6  123
8  6  578
9  7  2
(2) 7  56
9  8  24
7  8  6679
3  9  134
    
```

- b–c The distribution is not mound-shaped, but is rather bimodal with two peaks centred 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.

- 1.25 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 $0.1 < 3.1$, $3.1 < 6.1$, $6.1 < 9.1$, and so on, are convenient and the tally is shown below.

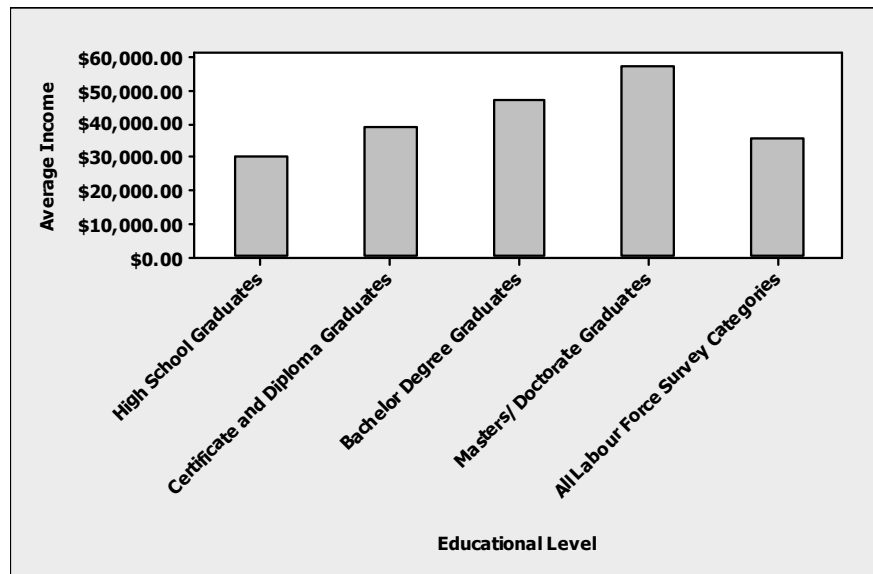
Class i	Class Boundaries	Tally	f_i	Relative Frequency, f_i/n
1	0.1 to < 3.1	11111 11111 11111	15	15/50
2	3.1 to < 6.1	11111 1111	9	9/50
3	6.1 to < 9.1	11111 11111	10	10/50
4	9.1 to < 12.1	111	3	3/50
5	12.1 to < 15.1	1111	4	4/50
6	15.1 to < 18.1	111	3	3/50
7	18.1 to < 21.1	11	2	2/50
8	21.1 to < 24.1	11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

The relative frequency histogram is shown below.



- 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 $36/50 = 0.72$.

- 1.26 a The data represent the average annual incomes of Albertans divided into five educational categories. A bar chart would be the most appropriate graphical method.
 b The bar chart is shown below.



- c The average salary for Albertans residents increases substantially as the person's educational level increases.

- 1.27 a Use the tens digit as the stem, and the ones digit as the leaf, dividing each stem into two parts.

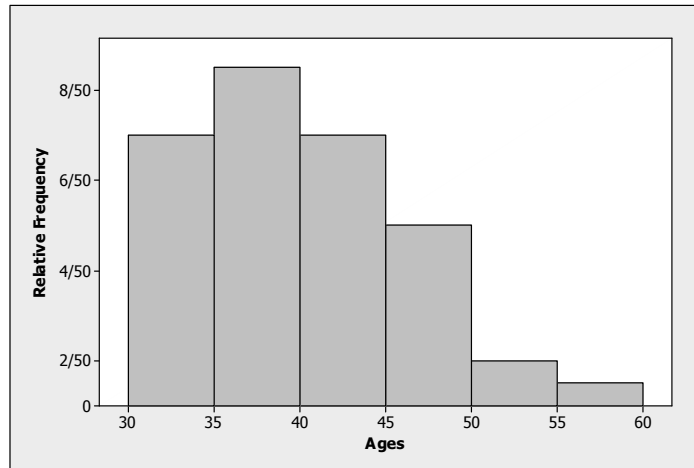
```

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

- b We use class intervals of length 5, beginning with the subinterval 30 to < 35. The tally is shown below.

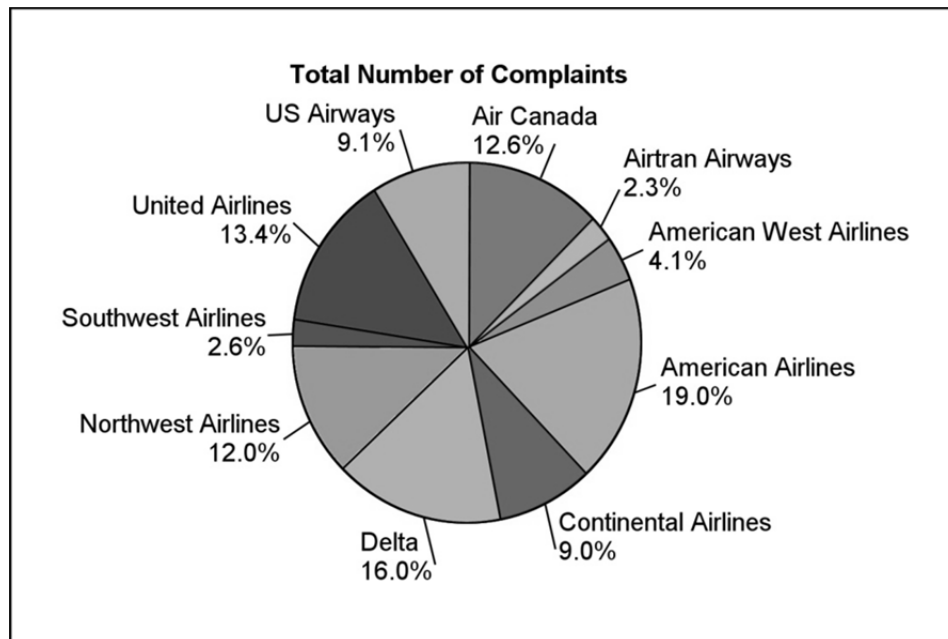
Class i	Class Boundaries	Tally	f_i	Relative Frequency, f_i/n
1	30 to < 35	11111 11111 11	12	12/50
2	35 to < 40	11111 11111 11111	15	15/50
3	40 to < 45	11111 11111 11	12	12/50
4	45 to < 50	11111 111	8	8/50
5	50 to < 55	11	2	2/50
6	55 to < 60	1	1	1/50

The relative frequency histogram is shown below.

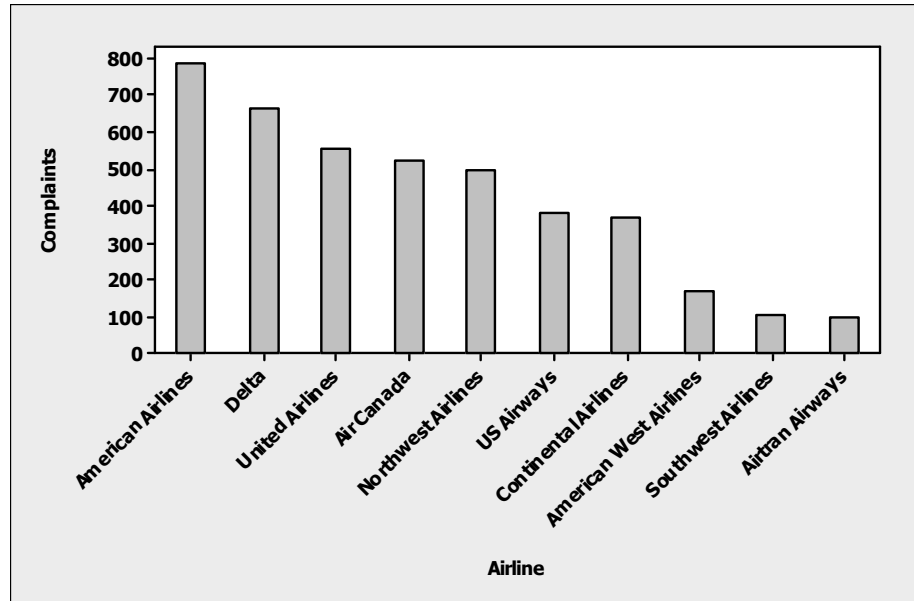


- c The two graphs are very similar, with the relative frequency histogram a bit more visually appealing. If the student chose to create the stem and leaf plot without splitting the stems into two parts, the stem and leaf plot would not be very helpful in describing the data set.
- d Use either the stem and leaf plot, the table, or the relative frequency histogram. The proportion of children in the interval 35 to < 45 is $(15 + 12)/50 = 0.54$.
- e The proportion of children aged less than 50 months is $(12 + 15 + 12 + 8)/50 = 0.94$.

1.28 a This is similar to previous exercises. The pie chart is shown below.



- b** 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.



- c** The larger the airline, the more difficult it may be to serve passengers without any complaints. Other variable might be “size of airline” or “number of passengers served” or “airfare.”

- 1.29 a** Use the ones digit as the stem, and the portion to the right of the ones digit as the leaf, dividing each stem into two parts.

```

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

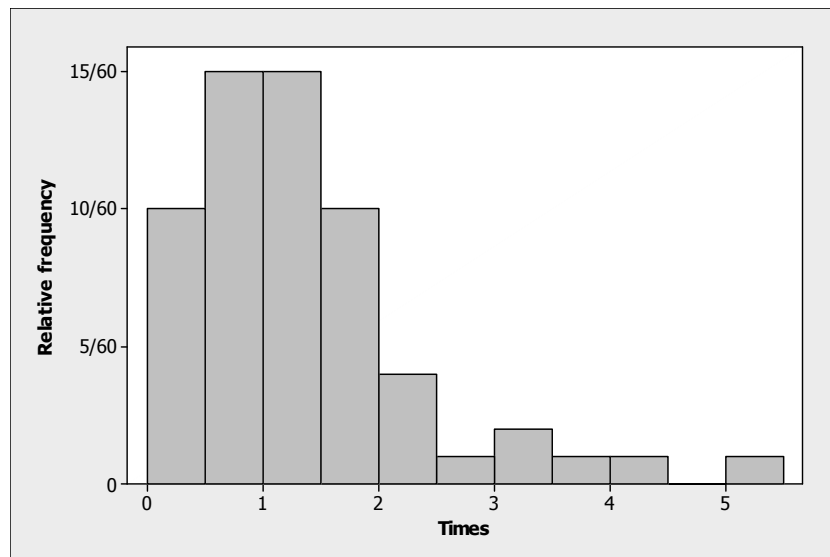
leaf digit = 0.1
1 2 represents 1.2

- b** Looking at the original data, we see that 25 customers waited 1 minute or less. Therefore, the fraction of service times less than or equal to 1 is $25/60 = 0.4167$.
- c** The smallest measurement is 0 2, which is translated as 0.2.

- 1.30 a** The data ranges from 0.2 to 5.2, or 5.0 units. Since the number of class intervals should be between 5 and 20, we chose to use 11 class intervals, with each class interval having length 0.50 ($5.0/11 = 0.45$, which rounded to the nearest convenient fraction is 0.50). We must now select interval boundaries such that no measurement can fall on a boundary point. The subintervals 0.1 to < 0.6 , $.6$ to < 1.1 , and so on, are convenient, and a tally is constructed.

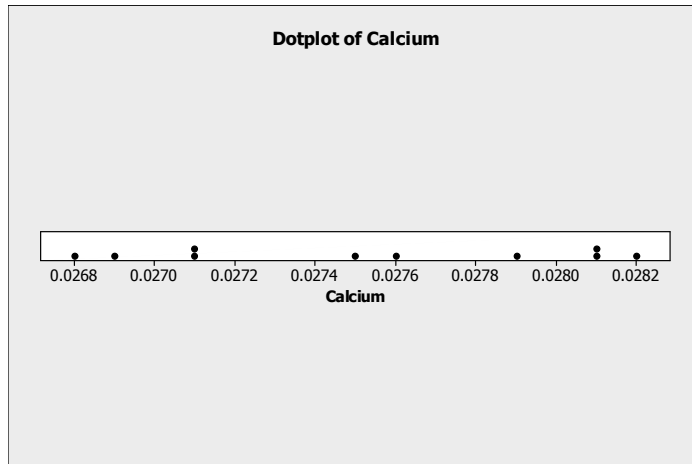
Class i	Class Boundaries	Tally	f_i	Relative Frequency, f_i/n
1	0.1 to < 0.6	11111 11111	10	0.167
2	0.6 to < 1.1	11111 11111 11111	15	0.250
3	1.1 to < 1.6	11111 11111 11111	15	0.250
4	1.6 to < 2.1	11111 11111	10	0.167
5	2.1 to < 2.6	1111	4	0.067
6	2.6 to < 3.1	1	1	0.017
7	3.1 to < 3.6	11	2	0.033
8	3.6 to < 4.1	1	1	0.017
9	4.1 to < 4.6	1	1	0.017
10	4.6 to < 5.1		0	0.000
11	5.1 to < 5.6	1	1	0.017

The relative frequency histogram is shown below.



- a** The distribution is skewed to the right, with several unusually large observations.
b 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.
c The two graphs convey the same information. The stem and leaf plot allows us to actually recreate the actual data set, while the histogram does not.

- 1.31 a–b The dotplot and the stem and leaf plot are drawn using *MINITAB*.



Stem and Leaf Plot: Calcium
 Stem and leaf of Calcium N = 10
 Leaf Unit = 0.00010

```

2  26  89
4  27  11
4  27
5  27  5
5  27  6
4  27  9
3  28  11
1  28  2
    
```

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

- 1.32 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. *MINITAB* divides each stem into two parts to create a better descriptive picture. Notice that the distribution is roughly mound-shaped.

Stem and Leaf Plot: Age
 Stem and leaf of Age N = 22
 Leaf Unit = 1.0

```

1  3  9
1  4
8  4  5666778
11 5  124
11 5  579
8  6  01
6  6  5669
2  7  04
    
```

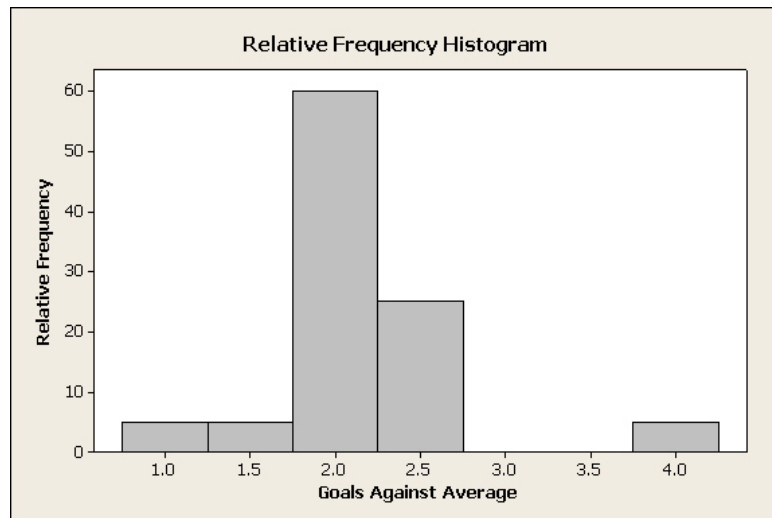
- c The five youngest prime ministers—Clark, Mulroney, Harper, Campbell, and Meighen—are all Conservative. Although this does not tell us much, we can guess that the trend of transferring the leadership to younger generations in Conservative party is faster.

- 1.33 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 below.

```

Stem and Leaf Plot: Cells
Stem and leaf of Cells  N  = 15
Leaf Unit = 0.010
 1  49  0
 2  50  0
 3  51  0
(5) 52  00000
 7  53  000
 4  54  000
 1  55  0
    
```

- b The data set is relatively mound-shaped, centred 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.
- 1.34 a Histograms will vary from student to student. A typical histogram, generated by *MINITAB*, is shown below.



- b Since 4 out of 20 players have averages above 2.4, the chance is 4 out of 20 or $4/20 = 0.2$ or 20%.
- 1.35 a The stem and leaf plot is shown below.

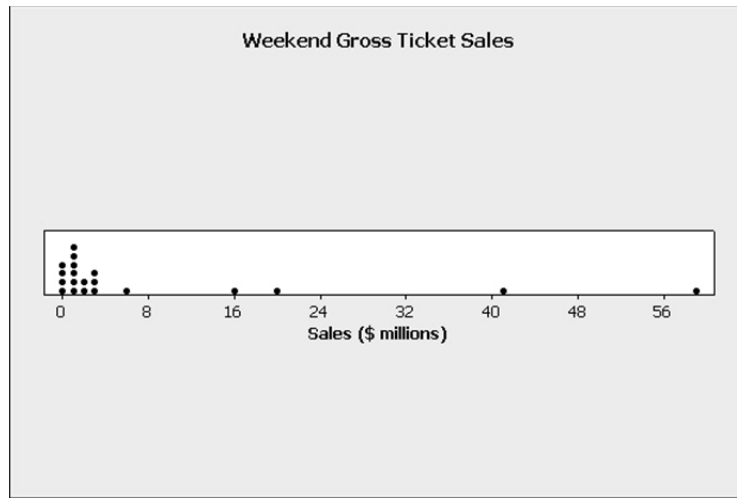
```

Stem-and-Leaf Plot: Weekend Gross Ticket Sales
Stem and leaf of Weekend  N  = 20
Leaf Unit = 0.10
 4  0  3444
 7  0  556
10  1  024
10  1  69
 8  2
 8  2  8
 7  3  11
 5  3
 5  4
 5  4
 5  5
 5  5
 5  6  2

HI 155, 201, 405, 593
    
```


The distribution is skewed to the right, with four outliers, marked by “HI” in the stem and leaf plot.

- b** The dotplot of the weekend gross ticket sales is given below.



The stem and leaf plot is more informative as compared to the dotplot. We can identify the numerical values of the potential outliers and also original data can be reconstructed from the stem and leaf plot.

- 1.36 a** The variable being measured is a discrete variable: the number of contaminated waste sites in each province or territory in Canada.

b Stem and Leaf Plot: Number of Contaminated Waste Sites

Stem and leaf of Sites N = 13
Leaf Unit = 1.0

```

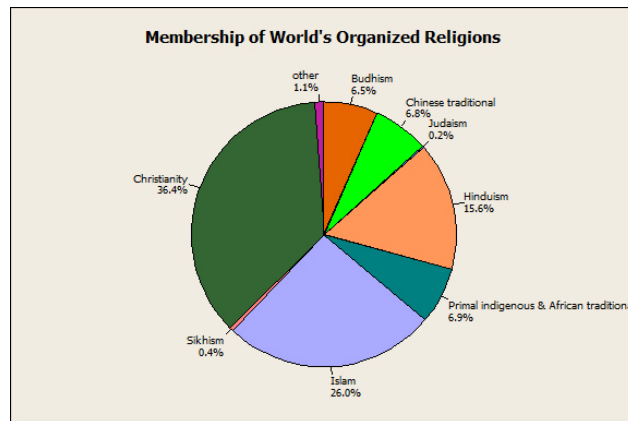
1  1  7
2  2  1
3  3  4
5  4  44
(3) 5  567
5  6
5  7  88
    
```

HI 205, 235, 300

The distribution is skewed to the right, with three unusually high numbers of contaminated waste sites (Ontario, British Columbia, and Quebec).

- c** In comparison to other provinces, the three largest provinces—Quebec, Ontario, and British Columbia—have very large numbers of contaminated sites. However, despite their large land sizes, Nunavut and Northwest Territories have relatively fewer contaminated sites. The pattern is not very clear. Some other variables such as population size, number of industries, etc. may help to explain the relationship better.

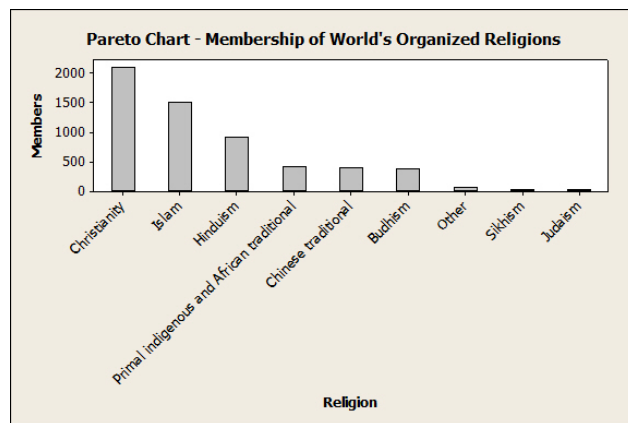
1.37 a The pie chart is given below.



b The bar chart is given below.



c The Pareto chart is given below.



Comparing graphs from parts a–c, the Pareto chart seems more effective since it is very easy to compare the relative membership of the organized religions.

- 1.38**
- 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 (McDonald’s, Burger King, or Subway) is measured.
 - d** “Mercury concentration” is a *quantitative variable* since a numerical quantity is measured.
- 1.39** 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 distribution of non-secured loan sizes might be skewed (a few extremely large loans are possible).
 - b** The distribution of secured loan sizes is not likely to contain unusually large or small values.
 - c** Not likely to be skewed.
 - d** Not likely to be skewed.
 - e** 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.
 - f** If an animal has one tick, it 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.
- 1.40**
- a** The number of homicides in Vancouver during a one-month period is a discrete random variable since it can take only the values 0, 1, 2, ...
 - b** 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.
 - c** The number of typing errors is a discrete random variable, since it can take only the values 0, 1, 2, ...
 - d** Again, this is a discrete random variable since it can take only the values 0, 1, 2, 3, 4.
 - e** The time required to finish an examination is a continuous random variable, similar to the random variable described in part **b**.
- 1.41**
- a** Weight is continuous, taking any positive real value.
 - b** Body temperature is continuous, taking any real value.
 - c** Number of people is discrete, taking the values 0, 1, 2,
 - d** Number of properties is discrete.
 - e** Number of claims is discrete.
- 1.42**
- a** Number of people 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.
- 1.43** Stem and leaf plots 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.

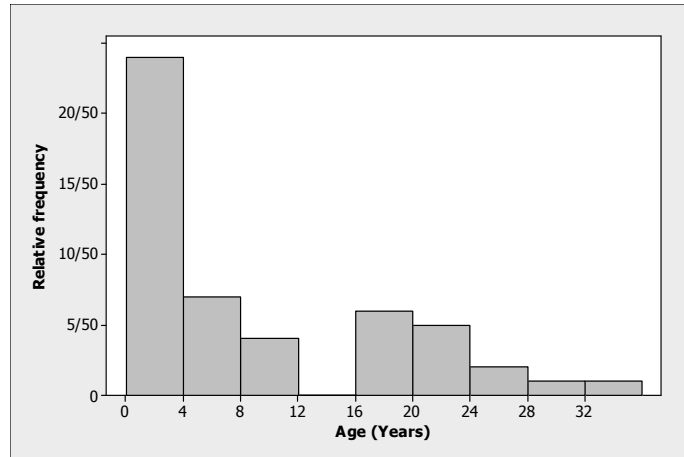
```

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
    
```

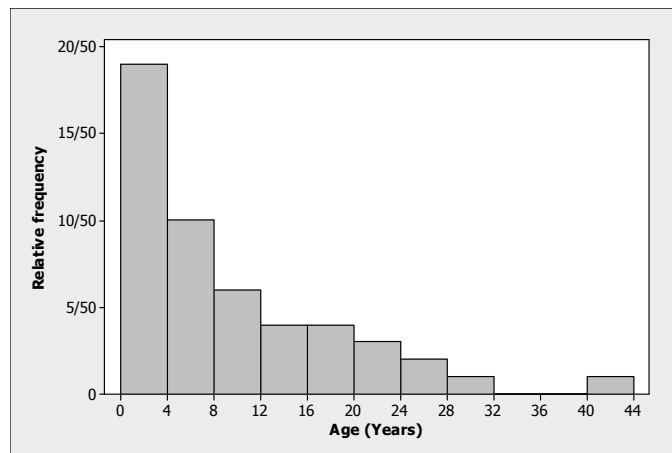
The display is fairly mound-shaped, with a large peak in the middle.

- 1.44**
- a** Side-by-side bar charts were used to compare EQAO scores pre- and post-JUMP training.
 - b** Answers will vary.

- 1.45 a–b Answers will vary from student to student. Students should notice that the distribution is skewed to the right with a few pennies being unusually old. A typical histogram is shown below.



- 1.46 a Answers will vary from student to student. A typical histogram is shown below. It looks very similar to the histogram from Exercise 1.45.



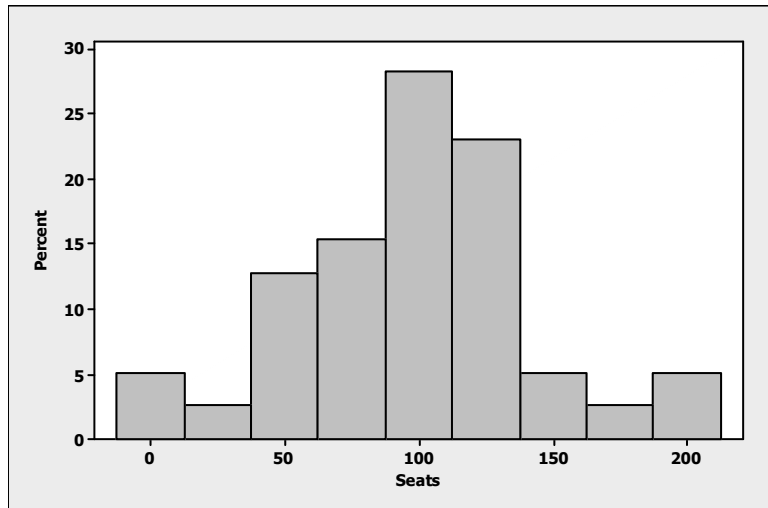
- b The stem and leaf plot is drawn using *MINITAB*. There is one outlier, $x = 41$.

```

Stem and Leaf Plot: Age (Years)
Stem and leaf of Age (Years)  N = 50
Leaf Unit = 1.0
 9  0  000000011
19  0  2223333333
(7) 0  4444555
24  0  777
21  0  88999
16  1  0
15  1  2
14  1  444
11  1  677
 8  1  9
 7  2  01
 5  2  3
 4  2  45
 2  2
 2  2  8
    
```

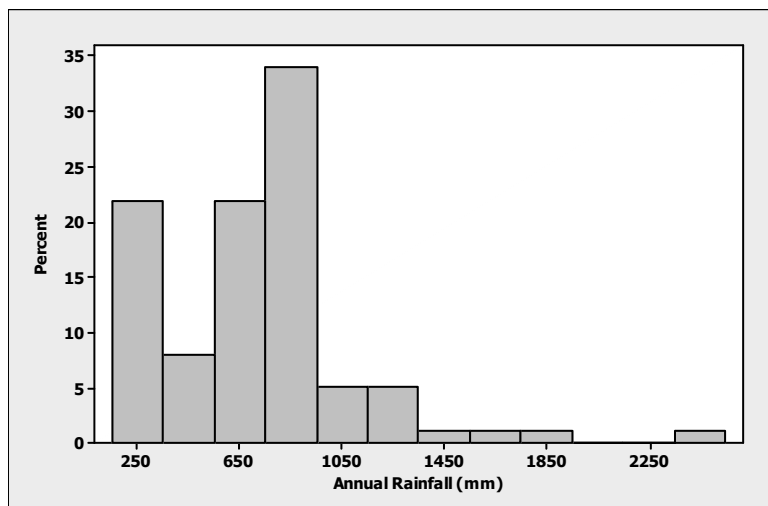
HI 41

- 1.47 Answers will vary from student to student. The students should notice that the distribution is mound-shaped with the number of seats won in most of the elections being between 75 and 125. There were an unusually high number of seats won in two of the elections (208 and 211 seats).



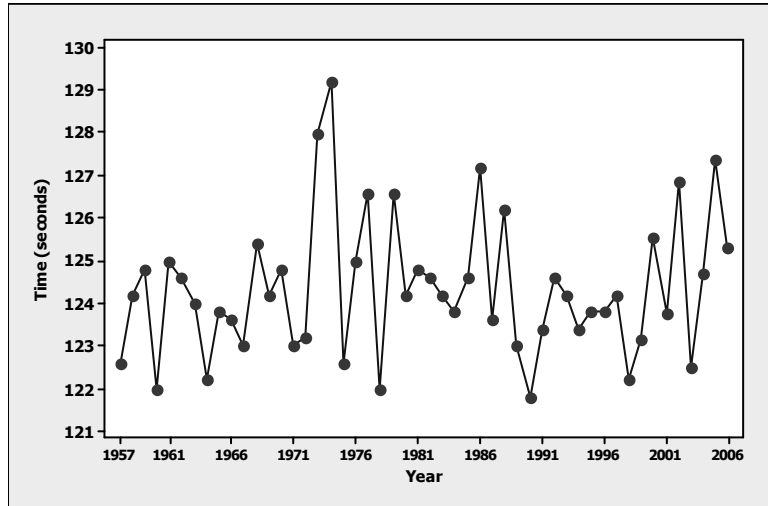
- 1.48 a Answers will vary from student to student. The relative frequency histogram below was constructed using classes of width 200 starting at $x = 150$. The value $x = 2468.53$ is not shown in the table, but appears on the graph.

Class i	Class Boundaries	f_i	Relative Frequency, f_i/n
1	150 to < 350	22	22/100
2	350 to < 550	8	8/100
3	550 to < 750	22	22/100
4	750 to < 950	34	34/100
5	950 to < 1150	5	5/100
6	1150 to < 1350	5	5/100
7	1350 to < 1550	1	1/100
8	1550 to < 1750	1	1/100
9	1750 to < 1950	1	1/100

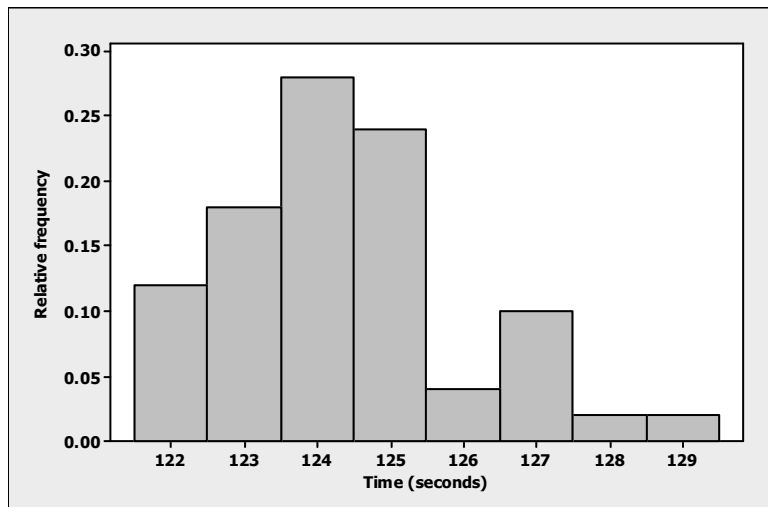


- b** Since Prince Rupert is located west of the Coastal mountain, it is not unusual that the average rainfall would be very high.
- c** The value $x = 1154.66$ (Vancouver) does not lie far from the centre of the distribution. It would not be considered unusually rainy.

1.49 a The line chart is shown below. The year does not appear to have an effect on his winning time.

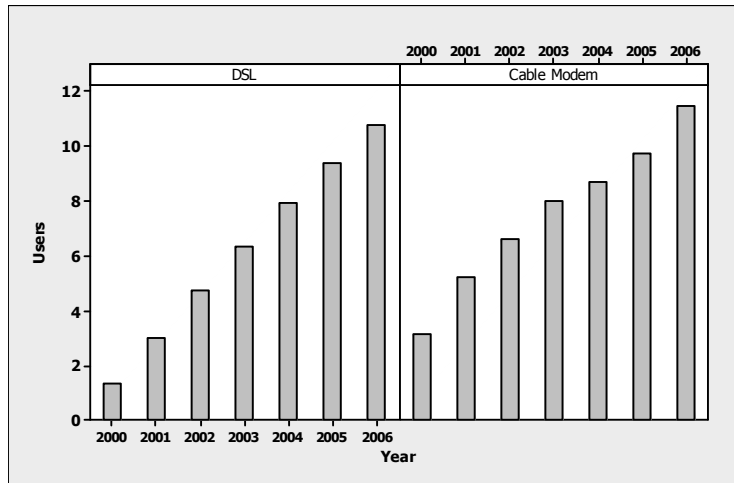
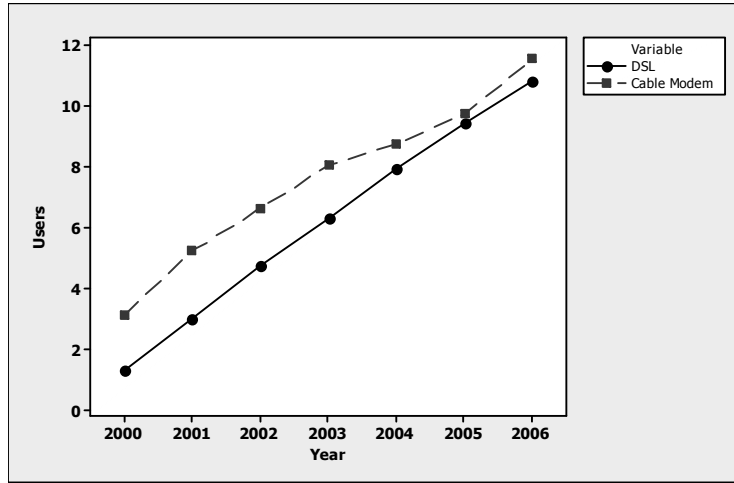


- b** Since the year of the race is not important in describing the data set, the distribution can be described using a relative frequency histogram. The distribution shown below is roughly mound-shaped with a few unusually slow ($x = 129.2$, $x = 128.0$) race times.

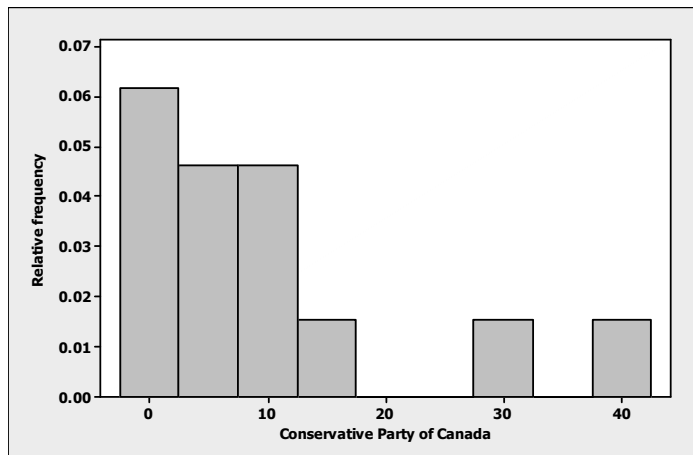


1.50 a The quantitative variable “number of high-speed Internet users” is measured over time for two different types of broadband Internet connections. Some sort of comparative bar charts or a line chart should be used.

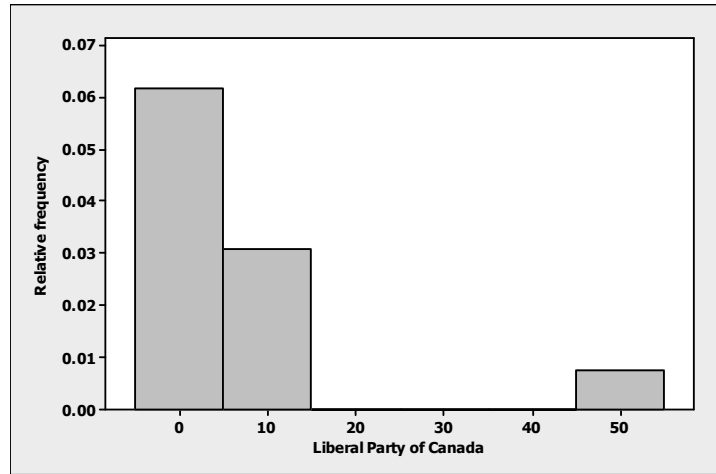
- b** The number of both DSL and cable modem Internet users can be expected to rise steadily in the future.



- 1.51 a** Most of the provinces/territories have very few Conservative seats (9 out of 13 have 10 or fewer seats won); 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 figure below. The distribution is indeed skewed to the right, with one outlier: Ontario ($x = 40$).



- 1.52** **a** Like Conservatives, most of the provinces/territories have very few Liberal seats (11 out of 13 have 9 or fewer seats), 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 figure below. The distribution is indeed skewed to the right, with one outlier: Ontario ($x = 54$).

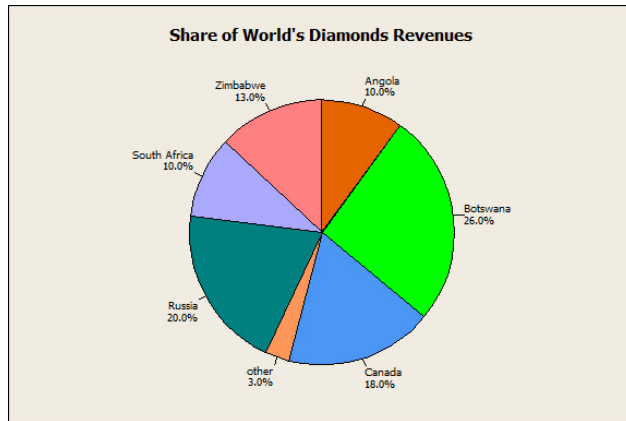


1.53 **Stem and Leaf Plot: Conservative Party of Canada, Liberal Party of Canada**

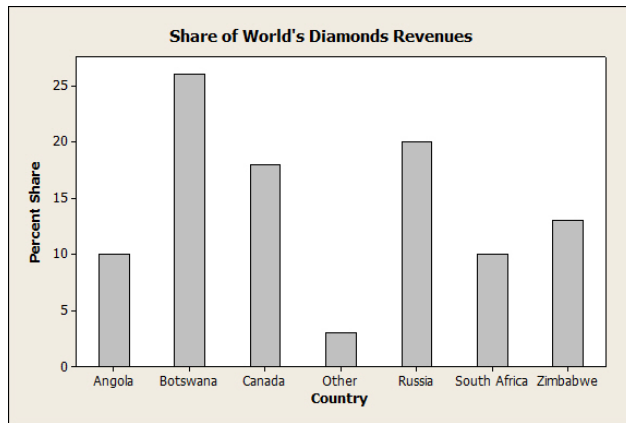
Stem and leaf of Conservative Party of Canada N = 13 Leaf Unit = 1.0	Stem and leaf of Liberal Party of Canada N = 13 Leaf Unit = 1.0
(7) 0 0000333	(8) 0 00112344
6 0 8	5 0 669
5 1 02	2 1 3
3 1 7	1 1
2 2	1 2
2 2 8	1 2
1 3	1 3
1 3	1 3
1 4 0	1 4
HI 40	1 4 4
	1 5 4
	HI 54

- a–b** As in the case of relative frequency histogram, both the distributions are skewed to the right with one outlier on each (Ontario). When the stem and leaf plots are turned 90°, the shapes are very similar to the histograms.
c Since the total of 308 House of Commons seats are distributed very disproportionately among different provinces and territories, with only four provinces having more than 15 seats, these graphs will be skewed right.

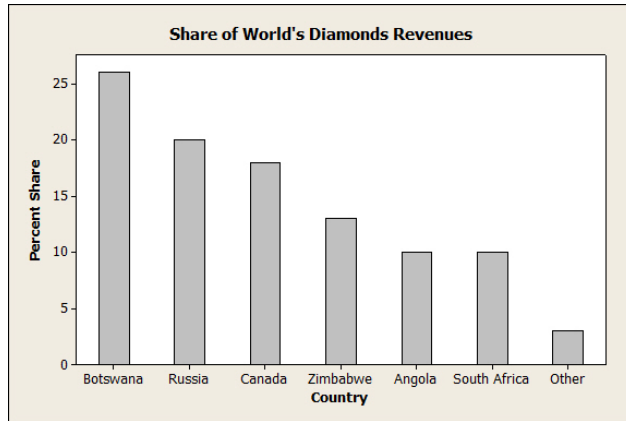
1.54 a The pie chart is given below.



b The bar chart is given below.

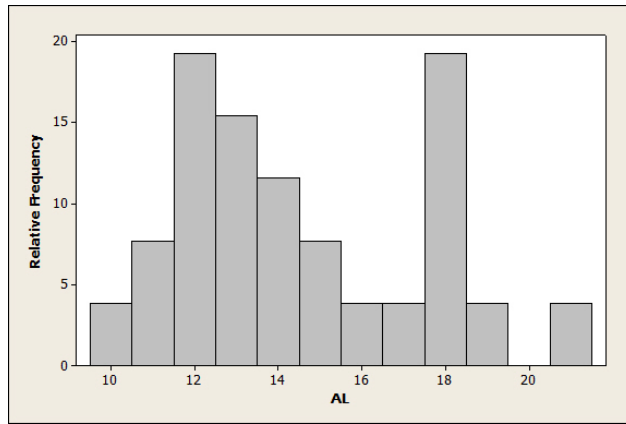


c The Pareto chart is given below.

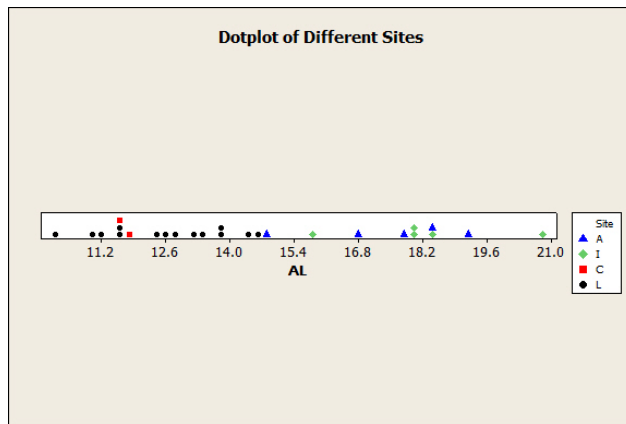


d The pie chart and the Pareto chart are both more effective than the bar chart. Note: The charts constructed in parts a–c may vary from student to student since Zimbabwe's share was given as a range (8–13%) and this share is used as 13% in constructing the above graphs.

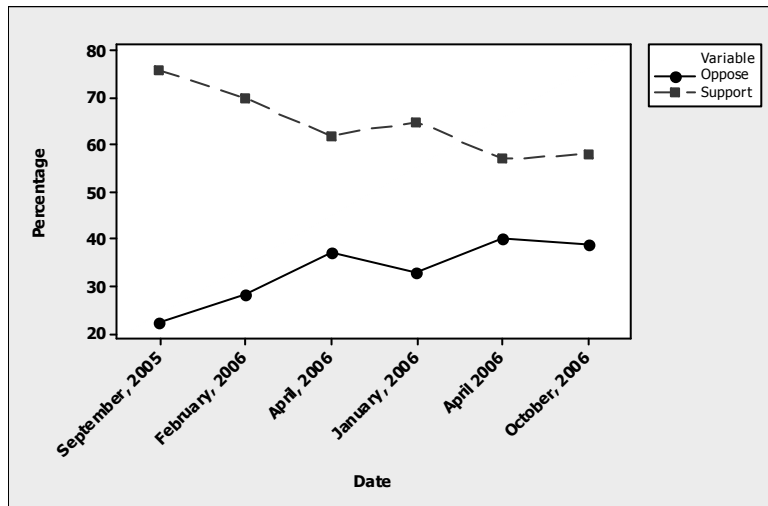
- 1.55 a–b Answers will vary. A typical relative frequency histogram is shown below. The gaps and bimodal nature of the histogram probably is due to the fact that the samples were collected at different locations.



- c The dotplot is shown below. The locations seem to be responsible for the unusual gaps and peaks in the relative frequency histogram given in part b.



- 1.56 a–c Answers will vary. The line chart should look similar to the one shown below.



d The horizontal axis on the EKOS chart is not an actual timeline, so that the time frame in which these changes occur may be distorted.

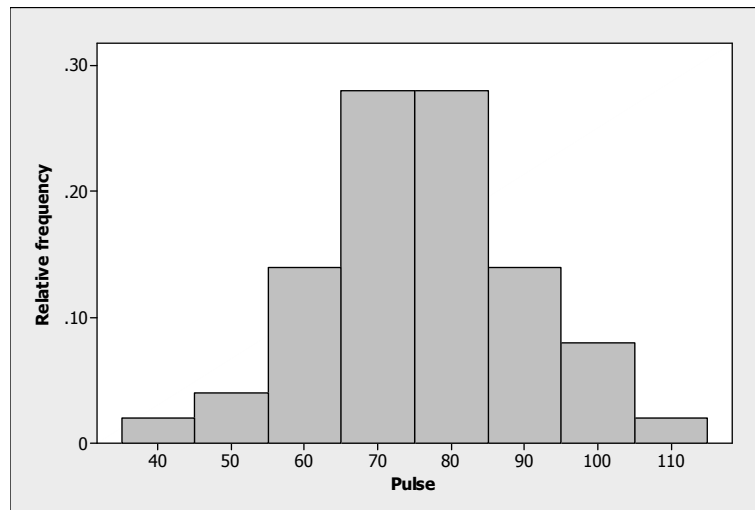
1.57 a The measurements are obtained by counting the number of beats for 30 seconds, and then multiplying by 2. Thus, the measurements should all be even numbers.

b The stem and leaf plot is shown below.

Stem and Leaf Plot: Pulse
 Stem and leaf of Pulse N = 50
 Leaf Unit = 1.0

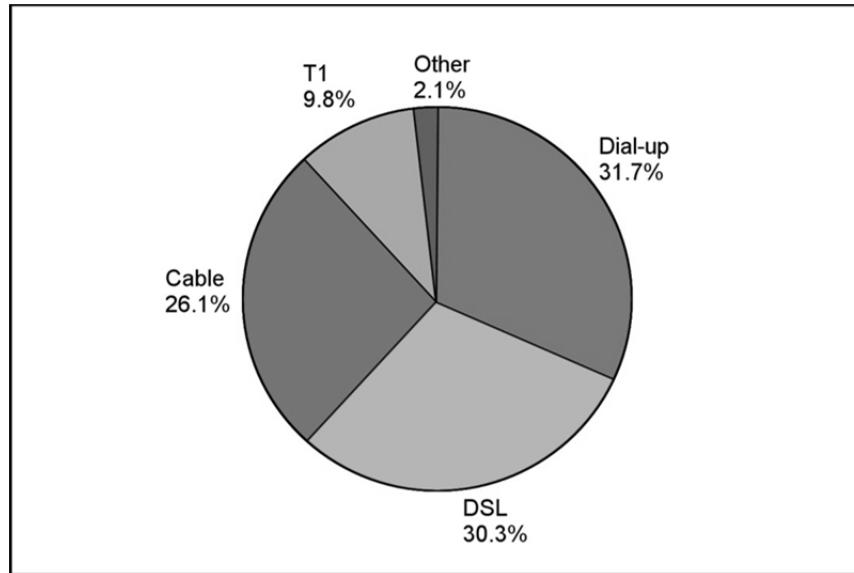
1	4	2
1	4	
3	5	24
6	5	688
10	6	0022
15	6	66668
24	7	000222224
25	7	8
25	8	0022444444444
12	8	68888
7	9	00
5	9	66
3	10	04
1	10	
1	11	0

c Answers will vary. A typical histogram, generated by *MINITAB*, is shown below.

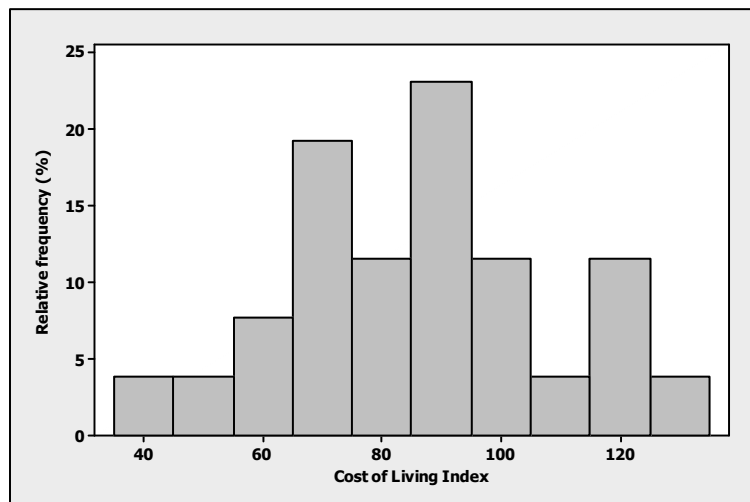


d The distribution of pulse rates is mound-shaped and relatively symmetric around a central location of 75 beats per minute. There are no outliers.

- 1.58 a–b This is similar to previous exercises. The percentages add up to 100%, and the pie chart is shown below.



- 1.59 Answers will vary. The distribution is roughly mound-shaped. A typical histogram is shown below.



- b Answers will vary. The stem and leaf plot generated by *MINITAB* uses the tens place as the stem and the ones place as the leaf.

Stem and Leaf Plot: Index

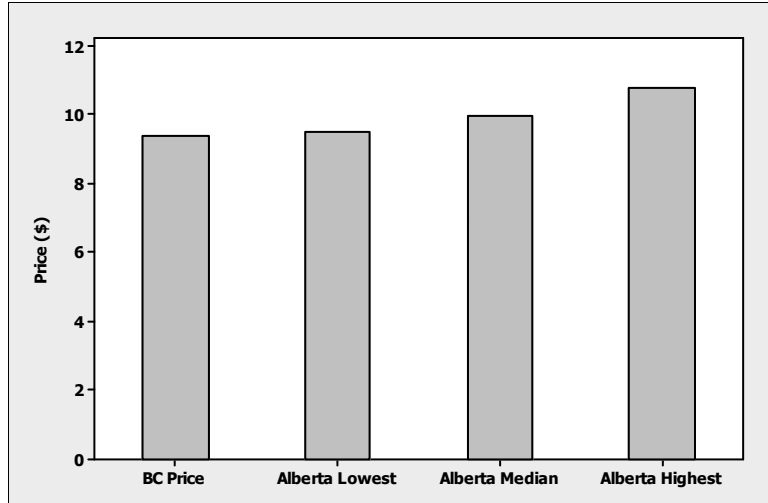
Stem and leaf of Index N = 26
Leaf Unit = 1.0

```

2  4  28
2  5
6  6  2469
9  7  134
(6) 8  344568
11 9  014
8  10 0119
4  11 679
1  12
1  13 0
    
```

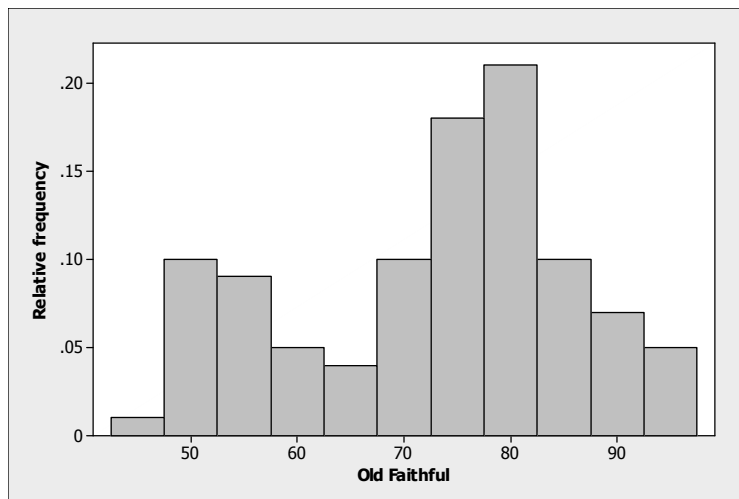
- c Since the data appears in Mercer's site, this global investment consulting agency may have chosen the cities of its business priorities.

- 1.60 a The heights of the six bars do increase as the prices increase, but not in the correct proportion to the actual prices. The price scale starts at \$8.50. So, although the height of "Alberta Highest = \$10.79" looks more than double of "Alberta Lowest = \$9.50," the actual price is only about 14% higher.
- b The bar graph that accurately portrays the retail price comparison is shown below. There are not much differences among four prices.



- 1.61 Answers will vary. Students should notice that the first distribution (12:00–1:30) is mound-shaped and the distribution (4:30–6:00) is slightly skewed.

- 1.62 Answers will vary. A typical relative frequency histogram is shown below. There is an unusual bimodal feature.



1.63 a–b The *MINITAB* stem and leaf plot is shown below. The distribution is slightly skewed to the left.

Stem and Leaf Plot: Total Tax Component
 Stem and leaf of Total Tax Component N = 15
 Leaf Unit = 1.0

```

1  2  2
2  2  4
4  2  66
4  2
7  3  000
(1) 3  3
7  3
7  3  667
4  3  8899
    
```

c There are no unusually high or low gasoline taxes in the data.

1.64 a The stem and leaf plot is shown below.

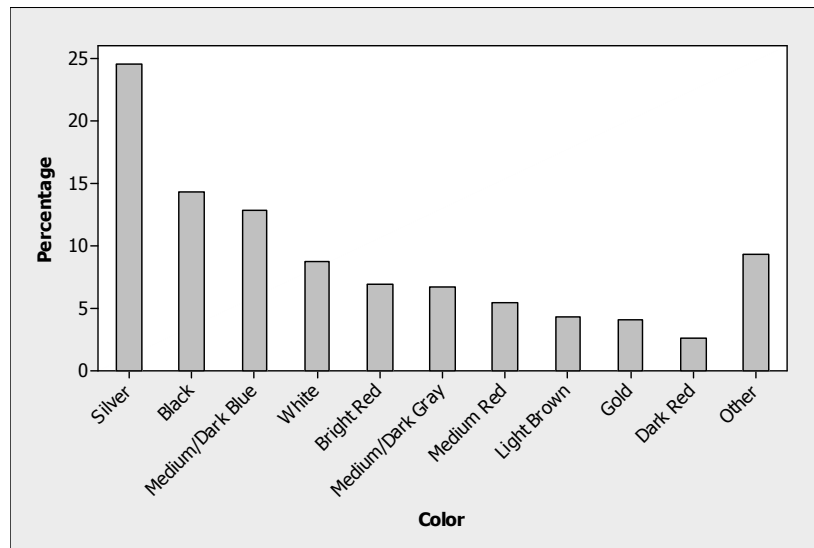
Stem and Leaf Plot: Megawatts
 Stem and leaf of Megawatts N = 20
 Leaf Unit = 1000

```

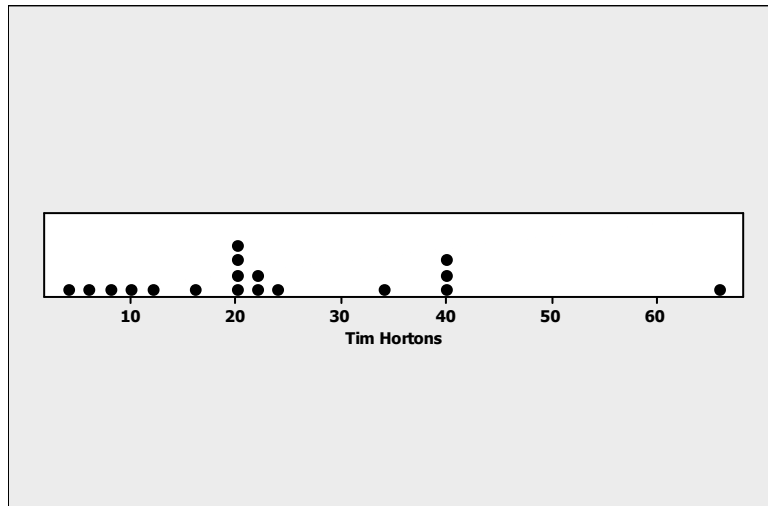
10  0  2222233333
10  0  444
7   0  666
4   0  8
3   1  0
2   1  2
1   1
1   1
1   1  8
    
```

b The distribution of planned rated capacities for the world's 20 largest plants is skewed to the right.

1.65 The data should be displayed with either a bar chart or a pie chart. Because of the large number of categories, the bar chart is probably more effective.

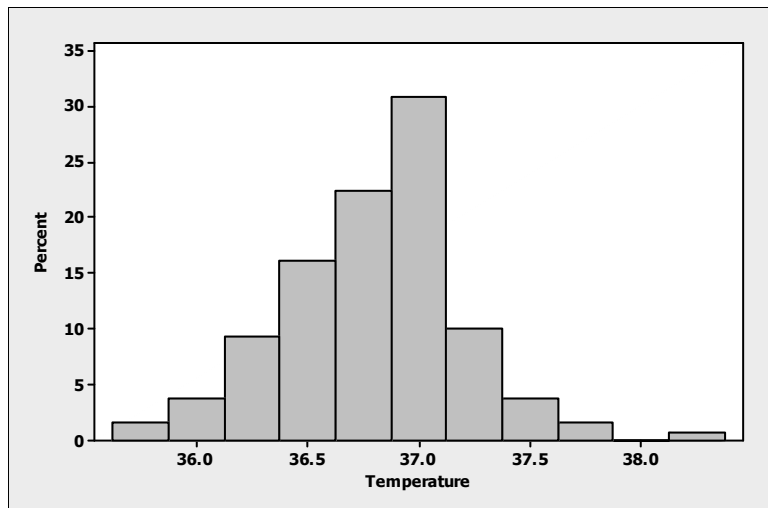


1.66 a–b The dotplot is shown below. The distribution is slightly skewed to the right.



c The number of Tim Hortons should be related to the number of customers. The cities with larger populations may be more likely to have a larger number of Tim Hortons shops.

1.67

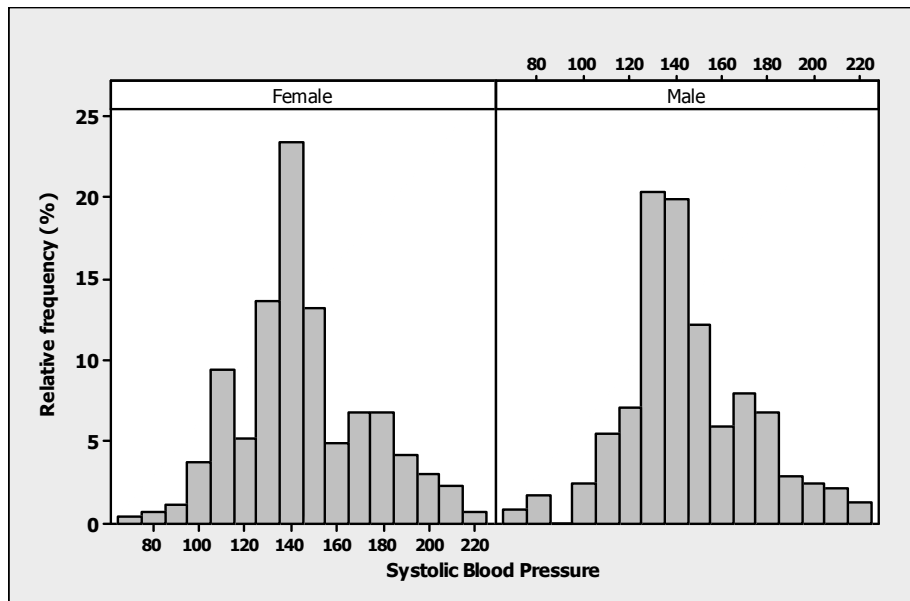


a–b The distribution is approximately mound-shaped, with one unusual measurement, in the class with midpoint at 38.5°C ($x = 38.22$). Perhaps the person whose temperature was 38.22°C had some sort of illness.

c The value 37°C is slightly to the right of centre.

Case Study: How Is Your Blood Pressure?

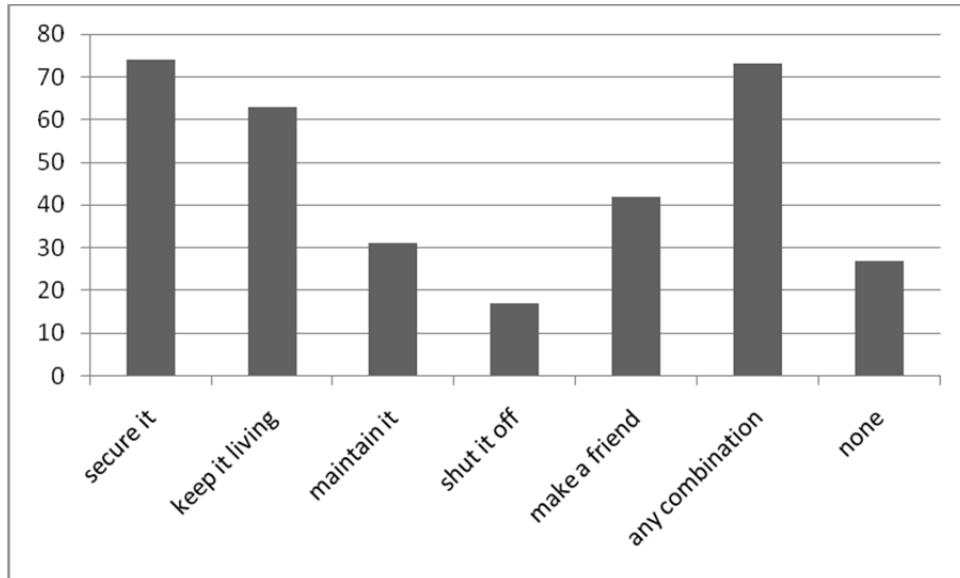
- 1 The following variables have been measured on the participants in this study: sex (qualitative); age in years (quantitative discrete); diastolic blood pressure (quantitative continuous, but measured to an integer value); and systolic blood pressure (quantitative continuous, but measured to an integer 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, depending on the choice of class boundaries or the software package which is used. The histograms should look fairly mound-shaped. A typical side-by-side histogram generated by *MINITAB* is shown below.



- 4 Answers will vary.
- 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.

Project 1-A: Five Tips for Keeping Your Home Safe This Summer

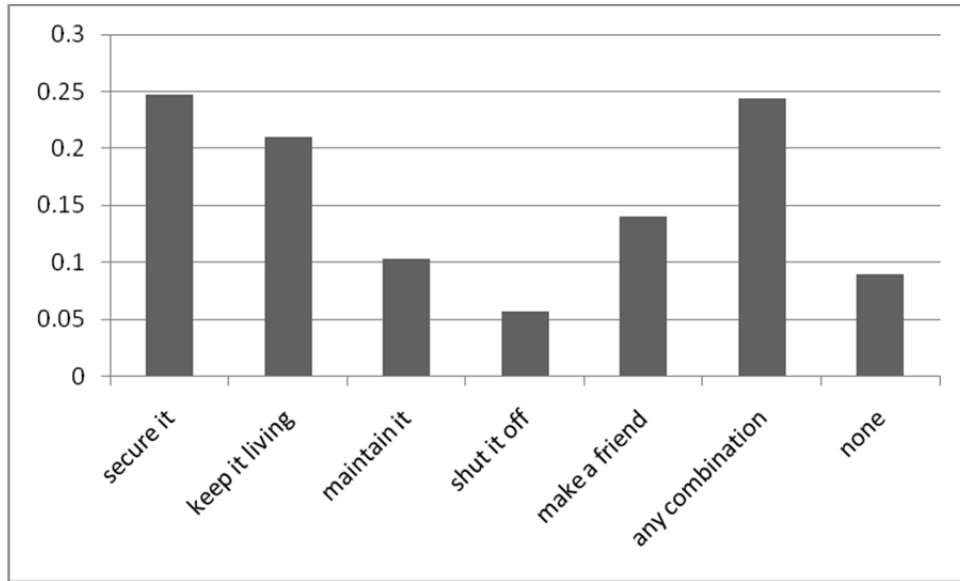
- a The population is all households in that particular subdivision in the city of North York. The sample is the 300 households in that subdivision that were surveyed.
- b The collected data is based on population, in the sense that the sample taken is randomly selected from the population and meant to be representative of the population.
- c The experimental units are the households.
- d The variable being measured is the type of tip employed.
- e The variable is qualitative.
- f Neither: The variable is qualitative. The counts, however, are discrete.
- g The bar chart is shown below. The chart graphically portrays the counts for each type of tip.



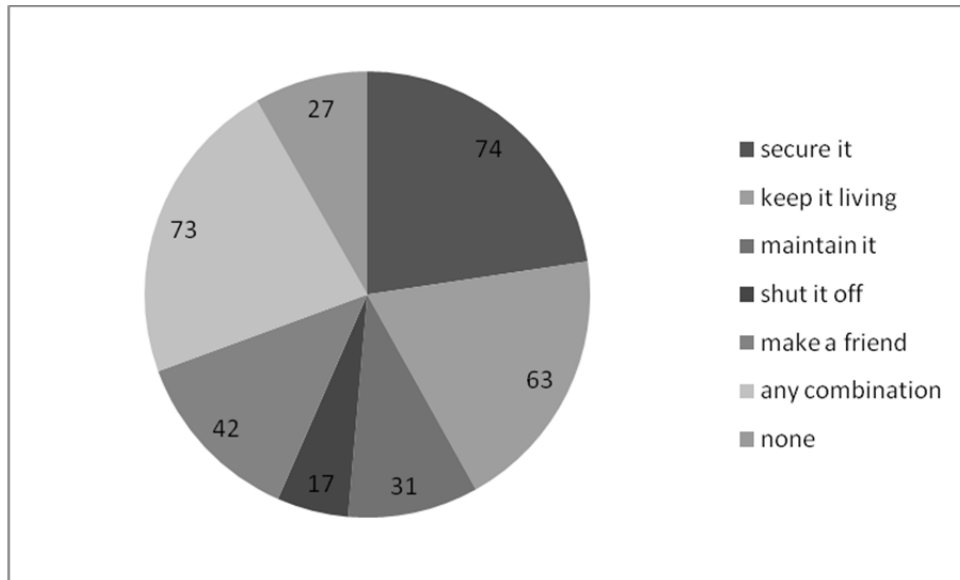
- h The relative frequencies are obtained by dividing the counts by the total sample size (300):

<u>Type of Tips</u>	<u>Number of Households</u>	<u>Relative Frequency</u>
secure it	74	0.246667
keep it living	63	0.210000
maintain it	31	0.103333
shut it off	17	0.056667
make a friend	42	0.140000
any combination	73	0.243333
none	27	0.090000

i A relative frequency bar chart is shown below.



j A pie chart (by count) is shown below.

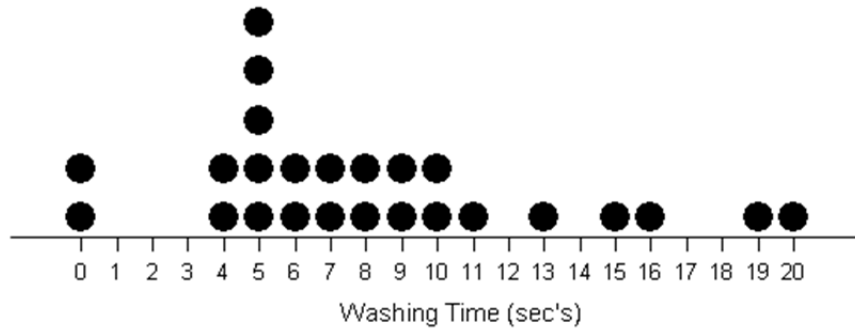


k The proportion of respondents who chose either “Secure it” or “Make a friend” is the sum of their respective relative frequencies: $0.246667 + 0.140000 = 0.386667$ or 38.67%.

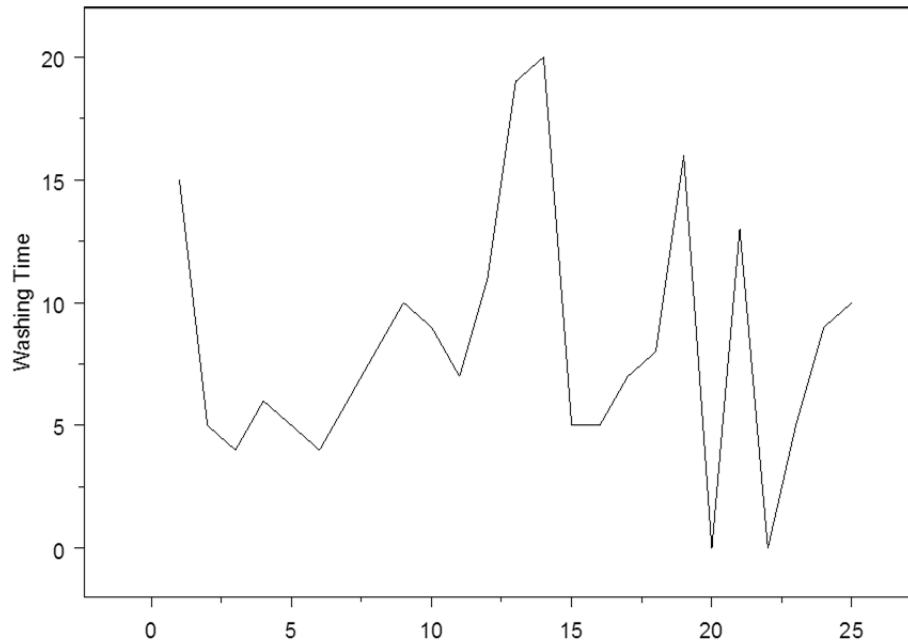
l Answers will vary.

Project 1-B: Handwashing Saves Lives: It's in Your Hands

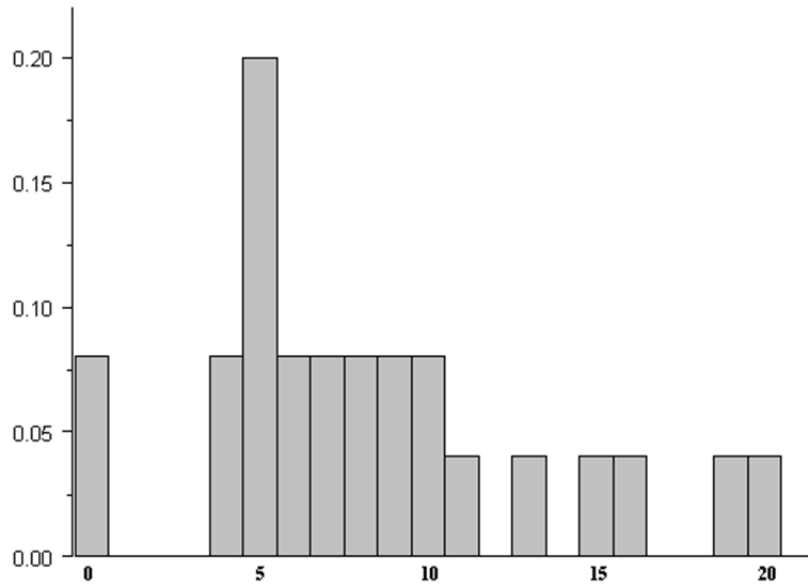
- a The experimental units are the students.
- b The variable is the time (in seconds) students take to wash their hands.
- c The variable is quantitative.
- d The variable is discrete because it is only measured to the nearest second.
- e A dotplot is shown below. The value that occurs most often is 5. The range of the data is from 0 to 20. The data is distributed more to the lower values. There are a few gaps in the data as well.



- f The distribution of data is skewed to the lower values, as most students took 10 seconds or less to wash their hands.
- g The line chart was constructed by using the students' numerical order (student 1, 2, ... 25) as the x -variable and the amount of time they washed their hands as the y -variable.



h The frequency histogram is shown below.



i We count 17 students that washed their hands for less than 10 seconds, or $17/25 = 68\%$.

j We count 21 students that washed their hands for at least 5 seconds, or $21/25 = 84\%$.

k No, we cannot comfortably state that most students wash their hands for 5 seconds or less. This is because only 9 students out of 25 (36%) wash their hands for 5 seconds or less. This leaves the majority of students washing their hands for more than 5 seconds.

l The stem and leaf plot is below. Note that the colon represents the decimal point in this case.

```

0 : 00
1 :
2 :
3 :
4 : 00
5 : 00000
6 : 00
7 : 00
8 : 00
9 : 00
10 : 00
11 : 0
12 :
13 : 0
14 :
15 : 0
16 : 0
17 :
18 :
19 : 0
20 : 0
    
```

- m** The data is skewed right, as the tail of the distribution is on the right (i.e., the higher values).
- n** Points 0, 19, and 20 appear to be potential outliers.
- o** Answers will vary.