SOLUTIONS MANUAL

JOHNSON / KUBY
STUDENT EDITION
STAT
WHAT'S INSIDE:
Student Tested, Faculty Approved
• PLUS •
ACCESS ONLINE INTERACTIVE PROBLEMS, FLASH CARDS, AND MORE WITH EVERY NEW BOOK!
A new approach to learning Introduction to Statistics built with you in mind
ONLINE RESOURCES AVAILABLE ONLY WITH THE PURCHASE OF A NEW BOOK

CHAPTER 2 DESCRIPTIVE ANALYSIS AND PRESENTATION OF SINGLE-VARIABLE DATA

Chapter Preview

Chapter 2 deals with the presentation of data that were obtained through the various sampling techniques discussed in Chapter 1. The four major areas for presentation and summary of the data are:

1. graphical displays,
2. measures of central tendency,
3. measures of variation, and
4. measures of position.

Show the students how the chapters are related as you work through the course. In Chapter 1, they learned how to randomly collect a set of data. In Chapter 2, they will learn the various presentations and measures needed to summarize data in a more manageable form.
Examples of bar graphs, circle graphs, pareto diagrams, dot plots and stem-and-leaf diagrams and their corresponding requirements need to be presented to the students.
Students have difficulty understanding the concept of partitioning a circle into percentages. When either a higher level or more accuracy is needed, present partitioning with respect to the 360 degrees in a circle.
See how many students constructed vertical and how many constructed horizontal bar graphs. Have them discuss their reasons for choosing.
How to pick increments for a dot plot should be presented. The entire concept of a number line may have to be reviewed.
Present the concept of sorting and summarizing data by using a stem-and-leaf display. Use a simple set of data values, such as test grades (2 digits), to show how to split the data values. Later, data sets with more complicated values can be shown. Where to break up the data values depends on the desired precision. Emphasize that knowledge of your data is invaluable and critical in these decisions.

OBJECTIVE 2.1 PROBLEMS

MINITAB - Statistical software
Data is entered by use of a spreadsheet divided into columns and rows. Data for each particular problem is entered into its own column. Each column represents a different set of data. Be sure to name the columns in the space provided above the first row, so that you know where each data set is located. (C1 = Column 1)

EXCEL – Spreadsheet software
Data is entered by use of a spreadsheet divided into columns and rows. Data for each particular problem is entered into its own column. Each column represents a different set of data. If needed, use the first row of a column for a title. (A1 = 1st cell of column A)

TI-83/84 Plus – Graphing calculator
Data is entered into columns called lists. Data for each particular problem is entered into its own list. Each list represents a different set of data. If needed, use the space provided above the first row for a title. Lists are found under STAT > 1:Edit.
(L1 = List 1)

Partitioning the circle:

1. Divide all quantities by the total sample size and turn them into percents.
2. 1 circle = 100%
   1/2 circle = 50%
   1/4 circle = 25%
   1/8 circle = 12.5%
3. Adjust other values accordingly.
4. Be sure that percents add up to 100 (or close to 100, depending on rounding).

Computer and calculator commands to construct a Pie Chart can be found on your Chapter Two Technology Card.

The TI-83 program 'CIRCLE' and others can be downloaded from the textbook companion website at 4ltrpress.cengage.com/stat. Select 'TI-83/84 Programs' from the book resource menu. Right-click on the zip file to save it to your computer and unzip the file using a zip utility, such as Winzip. Then download the programs to your calculator using TI-Graph Link Software.

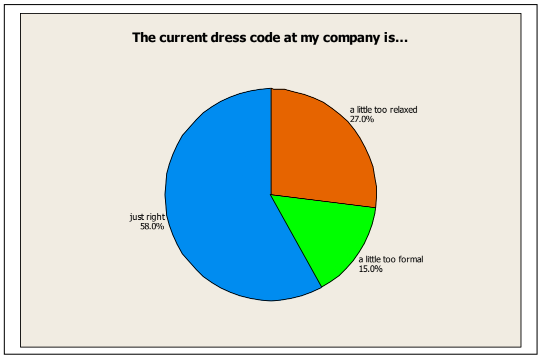
MINITAB commands to construct a bar graph can be found in the GRAPH > BAR CHART pull-down menu. With the categories in C1 and the corresp. freq. in C2, select 'Values from a table'; '1 col of values: Simple'.

EXCEL commands to construct a bar graph can be found under Chart Wizard. Input the categories into column A and the corresponding frequencies into column B.

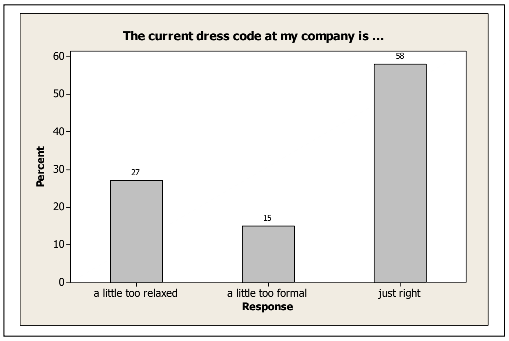
TI-83/84 commands to construct a bar graph can be found using the STATPLOT command. Input the categories as numbers into list 1 and the corresponding frequencies into list 2. Adjust the *x*-scale in the WINDOW to 0.5 to allow for gaps between bars.

NOTE: Bar graphs may be vertical (as shown below) or horizontal (as shown in the chapter opener). Information typically on the axes may also be printed inside the bars themselves. There is much flexibility in constructing a bar graph. Remember to leave a space between the bars. Be sure to label both axes and give a title to the graph.

2.1 a.



b.



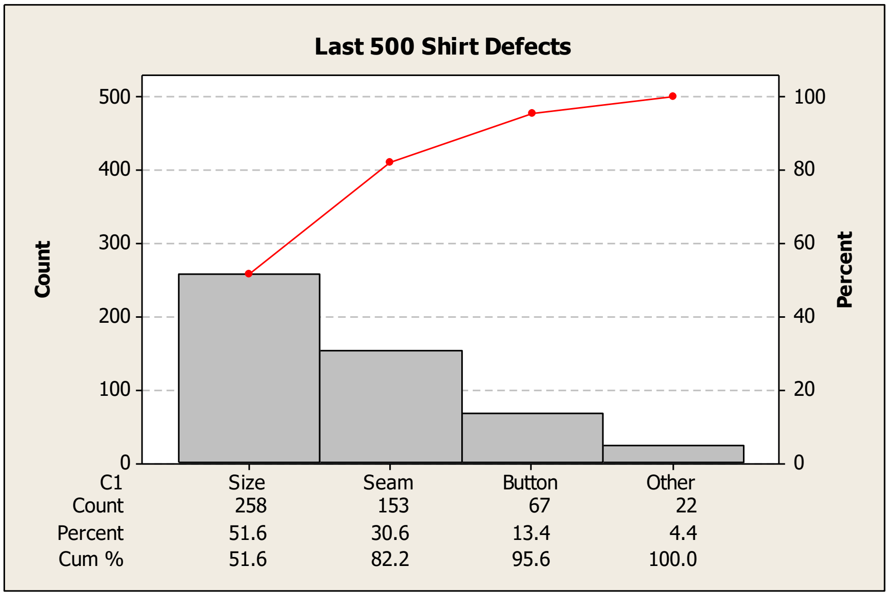
c. A circle graph better displays the proportions of the responses than a bar graph. It allows their relative frequencies to be compared.

The Pareto command generates bars, starting with the largest category.

NOTE: Pareto diagrams are primarily used for quality control applications and therefore MINITAB's PARETO command identifies the categories as "Defects", even when they may not be defects.

Computer and calculator commands to construct a Pareto diagram can be found on your Chapter 2 Technology Card.

2.2



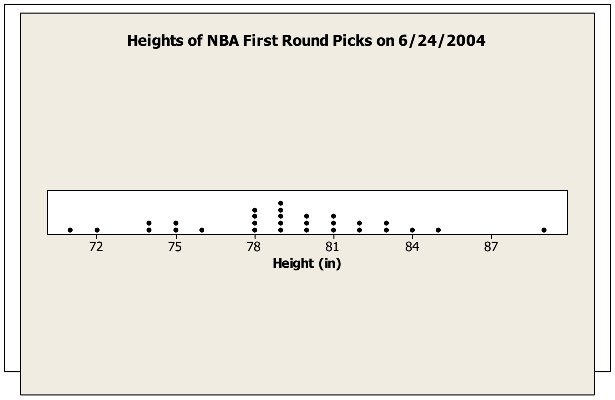
2.3 a. 150 defects

b. percent of scratch = *n*(scratch)/150 = 45/150 = 0.30 or 30.0%

c. 90.7% = [37.3 + 30.0 + 15.3 + 8.0]% [round-off error] 90.7% is the sum of the percentages for all defects that occurred more often than Bend, including Bend.

d. Two defects, Blem and Scratch, total 67.3%. If they can control these two defects, the goal should be within reach.

2.4 **a.**



b. shortest – 71 inches, tallest – 89 inches

c. most common – 79 inches, 5 players share that height

d. most common height = tallest column of dots

STEM-AND-LEAF DISPLAYS

1. Find the lowest and highest data values.
2. Decide in what "place value" position the data values will be split.
3. Stem = leading digit(s)
4. Leaf = trailing digit(s) (if necessary, data is first "rounded" to the desired position)
5. Sort stems and list.
6. Split data values accordingly, listing leaves on the appropriate stem.

The column on the left of the stem-and-leaf display is the cumulative count of the data from the top (low-value) down and the bottom (high-value) up until the class containing the median is reached. The number of data values for the median class is in parentheses.

Computer and calculator commands to construct a stem-and-leaf display can be found on youe technology card.

2.5 Points scored per game

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

Have a discussion about the four different graphical displays presented thus far. Which ones do the students find more informative? Talk about which types of graphs are used with attribute/qualitative data (circle or bar graphs) and which are used with quantitative data (stem-and-leaf, dot plot). As an exercise, the students could find an example of one or of each display from a newspaper or magazine. The students could then do a critical review with respect to the best display for that particular set of data.

Another technique for summarizing data is the grouped frequency distribution. Use the same set of values (test grades) that were used to introduce the stem-and-leaf display. It is a type of data students are familiar with and they would have a feel for how it should be grouped into classes. Test grades are often divided into classes such as 70s, 80s, 90s making for a convenient class width of 10 units. Use this set of data to demonstrate a frequency distribution and histogram, a relative frequency distribution and histogram, and a cumulative relative frequency distribution and its corresponding ogive.

Problems 2.9 and 2.10 are excellent problems to assign. Each covers the key points of data in a frequency distribution such as: class width, class midpoints, and class boundaries; the sum of the frequencies equals the sample size.

OBJECTIVE 2.2 PROBLEMS

Frequency distributions can be either grouped or ungrouped. Ungrouped frequency distributions have single data values as *x* values. Grouped frequency distributions have intervals of *x* values, therefore, use the class midpoints (class marks) as the *x* values.

Histograms can be used to show either type of distribution graphically. Frequency or relative frequency is on the vertical axis. Be sure the bars touch each other (unlike bar graphs). Increments and widths of bars should all be equal. A title should also be given to the histogram.

Computer or calculator commands to construct a histogram can be found on your Chapter 2 Technology Card. Note the two methods, depending on the form of your data.

2.6 a.

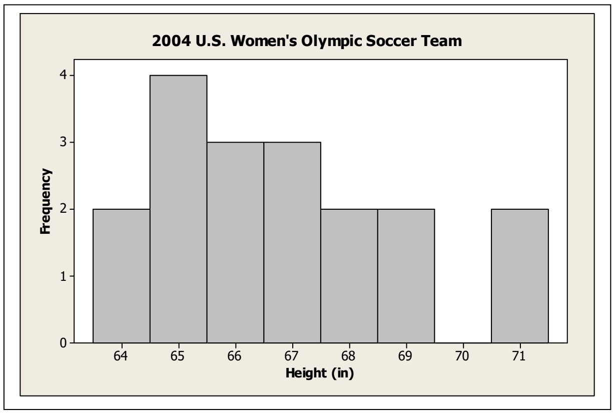
x	*f*
0	2
1	5
2	3
3	0
4	2
	12

b. *f* is frequency, therefore value of 1 occurred 5 times.
c. 2 + 5 + 3 + 0 + 2 = 12
d. The sum represents the sum of all the frequencies, which is the number of data, or the sample size.

2.7 a.

Height (in)	Frequency
64	2
65	4
66	3
67	3
68	2
69	2
71	2

b



c.

Height (in)	Rel. Freq.
64	0.111
65	0.222
66	0.167
67	0.167
68	0.111
69	0.111
71	0.111

d. 0.167 + 0.167 + 0.111 + 0.111 + 0.111 = 0.667 = 66.7%

An ogive is a line graph of a cumulative frequency or cumulative relative frequency distribution. Start the line at zero for a class below the smallest class. Plot the upper class boundary points from the remaining values of the cumulative (relative) frequency distribution. Connect all of the points with straight line segments. The last point (class) is at the value of one (vertically).

Computer and calculator commands to construct an ogive can be found on your Technology Card.

Parts of a grouped frequency distribution -

class boundaries = the low and the high endpoints of the interval

class width = distance from any point in one class to the same position point in the next class or the difference between the upper and lower class boundaries

class midpoint (mark) = (lower boundary + upper boundary)/2, midpoint of the interval

Example: with respect to the second class interval

30 - 40   form: (40 ≤ x < 50)
40 - 50   l 40 - 50 s boundary = 40
50 - 60   upper class boundary = 50
class width = 50 - 40 = 10
class midpoint = (40 + 50)/2 = 45

2.8 a. 35-45

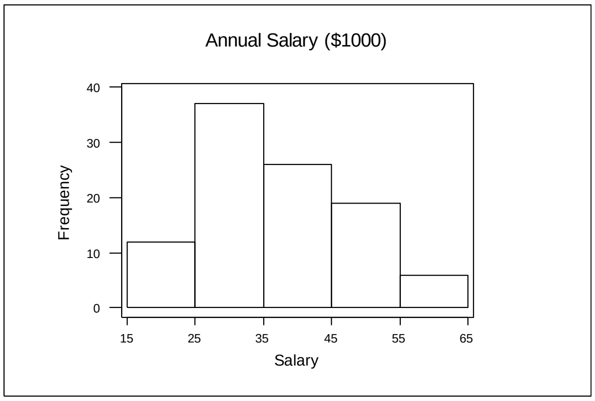
b. Values greater than or equal to 35 and also less than 45 belong to the class 35-45.

c. Difference between upper and lower class boundaries.

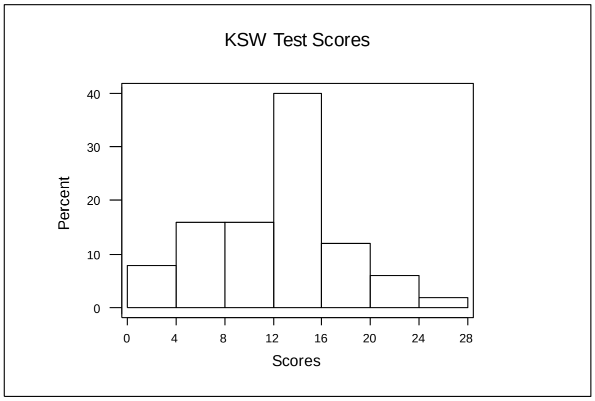
i. Subtracting the lower class boundary from the upper class boundary for any one class

ii. Subtracting a lower class boundary from the next consecutive lower class boundary

iii. Subtracting an upper class boundary from the next consecutive upper class boundary d.



2.9 a. 12 and 16
b. 2, 6, 10, 14, 18, 22, 26
c. 4.0
d. 0.08, 0.16, 0.16, 0.40, 0.12, 0.06, 0.02



e.

Refer to the frequency distribution information before problem 2.8 if necessary. Either class boundaries or class midpoints may be used to determine increments along the horizontal axis for histograms of grouped frequency distributions.

2.10 a.

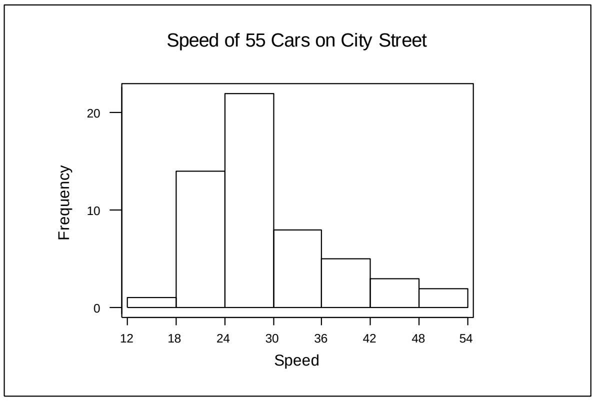
Class limits	frequency
12 - 18	1
18 - 24	14
24 - 30	22
30 - 36	8
36 - 42	5
42 - 48	3
48 - 54	2

b. class width = 6

c. class midpoint = (24+30)/2
= 27

lower class
boundary = 24

upper class
boundary = 30



d.

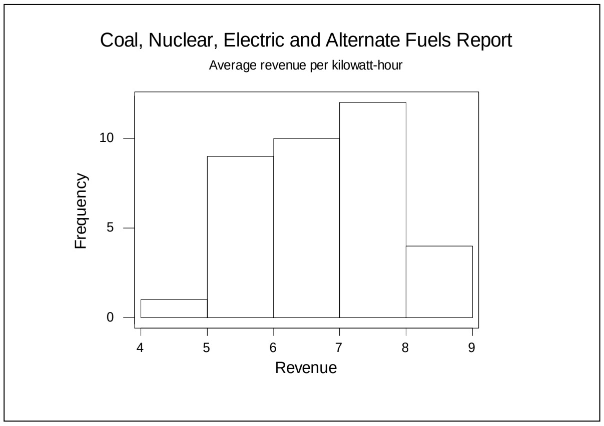
2.11 a.

4 – 5	1
5 – 6	9
6 – 7	10
7 – 8	12
8 – 9	4

b. 1

c. 4.5, 5.5, 6.5, 7.5, 8.5,

d.

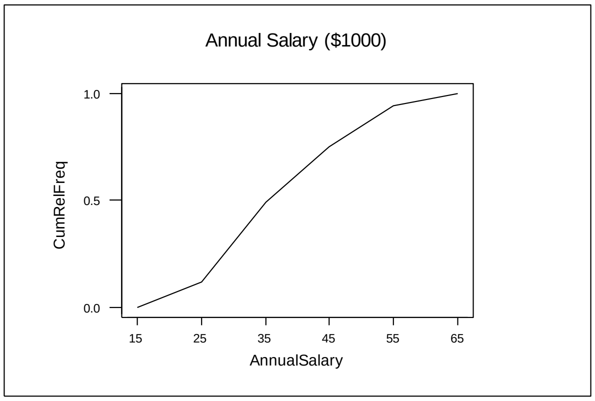


2.12 a.

Class Boundaries	Cumulative Frequency
15 ≤ x < 25	12
25 ≤ x < 35	49
35 ≤ x < 45	75
45 ≤ x < 55	94
55 ≤ x ≤ 65	100

b.

Class Boundaries	Cum. Rel. Frequency
15 ≤ x < 25	0.12
25 ≤ x < 35	0.49
35 ≤ x < 45	0.75
45 ≤ x < 55	0.94
55 ≤ x ≤ 65	1.00



c.

Using two small sets of data, one with an even sample size and one with an odd sample size, demonstrate each of the measures of central tendency. Use one of the sets for demonstration and the other for individual in-class practice. This is usually the first time students have seen the summation (Σ) notation.

Rework problem 2.16 with a 20 in place of the 9. Show the effects of an extreme value on the mean, median, mode and midrange. Cite possible examples where median may be more appropriate than the mean (a certain area's home prices, salaries in a company, etc.).

OBJECTIVE 2.3 PROBLEMS

NOTE: A measure of central tendency is a value of the variable. It is that value which locates the "average" value for a set of data. The "average" value may indicate the "middle" or the "center" or the most popular data value.

NOTATION AND FORMULAS FOR MEASURES OF CENTRAL TENDENCY

$\sum x$ = sum of data values
n = # of data values in the sample
$\bar{x}$ = sample mean = $\sum x/n$
$\tilde{x}$ = sample median = middle data value
$d(\tilde{x})$ = depth or position of median = $(n + 1)/2$
mode = the data value that occurs most often
midrange = (highest value + lowest value)/2

NOTE: REMEMBER TO RANK THE DATA BEFORE FINDING THE MEDIAN.
$d(\tilde{x})$ only gives the depth or position, not the value of the median. If n is even, $\tilde{x}$ is the average of the two middle values.

See Introductory Concepts for additional information about the Σ (summation) notation.

Computer and calculator commands to find the mean and median can be found your Chapter 2 Technology Card.

2.13 $\bar{x} = \sum x/n = (1+2+1+3+2+1+5+3)/8 = 18/8 = \underline{2.25}$

2.14 Ranked data: 70, 72, 73, 74, 76
$d(\tilde{x}) = (n+1)/2 = (5+1)/2 = $ 3rd; $\tilde{x} = \underline{73}$

2.15 Ranked data: 4.15, 4.25, 4.25, 4.50, 4.60, 4.60, 4.75, 4.90
$d(\tilde{x}) = (n+1)/2 = (8+1)/2 = $ 4.5th; $\tilde{x} = (4.50+4.60)/2 = \underline{4.55}$

2.16 a. $\bar{x} = \sum x/n = (2+4+7+8+9)/5 = 30/5 = \underline{6.0}$
b. $d(\tilde{x}) = (n+1)/2 = (5+1)/2 = $ 3rd; $\tilde{x} = \underline{7}$
c. no mode, no value repeats
d. midrange = $(H+L)/2 = (9+2)/2 = 11/2 = \underline{5.5}$

2.17 {28, 29, 33, 40, 41, 42, 44, 48, 48, 49}
a. $\bar{x} = \sum x/n = 402/10 = \underline{40.2}$
b. $d(\tilde{x}) = (n+1)/2 = (10+1)/2 = $ 5.5th; $\tilde{x} = \underline{41.5}$
c. midrange = $(H+L)/2 = (28+49)/2 = \underline{38.5}$
d. mode = $\underline{48}$

The 2.6 formula for variance is good for demonstrating to the students what variance and standard deviation are. The 2.10 formula is strongly suggested for use; it avoids the possibility of accumulating round-off errors.
The difference between the two formulas is emphasized in problem 2.22.
Students often have difficulty working through these standard deviation formulas. Review the order of operations, including the steps involved using the calculator. Have the students work out the problem individually. Check their answers. This approach avoids mistakes that are just calculator input errors. Emphasize the *SS* notation, as it will be used frequently throughout Chapter 3. The difference between *SS*(*x*) and Σx^2, both in name, notation and calculation, must be pointed out to the students. These two sums can cause much confusion.
Problem 2.24 is a good homework problem for demonstrating the relationship between range and standard deviation and the differences between sets of data.
Problem 2.25 is good for class discussion. Hopefully the students notice the negative sign.

OBJECTIVE 2.4 PROBLEMS

NOTE: A measure of dispersion is a value of the variable. It is that value which describes the amount of variation or spread in a data set. A small measure of dispersion indicates data that are closely grouped, whereas a large value indicates data that are more widely spread.

MEASURES OF DISPERSION - THE SPREAD OF THE DATA

Range = highest value - lowest value

Standard Deviation (s) = the average distance a data value is from the mean

$$s = \sqrt{\Sigma(x-\bar{x})^2/(n-1)}$$

Variance (s^2) = the square of the standard deviation
(i.e., before taking the square root)

For problems 2.21 and 2.22, be sure that the $\sum(x - \bar{x}) = 0$.

NOTE: Standard deviation and/or variance cannot be negative. This would indicate an error in sums or calculations.

See Introductory Concepts for additional information about Rounding Off. Computer and calculator commands to find the range and standard deviation can be found on your Chapter 2 Technology Card.
If using a non-graphing statistical calculator (one that lets you input the data points) to find the standard deviation of a sample, use the $\sigma(n\text{-}1)$ or s_x key. $\sigma(n)$ or σ_x would give the population standard deviation; that is, divide by "n" instead of "n-1".

2.18 a. The data value $x = 45$ is 12 units above the mean; therefore the mean must be 33.
b. The data value $x = 84$ is 20 units below the mean; therefore the mean must be 104.

2.19 The mean is the 'balance point' or 'center of gravity' to all the data values. Since the weights of the data values on each side of $\bar{x}$ are equal, $\Sigma(x-\bar{x})$ will give a positive amount and an equal negative amount, thereby canceling each other out.

Algebraically: $\Sigma(x-\bar{x}) = \Sigma x - n\bar{x} = \Sigma x - n\cdot(\Sigma x/n) = \Sigma x - \Sigma x = 0$

2.20 a. "Only zero and positive values can occur." Or "The smallest value would be 'zero'" or "All non-zero values are positive"
b. There would be "no variation." That is, "all values would be the same."
c. The existence of any variation in the data, that is, at least one value is different from the others.

2.21 a. range = H - L = 9 - 2 = 7

b. 1st: find mean, $\bar{x}$ = $\sum x/n$ = 30/5 = 6

	x	$x-\bar{x}$	$(x-\bar{x})^2$
	2	-4	16
	4	-2	4
	7	1	1
	8	2	4
	9	3	9
$\sum$	30	0	34

$s^2 = \sum(x-\bar{x})^2/(n-1)$

= 34/4 = 8.5

c. s = = = 2.915 = 2.9

An easier formula for s - sample standard deviation

1. Calculate "the sum of squares for x", $SS(x)$: $SS(x) = \sum x^2 - ((\sum x)^2/n)$

2. $s = \sqrt{SS(x)/n-1}$

This formula eliminates the problem of accumulating round-off errors.
NOTE: $SS(x)$ is formed from the "sum of squared deviations from the mean", $\sum(x-\bar{x})^2$. $\sum x^2$ is the "sum of the squared x's". $SS(x) \neq \sum x^2$.

2.22 a. 1st: find mean, $\bar{x}$ = $\sum x/n$ = 104/15 = 6.9

	x	$x-\bar{x}$	$(x-\bar{x})^2$
	4	-2.9	8.41
	5	-1.9	3.61
	5	-1.9	3.61
	6	-0.9	0.81
	6	-0.9	0.81
	6	-0.9	0.81
	7	0.1	0.01
	7	0.1	0.01
	7	0.1	0.01
	7	0.1	0.01
	8	1.1	1.21
	8	1.1	1.21
	8	1.1	1.21
	9	2.1	4.41
	11	4.1	16.81
$\sum$	104	+0.5*	42.95

$s^2 = \sum(x-\bar{x})^2/(n-1)$

= 42.95/14

= 3.0679

= 3.1

*The 0.5 is due to the round-off error introduced by using $\bar{x}$ = 6.9 instead of 6.933333.

b.

	x	x^2
	4	16
	5	25
	5	25
	6	36
	6	36
	6	36
	7	49
	7	49
	7	49
	7	49
	8	64
	8	64
	8	64
	9	81
	11	121
$\sum$	104	764

$SS(x) = \sum x^2 - ((\sum x)^2/n)$

$= 764 - ((104)^2/15)$

$= 764 - 721.0667 = 42.93333$

$s^2 = SS(x)/(n-1)$

$= 42.9333/14 = 3.0667 = \underline{3.1}$

c. $s = \ = \ = 1.751 = \underline{1.8}$

2.23 a.

```
            Police Recruits
                      • s = 2.8
                      •←--------→|
                      •       •
  •       •           •   •   •   •   •
  •   •   •       •   •   •   •   •   •
--+---+---+---+---+---+---+---+---+---+--
  25  26  27  28  29  30  31  32  33  34
         exercise capacity (minutes)
  ←-----------------------------------→
                  range = 9
```

b. $\bar{x} = 601/20 = 30.05$

c. range = $H - L = 34 - 25 = \underline{9}$

d. $n = 20$, $\sum x = 601$, $\sum x^2 = 18{,}209$
$SS(x) = \sum x^2 - ((\sum x)^2/n) = 18{,}209 - (601^2/20) = 148.95$
$s^2 = SS(x)/(n-1) = 148.95/19 = 7.83947 = \underline{7.8}$

e. $s = \ = \ = 2.7999 = \underline{2.8}$

f. See graph in (a).

g. Except for the value $x = 30$, the distribution looks rectangular. Range is a little more than 3 standard deviations.

2.24 Set 1: $\bar{x}$ = 250/5 = 50 | Set 2: $\bar{x}$ = 250/5 = 50

x	$x-\bar{x}$	$(x-\bar{x})^2$	x	$x-\bar{x}$	$(x-\bar{x})^2$	
46	-4	16	30	20	400	
55	+5	25	55	+5	5	25
50	0	0	65	+15	15	225
47	-3	9	47	-3	3	9
52	+2	4	53	+3	3	9
250	0	54	250	0	46	668

	$\sum x$	$\sum(x-\bar{x})$	$\sum(x-\bar{x})^2$	Range
Set 1:	250	0	54	9
Set 2:	250	0	668	35

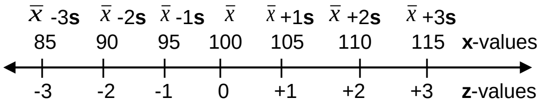
The values of $SS(x)$ [recall $SS(x) = \sum(x-\bar{x})^2$] and range reflect the fact that there is more variability in the data forming set 2 than in the data of set 1.

2.25 The statement is incorrect. The standard deviation can never be negative. There has to be an error in the calculations or a typographical error in the statement.

Stress the fact that data must be ranked before calculating any measure of position. Note also to the students that measures of position are not always equal to one of the actual data values. The measure may occur halfway between two values.
Discussion of different box-and-whisker displays may be helpful in leading into various distribution shapes. Problem 2.28 works well as an in-class or homework problem.

Demonstrate the meaning of *z* with an example and a number line. Have the students note the positive and negative *z*-values. Remind them that the signs are actually indicators of position with respect to the mean.

ex. let $\bar{x}$ = 100, *s* = 5, *z* = (*x*- $\bar{x}$)/*s* (see below)



Problem 2.31 presents a good use for *z*. It may be used to compare two different sets of data.

OBJECTIVE 2.5 PROBLEMS

NOTE: A measure of position is a value of the variable. It is that value which divides the set of data into two groups: those data smaller in value than the measure of position, and those larger in value than the measure of position.

To find any measure of position:

1. Rank the data - DATA MUST BE RANKED LOW TO HIGH
2. Determine the depth or position in two separate steps:
   a. Calculate $nk/100$, where n = sample size,
   k = desired percentile
   b. Determine $d(P_k)$:
   If $nk/100$ = integer $\Rightarrow$ add .5 (value will be halfway between 2 integers)
   If $nk/100$ = decimal $\Rightarrow$ round up to the nearest whole number
3. Locate the value of P_k

REMEMBER:
Q_1 = P_{25} = 1st quartile - 25% of the data lies below this value
Q_2 = P_{50} = $\tilde{x}$ = 2nd quartile - 50% of the data lies below this value
Q_3 = P_{75} = 3rd quartile - 75% of the data lies below this value

2.26 a. 91 is in the 44th position from the Low value of 39
91 is in the 7th position from the High value of 98

b. $nk/100 = (50)(20)/100 = 10.0$;
therefore $d(P_{20})$ = 10.5th from L
$P_{20} = (64+64)/2 = \underline{64}$

$nk/100 = (50)(35)/100 = 17.5$;
therefore $d(P_{35})$ = 18th from L
$P_{35} = \underline{70}$

c. $nk/100 = (50)(20)/100 = 10.0$;
therefore $d(P_{80})$ = 10.5th from H
$P_{80} = (88+89)/2 = \underline{88.5}$

$nk/100 = (50)(5)/100 = 2.5$;
therefore $d(P_{95})$ = 3rd from H
$P_{95} = \underline{95}$

2.27 Ranked data:

2.6 2.7 3.4 3.6 3.7   3.9 4.0 4.4 4.8 4.8
4.8 5.0 5.1 5.6 5.6   5.6 5.8 6.8 7.0 7.0

a. $nk/100 = (20)(25)/100 = 5.0$; therefore $d(P_{25})$ = 5.5th
$Q_1 = P_{25} = (3.7 + 3.9)/2 = \underline{3.8}$

$nk/100 = (20)(75)/100 = 15.0$; therefore $d(P_{75}) = 15.5$th

$Q_3 = P_{75} = (5.6 + 5.6)/2 = \underline{5.6}$

b. midquartile = $(Q_1 + Q_3)/2 = (3.8 + 5.6)/2 = \underline{4.7}$

c. $nk/100 = (20)(15)/100 = 3.0$; therefore $d(P_{15}) = 3.5$th

$P_{15} = (3.4+3.6)/2 = \underline{3.5}$

$nk/100 = (20)(33)/100 = 6.6$; therefore $d(P_{33}) = 7$th

$P_{33} = \underline{4.0}$

$nk/100 = (20)(90)/100 = 18.0$; therefore $d(P_{90}) = 18.5$th

$P_{90} = (6.8+7.0)/2 = \underline{6.9}$

> Box-and-whisker displays may be drawn horizontal or vertical.
> The Student Suite CD contains the Excel macro, Data Analysis Plus, for constructing box-and-whisker displays.

2.28 ranked data:

1.4	2.3	2.4	2.6	2.6	2.7	2.7	2.8	2.8	2.9	2.9
2.9	3.0	3.1	3.1	3.2	3.3	3.4	3.5	3.5	3.6	3.7
3.7	3.9	3.9	4.0	4.0	4.0	4.1	4.1	4.2	4.2	4.2
4.4	4.4	4.5	4.6	4.6	4.6	4.7	4.8	4.8	4.8	4.9
5.2	5.2	5.5	5.6	5.7	6.5	7.0	13.3			

a. $nk/100 = (52)(25)/100 = 13$; therefore $d(Q_1) = 13.5^{th}$
$Q_1 = (3.0+3.1)/2 = 3.05$

b. d(median) = $(52+1)/2 = 26.5^{th}$, Q_2 = median = $(4.0+4.0)/2 = 4.0$

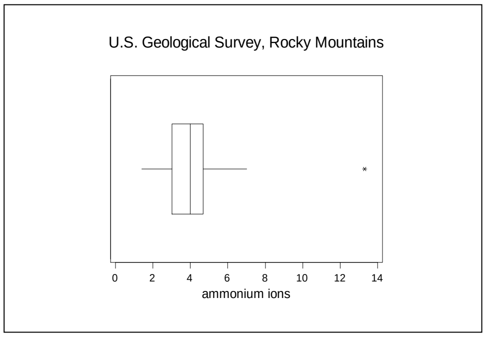
c. $nk/100 = (52)(75)/100 = 39$; therefore $d(Q_3) = 39.5^{th}$
$Q_3 = (4.6+4.7)/2 = 4.65$

d. midquartile = $(3.05+4.65)/2 = 3.85$

e. $nk/100 = (52)(30)/100 = 15.6$; therefore $d(P_{30}) = 16^{th}$
$P_{30} = 3.2$

f. 5-number summary: 1.4, 3.05, 4.0, 4.65, 13.3

g.



2.29 $z = (x - \text{mean})/\text{st.dev.}$

a. for $x = 54$, $z = (54 - 74.2)/11.5 = \underline{-1.76}$

b. for $x = 68$, $z = (68 - 74.2)/11.5 = \underline{-0.54}$

c. for x = 79, z = (79 - 74.2)/11.5 = 0.42
d. for x = 93, z = (93 - 74.2)/11.5 = 1.63

2.30 If z = (x - mean)/st.dev; then x = (z)(st.dev) + mean

a. for z = 0.0, x = (0.0)(20.0) + 120 = 120
b. for z = 1.2, x = (1.2)(20.0) + 120 = 144.0
c. for z = -1.4, x = (-1.4)(20.0) + 120 = 92.0
d. for z = 2.05, x = (2.05)(20.0) + 120 = 161.0

2.31 for A: z = (85 - 72)/8 = 1.625
for B: z = (93 - 87)/5 = 1.2
Therefore, A has the higher relative position.

OBJECTIVE 2.7 PROBLEMS

The presentation of Chebyshev's theorem and the empirical rule are critical. These two rules can be referred to many times in the following chapters; therefore, time spent emphasizing and explaining the rules is invaluable.
Some students have difficulty with fractions. A review of fractions and calculator input is useful. If there is a wide array of calculators, from very basic to graphing calculators, offer office time rather than class time to help individual students with their calculators.
Problem 2.36 is an excellent homework problem because it requires the use of many of the concepts learned throughout the chapter.

The empirical rule applies to a normal distribution.
Approximately 68% of the data lies within 1 standard deviation of the mean.
Approximately 95% of the data lies within 2 standard deviations of the mean.
Approximately 99.7% of the data lies within 3 standard deviations of the mean.

Chebyshev's theorem applies to any shape distribution.
At least 75% of the data lies within 2 standard deviations of the mean.
At least 89% of the data lies within 3 standard deviations of the mean.

2.32 a. ≈ 68% b. ≈ 95% c. ≈ 99.7%

2.33 The interval 22,500 to 37,500 represents the mean plus or minus three standard deviations.

a. If the distribution is normal, then approximately 99.7% of the distribution is contained within the interval.

b. If nothing is known about the shape of the distribution, then we can be sure that at least 89% of the distribution is contained within the interval.

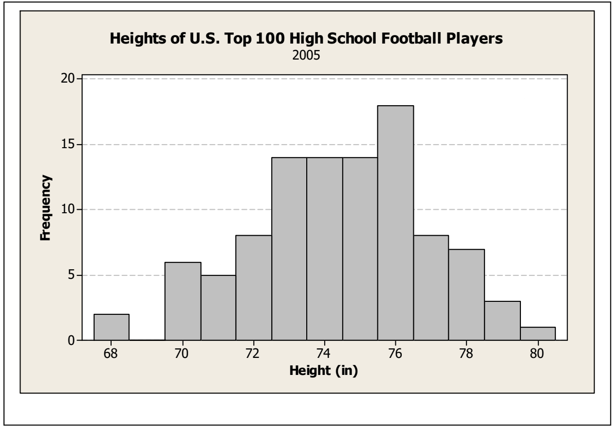
2.34 a. <u>50%</u> b. 0.50 – 0.34 = 0.16 = <u>16%</u>
c. 0.50 + 0.34 = 0.84 = <u>84%</u> d. 0.34 + 0.475 = 0.815 = <u>81.5%</u>

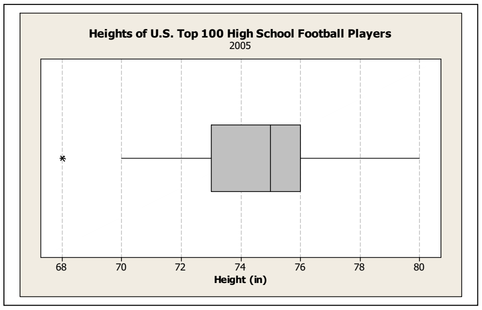
Chebyshev's theorem

At least $\left(1-\frac{1}{k^2}\right)\%$ of the data lies within k standard deviations of the mean. ($k > 1$)

2.35 a. at least 75% b. at least 89%

2.36 a. The second graph is the student’s choice, so answers will vary. The box and whisker display below the histogram is one option.





b. $\bar{x} = \sum x/n = 7442/100 = 74.42 = \underline{74.4}$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 554456 - (7442^2/100) = 622.36$
$s = \sqrt{SS(x)/(n-1)} = \sqrt{622.36/99} = \sqrt{6.2868} = 2.507285 = \underline{2.5}$

c. 68 68 70 70 70 70 70 70 71 71 71 71 71
72 72 72 72 72 72 72 72 73 73 73 73 73
73 73 73 73 73 73 73 73 73 74 74 74 74
74 74 74 74 74 74 74 74 74 74 75 75 75
75 75 75 75 75 75 75 75 75 75 75 76 76
76 76 76 76 76 76 76 76 76 76 76 76 76
76 76 76 77 77 77 77 77 77 77 77 78 78
78 78 78 78 78 79 79 79 80

d. $\bar{x} \pm 1s = 74.4 \pm (2.5)$ or 71.9 to 76.9
68% of the data (68/100) is between 71.9 and 76.9.

$\bar{x} \pm 2s = 74.4 \pm 2(2.5) = 74.4 \pm 5$ or 69.4 to 79.4
97% of the data (97/100) is between 69.4 and 79.4.

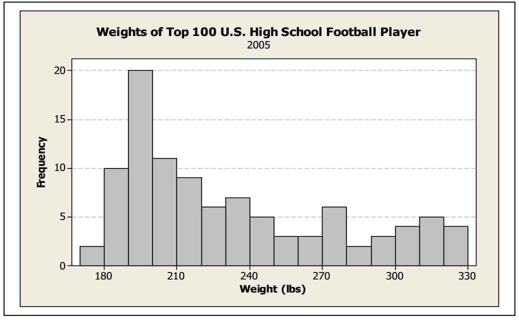
$\bar{x} \pm 3s = 74.4 \pm 3(2.5) = 74.4 \pm 7.5$ or 66.9 to 81.9
100% of the data (100/100) is between 66.9 and 81.9.

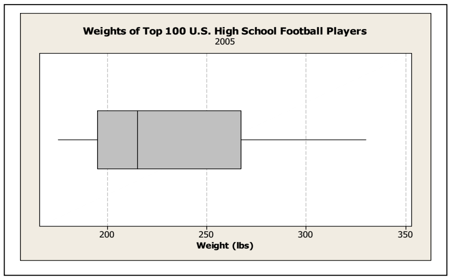
e. The empirical rule says approximately 68%, 95%, and 99.7% of the data are within one, two, and three standard deviations, respectively; the 68%, 97% and 100% do somewhat agree with the rule; based solely on this information, the distribution can be considered "approximately" normal.

f. Chebyshev's theorem says at least 75%, and 89%, of the data are within two, and three standard deviations, respectively; 97% and 100% both satisfy the theorem.

g. The graphs indicate a skewed right distribution, and therefore not normal. The boxplot shows an outlier to the left, and the right side of the box is much shorter than the left side, indicating skewness; the histogram shows one value to the far left and modal class is to right within the center cluster, thus both graphs show a slight skewness to the left that the empirical rule alone cannot detect.

2.37 a. The second graph is the student's choice, so answers will vary. The box and whisker display below the histogram is one option.





b. $\bar{x} = \sum x/n$ = 23032/100 = 230.32 = <u>230.3</u>

$SS(x) = \sum x^2 - ((\sum x)^2/n)$ = 5497014 - (23032^2/100) = 192283.76

$s = \sqrt{SS(x)/(n-1)} = \sqrt{192283.76/99}$

= 44.07108 = <u>44.1</u>

c.

175	175	180	180	180	180	182	185	185	185	185	187
190	190	190	190	190	190	190	190	190	191	192	193
195	195	195	195	195	195	196	197	200	200	200	200
204	205	205	205	205	207	208	210	210	210	212	214
215	215	215	218	220	220	220	220	221	225	230	230
230	230	233	235	237	240	240	240	240	242	250	250
250	262	265	268	270	270	275	275	275	275	280	280
290	290	297	300	300	300	305	310	310	310	315	316
322	323	330	330								

d. $\bar{x} \pm 1s$ = 230.3 ± (44.1) or <u>186.2</u> <u>to</u> <u>274.4</u>
67% of the data (67/100) is between 186.2 and 274.4.

$\bar{x} \pm 2s$ = 230.3 ± 2(44.1) = 230.3 ± 88.2 or <u>142.1</u> <u>to</u> <u>318.5</u>
96% of the data (96/100) is between 142.1 and 318.5.

$\bar{x} \pm 3s$ = 230.3 ± 3(44.1) = 230.3 ± 132.3 or <u>98</u> <u>to</u> <u>362.6</u>
100% of the data (100/100) is between 98 and 362.6.

e. The empirical rule says approximately 68%, 95%, and 99.7% of the data are within one, two, and three standard deviations, respectively; the 67%, 96% and 100% do somewhat agree with the rule; based solely on this information, the distribution can be considered "approximately" normal.

f. Chebyshev's theorem says at least 75%, and 89%, of the data are within two, and three standard deviations, respectively; 96%, and 100% both satisfy the theorem.

g. The graphs do not agree with above answers. The graphs show a skewed right distribution. The histogram has a long right tail with over 10 classes to the right of the main clustering and only one class to the left. The histogram even shows some clustering above 300. The boxplot shows a skewness to the right; the right side of the box is much wider than the left and the right whisker is much longer than the left whisker. Both graphs clearly show this distribution to be skewed right. We need both the graphs and empirical rule to hold in order to conclude that the distribution is normal.

Helpful hint to use when expecting to count data on histogram:

Minitab: While on the Histogram dialogue box,
    Select: **Labels > Data Labels...**
    Select: **Use *y*-value labels**

This will direct the computer to print the frequency of each class above its corresponding bar.

Excel: Returns a frequency distribution with the histogram.
TI-83/84 Plus: Use the TRACE and arrow keys.

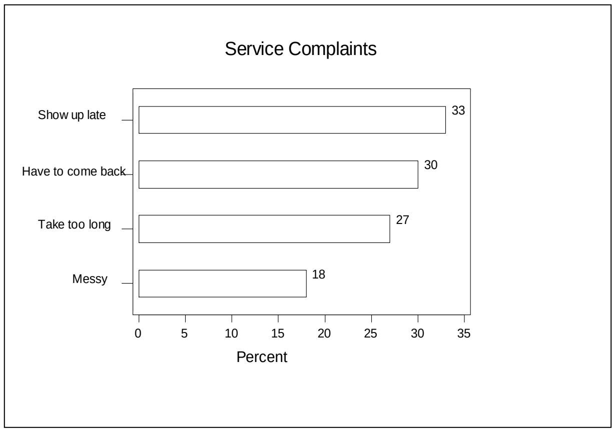
OBJECTIVE 2.7 PROBLEMS

Make students aware of misrepresentations that can occur in statistical graphs by bringing several examples to class. Perhaps bonus points could be given for students finding such a graph on their own.

2.38 a. The graph is a bar graph. The classes are a mix of numerical and categorical variables.
b. The graph has nonuniform classes, a requirement for histograms.

2.39 a. Answers will vary. Here are a few possibilites:
1. The response "Take too long" is twice as likely as "Messy"
2. The response "Have to come back" is three times as likely as "Messy"
3. The response "Show up late" is four times as likely as "Messy"
4. The response "Show up late" is twice as likely as "Take too long"
5. The percentage answering with each response decreases the same amount from response to response as you read down the Snapshot.

b.



c. Comments – "Messy" is a substantial percent when displayed against the other complaints in the proper format. "Show up late" is not even twice that of "Messy". The difference between "Show up late" and "Take too long" is minimal.

Solutions for Extra Problems Online

Below you will find the solutions to the extra problems, which can be accessed online at 4ltrpress.cengage.com/stat.

OBJECTIVE 2.1 EXTRA PROBLEMS

2.40 a. Answers will vary. Possibilities might include: sort data, frequency of each value, average.

b. Answers will vary. Possibilities might include: location of my data value with respect to the other values, proximity to average.

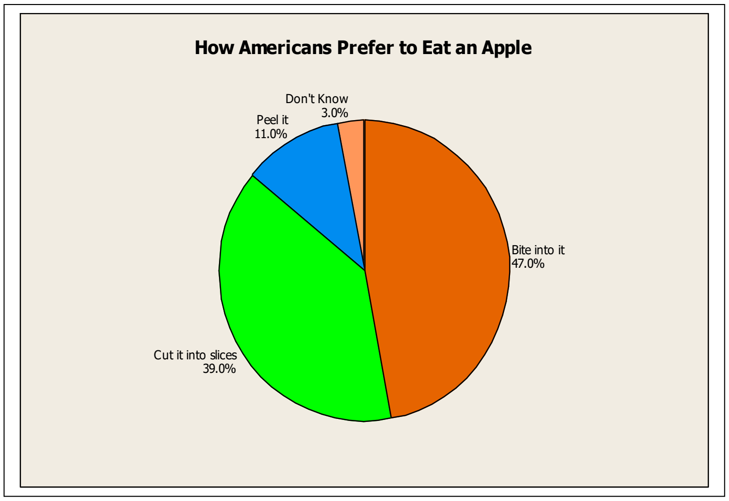
2.41 a. through c. Answers will vary.

2.42 a. Answers will vary but both graphs show relative size with respect to the individual answers.

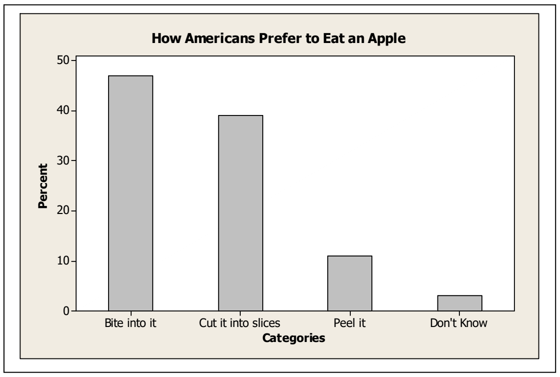
b. Answers will vary. The circle graph does a better job of representing the relative proportions of the answers to the group as a whole.

c. The bar graph is more dramatic in representing the relative proportions between the individual answers.

2.43 a.

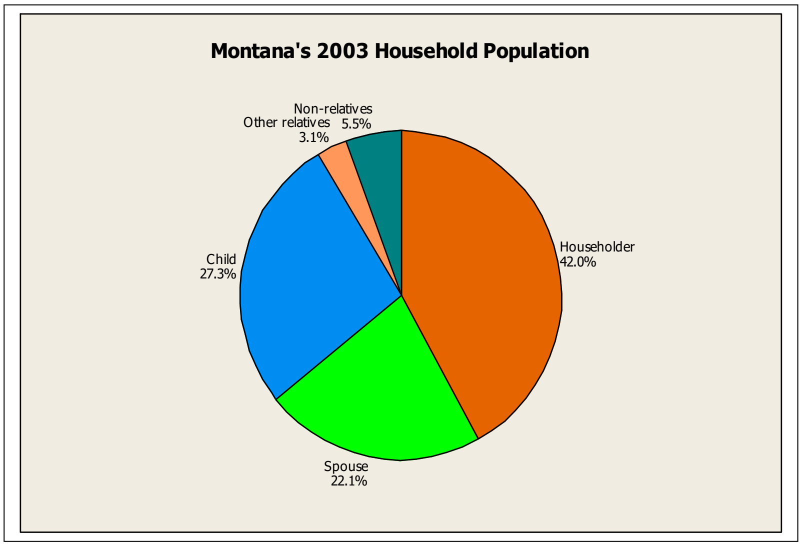


b.

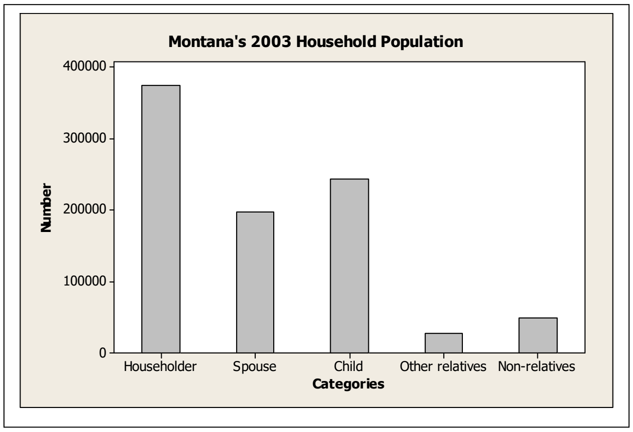


c. Answers will vary. The circle graph shows relative size with respect to the whole, whereas the bar graph shows relative size with respect to the other categories.

2.44 a.

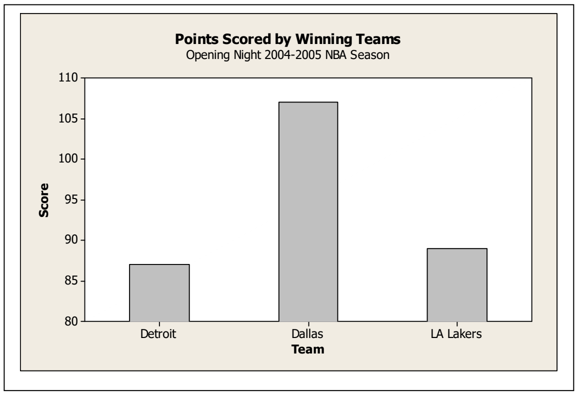


b.

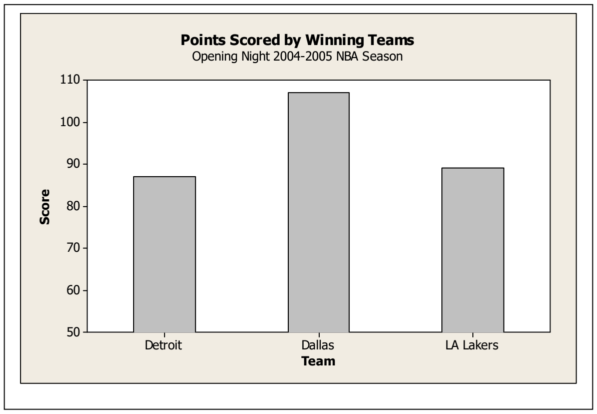


c. Answers will vary. The pie chart appears to be more informative about Montana’s population. The relative proportions with respect to the whole gives the information most likely needed: where the majority is, etc.

2.45 a.



b.



c. The bar graph in part a makes the NBA scores appear that they vary more. The Dallas team seems like it outscored the other team by 3 times as many points.

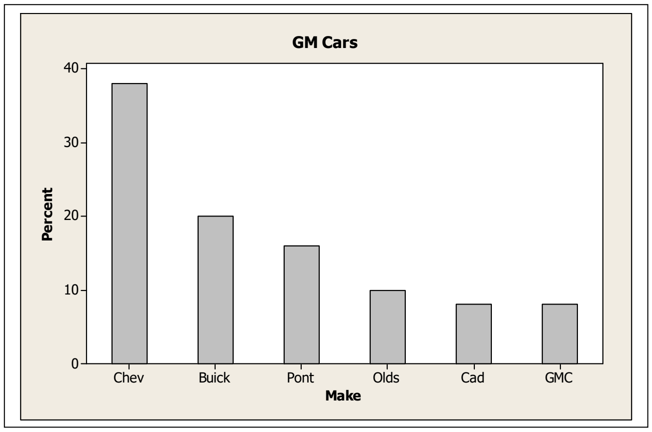
d. To make a more accurate representation, begin the vertical scale at zero.

$$\text{Percentage} = \frac{\text{\# of a particular make of car}}{\text{total \# of cars}}$$

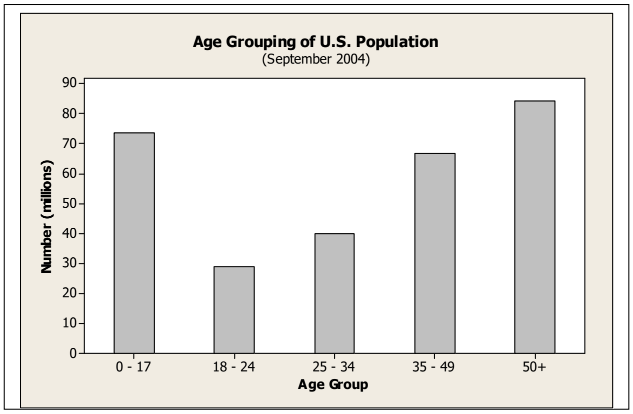
2.46 a. Chev - 19, Pont - 8, Olds - 5, Buick - 10,
Cad – 4, GMC - 4

b. Chev - 38%, Pont - 16%, Olds - 10%, Buick - 20%,
Cad - 8%, GMC – 8%

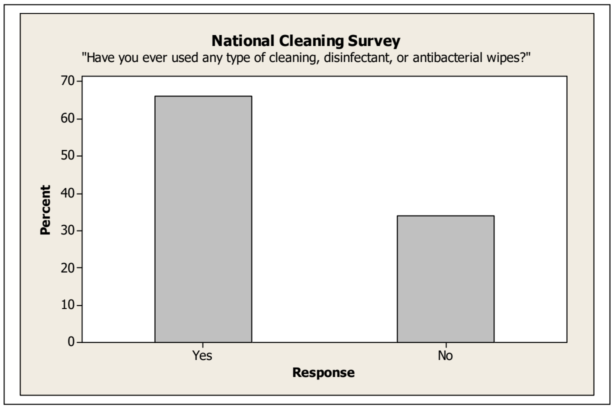
c.



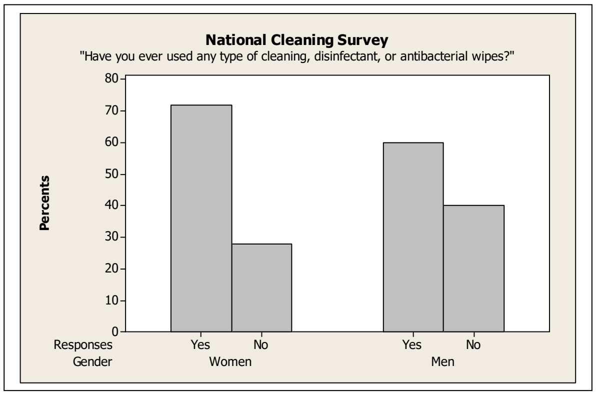
2.47


Age Grouping of U.S. Population
(September 2004)
Number (millions)
0
10
20
30
40
50
60
70
80
90
0 - 17
18 - 24
25 - 34
35 - 49
50+
Age Group

2.48 a.

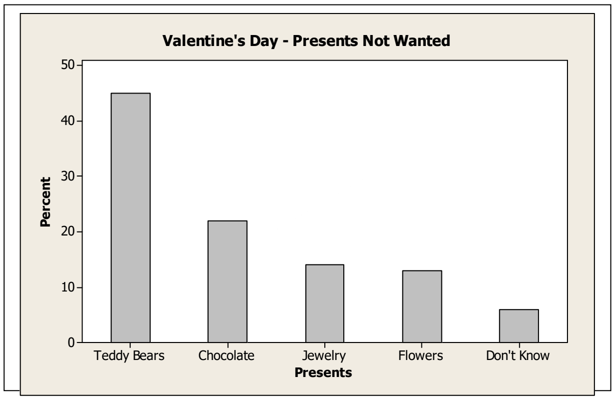

National Cleaning Survey
"Have you ever used any type of cleaning, disinfectant, or antibacterial wipes?"
Percent
0
10
20
30
40
50
60
70
Yes
No
Response

b.

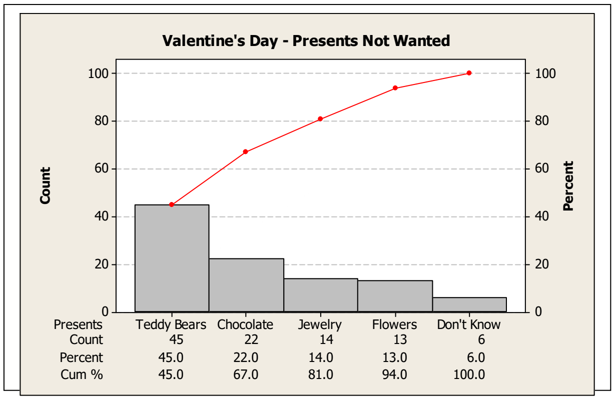

National Cleaning Survey
"Have you ever used any type of cleaning, disinfectant, or antibacterial wipes?"
Percents
0
10
20
30
40
50
60
70
80
Responses
Gender
Yes
No
Women
Yes
No
Men

c. Both graphs demonstrate that more people have used cleaning wipes than have not. The first graph shows the data for all people, whereas the second graph demonstrates the difference in genders. Since the sample sizes per women and men were close, the comparative proportions shown give an accurate picture.

2.49 a.



b.

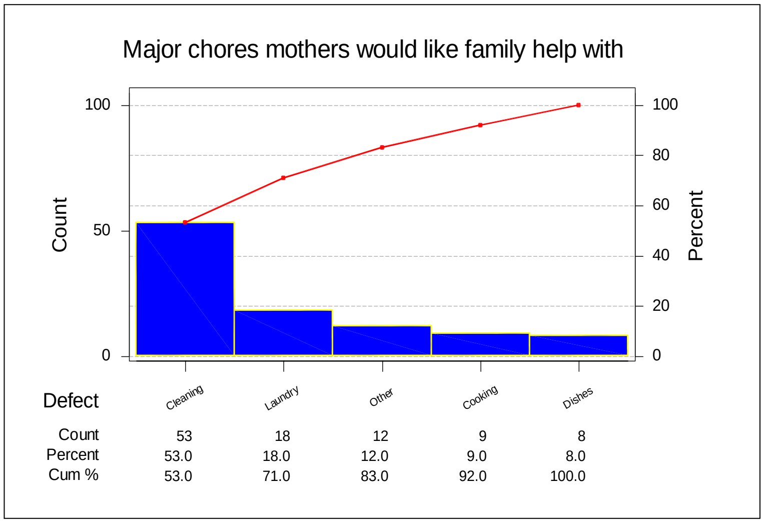


c. Avoid buying teddy bears, chocolates, and jewelry. These presents are listed first due to their having the highest percentages.

Frequency = (total #)(percentage)

d. 135 – teddy bears, 66 – chocolates, 42 – jewelry, 39 – flowers

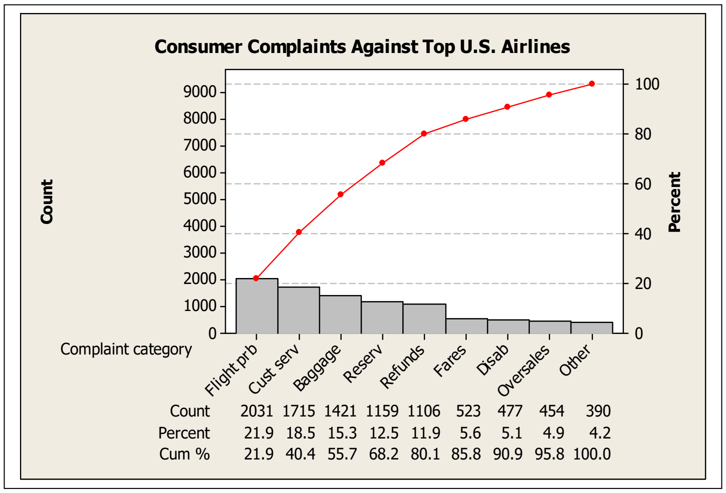
2.50 a.



Defect	Cleaning	Laundry	Other	Cooking	Dishes
Count	53	18	12	9	8
Percent	53.0	18.0	12.0	9.0	8.0
Cum %	53.0	71.0	83.0	92.0	100.0

b. The "Other" category is too large; it is a collection of several answers and as such is larger than two of the categories. If it were broken down, then the Pareto diagram would have the categories in order of mother's wishes.

2.51 a.

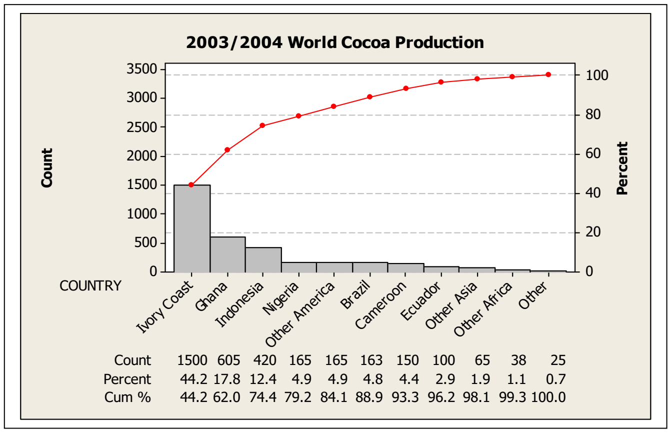


Complaint category	Flight prb	Cust serv	Baggage	Reserv	Refunds	Fares	Disab	Oversales	Other
Count	2031	1715	1421	1159	1106	523	477	454	390
Percent	21.9	18.5	15.3	12.5	11.9	5.6	5.1	4.9	4.2
Cum %	21.9	40.4	55.7	68.2	80.1	85.8	90.9	95.8	100.0

b. Flight problem, Customer service, Baggage, Reservations and Refunds would cover 80% of the airlines' problems. These categories are listed first to show the most effect, and the cumulative line graph shows that 80% includes up to and including Refunds.

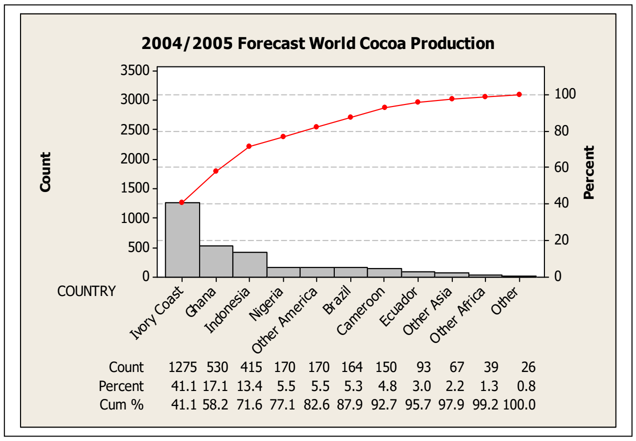
2.52 a. Total for 2003/2004 = 3396;
Total for 2004/2005 forecast = 3099
3396 – 3099 = 297 = 297,000 tonnes
percent decrease = 297/3396 = 0.087 = 8.7% decrease

b.



	Ivory Coast	Ghana	Indonesia	Nigeria	Other America	Brazil	Cameroon	Ecuador	Other Asia	Other Africa	Other
Count	1500	605	420	165	165	163	150	100	65	38	25
Percent	44.2	17.8	12.4	4.9	4.9	4.8	4.4	2.9	1.9	1.1	0.7
Cum %	44.2	62.0	74.4	79.2	84.1	88.9	93.3	96.2	98.1	99.3	100.0

c.



	Ivory Coast	Ghana	Indonesia	Nigeria	Other America	Brazil	Cameroon	Ecuador	Other Asia	Other Africa	Other
Count	1275	530	415	170	170	164	150	93	67	39	26
Percent	41.1	17.1	13.4	5.5	5.5	5.3	4.8	3.0	2.2	1.3	0.8
Cum %	41.1	58.2	71.6	77.1	82.6	87.9	92.7	95.7	97.9	99.2	100.0

d. Africa – 41.1 + 17.1 + 5.5 + 4.8 + 1.3 = 69.8%
Americas – 5.5 + 5.3 + 3.0 = 13.8%
Asia – 13.4 + 2.2 + 0.8 = 16.4%

Picking increments (spacing between tick marks) for a dot plot

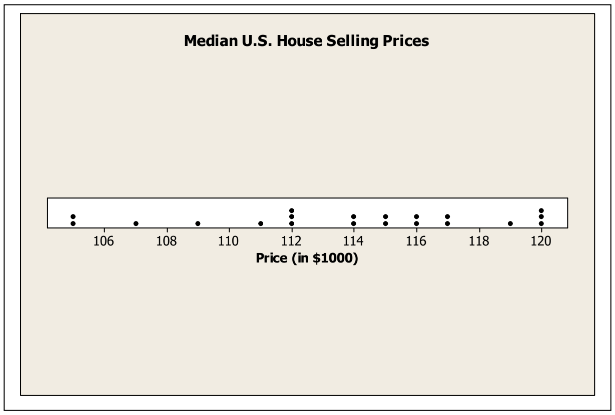
1. Calculate the spread (highest value minus the lowest value).
2. Divide this value by the number of increments you wish to show (no more than 7 usually).
3. Use this increment size or adjust to the nearest number that is easy to work with (5, 10, etc.).

Computer and calculator commands to construct a dotplot can be found on your Chapter 2 Technology Card.

2.53 Points scored per game by basketball team

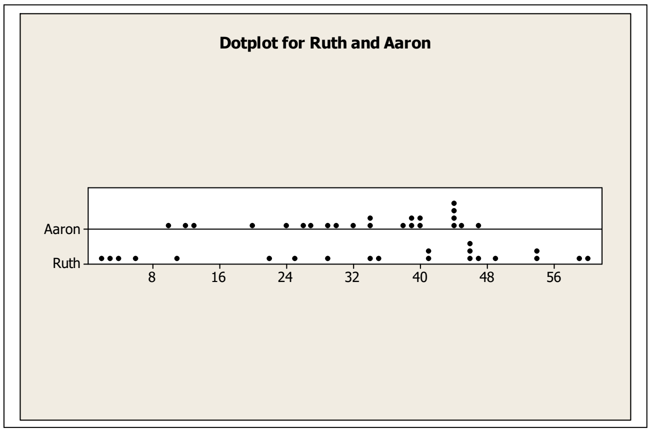
```
                                       .
              .         .    .. :..   .:  . . .  .
     ---+---------+---------+---------+---------+---------+--- points
       30        40        50        60        70        80
```

2.54 a.



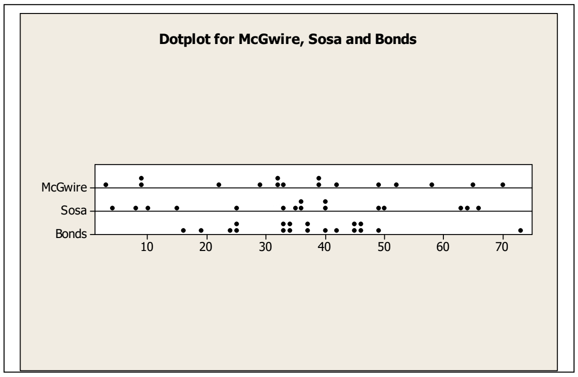
b. The majority of the median home prices across the U.S. are between $112,000 and $120,000. The maximum price for this sample was $120,000 and the lowest was $105,000.

2.55 a.



b. Aaron: 755 total homeruns, 30 or more homeruns in each of 15 seasons; Ruth: 40 or more homeruns in each of 11 seasons, including 50 or more in 4 of those seasons

c.



d. McGwire: 70 homeruns in one season, 50 or more homeruns in each of 4 seasons, 40 or more in each of 5 of those seasons;

Bonds: 73 homeruns in one season, 40 or more homeruns in each of 8 seasons;

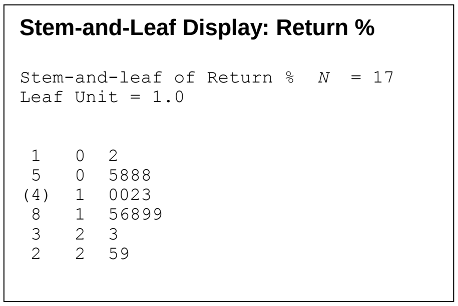
Sosa: no seasons with 70 or more homeruns.

2.56 Overall length of commutators

```
                        •
                        •
                   • •  ••
             •   • •••  •• • •
     • ••    ••• • •••• •••• • • ••  •
---+----+----+----+----+----+----+----+---
 18.740    18.780    18.820    18.860    length
```

2.57 a. 15
b. 11.2, 11.2, 11.3, 11.4, 11.7
c. 15.6
d. 13.7; 3

2.58 a.



Stem-and-Leaf Display: Return %

```
Stem-and-leaf of Return %  N  = 17
Leaf Unit = 1.0


 1   0  2
 5   0  5888
(4)  1  0023
 8   1  56899
 3   2  3
 2   2  59
```

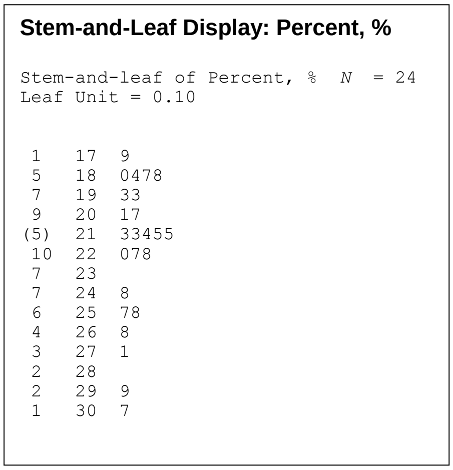
b. The bulk of the total returns are between 5 and 19. The distribution is skewed to the right, since there is more data above 19 than below 5.

2.59 a. Quik Delivery's delivery charges

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

b. The distribution is slightly skewed to the right; the bulk of the distribution is between 2.8 and 5.4, with only one smaller value and 6 larger values that create a longer right-hand tail.

2.60 a.



Stem-and-Leaf Display: Percent, %

```
Stem-and-leaf of Percent, %  N  = 24
Leaf Unit = 0.10


 1   17  9
 5   18  0478
 7   19  33
 9   20  17
(5)  21  33455
 10  22  078
 7   23
 7   24  8
 6   25  78
 4   26  8
 3   27  1
 2   28
 2   29  9
 1   30  7
```

b. The distribution is strongly skewed right; the bulk of the distribution is below 23, while the values above 23 spread out to 30, creating a relatively long right-hand tail.

2.61 a. The place value of the leaves is in the hundredths place; i.e., 59|7 is 5.97.
b. 16
c. 5.97, 6.01, 6.04, 6.08

d. The left hand column shows cumulative frequencies starting at the top and the bottom until it reaches the class that contains the median. The number in parentheses is the frequency for just the median class.

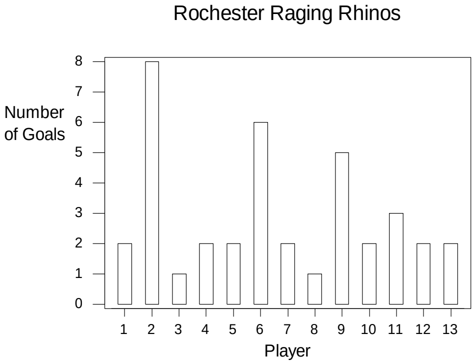
2.62 a. The place value of the leaves is in the tens place; i.e., 60|7 is 6070.
b. 6070, 6080, 6100, 6130
c. 6190, 6430

OBJECTIVE 2.2 EXTRA PROBLEMS

2.63 Answers will vary. Bar graphs are typically used for qualitative data; spacing between bars separates the non-sequential categories. Histograms are used for numerical data; the bars are adjacent due to sequential data.

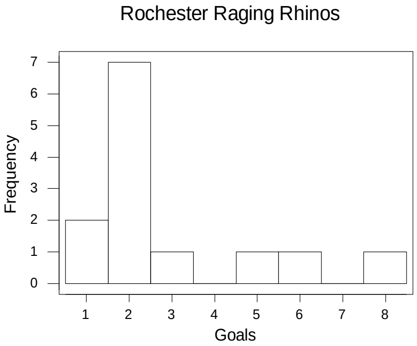
2.64 a. Bar graph, because the player's name is qualitative.

b.



c. Histogram, because you are working only with numerical data and want to show sequence.

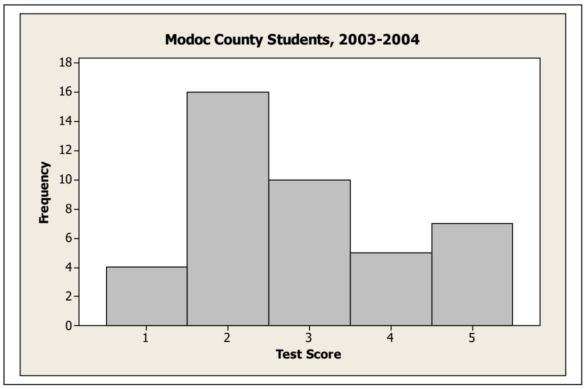
d.



2.65 a.

Test Score	Count
1	4
2	16
3	10
4	5
5	7

b.



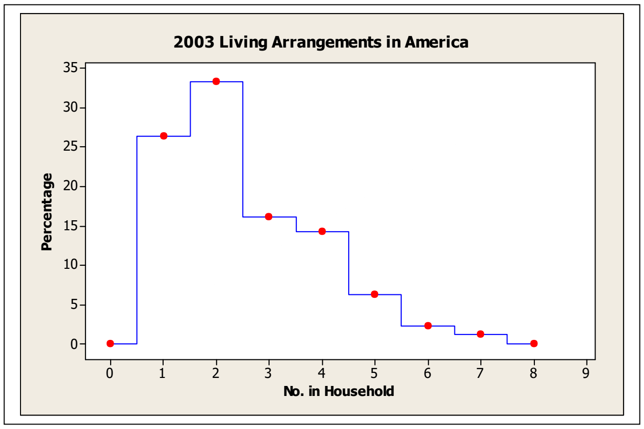
Relative frequency = $\frac{\text{frequency}}{\text{sample size}}$

c.

Test Score	Rel. Freq.
1	0.0952
2	0.3810
3	0.2381
4	0.1190
5	0.1667

d. 0.2381 + 0.1190 + 0.1667 = 0.5238 = 52.4%

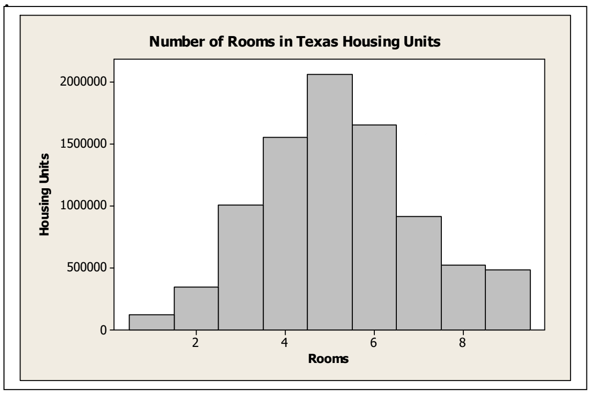
2.66 a.



b. skewed right

c. 1 and 2 member households account for most of the distribution.

2.67 a.



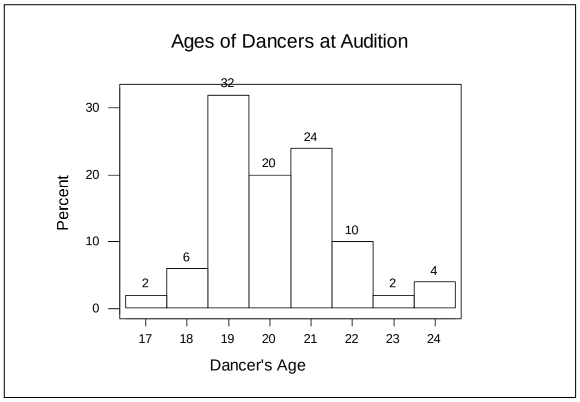
b. Mounded, truncated on right due to 9+ class

c. Centered on 5 rooms, there are 3 to 7 rooms in most households.

2.68 a.

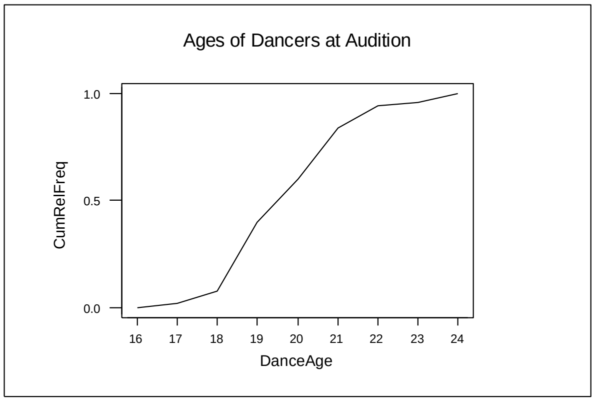
age	frequency	b. age	rel.freq.	d. age	cum.rel.freq.
17	1	17	0.02	17	0.02
18	3	18	0.06	18	0.08
19	16	19	0.32	19	0.40
20	10	20	0.20	20	0.60
21	12	21	0.24	21	0.84
22	5	22	0.10	22	0.94
23	1	23	0.02	23	0.96
24	2	24	0.04	24	1.00
	50		1.00		

```
CHECKS:  sum of frequencies = sample size
         sum of relative frequencies = 1.00
```



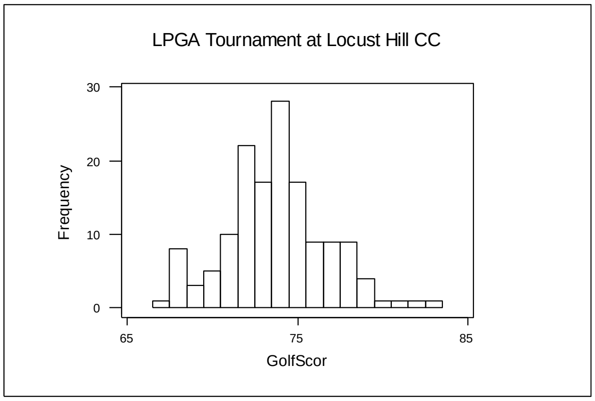
c.

e.



2.69 a.

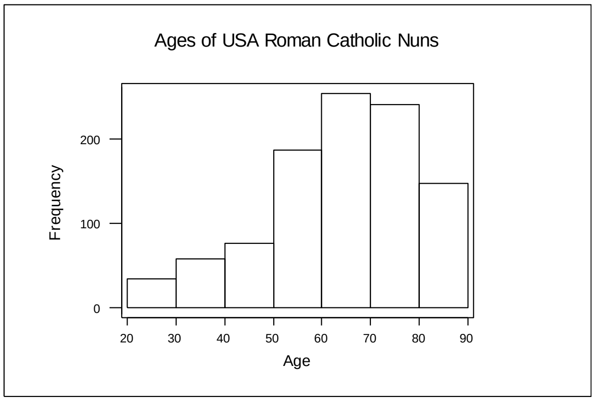
x	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
f	1	8	3	5	10	22	17	28	17	9	9	9	4	1	1	1	1



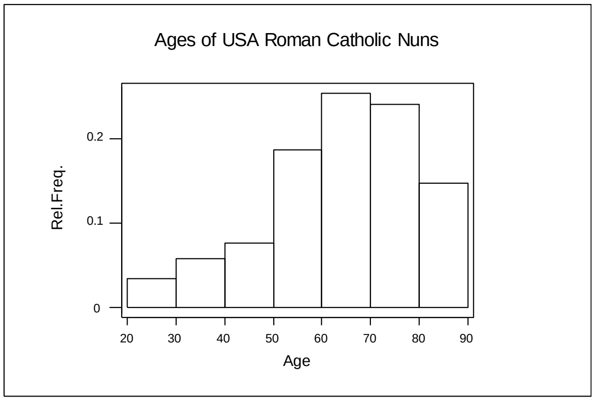
b.

2.70 a. Time of strike
b. One-hour interval
c. Most strikes occur around noon and none at night.
d. The tallest bar occurred around noon and no data occurred at night.

2.71 Similarities: same classes of data, same shape
Differences: vertical vs. horizontal, individual data points vs. grouped



2.72 a.



b.

c. The relative frequency histogram may be easier to understand because it will be in percentages that can be used to compare to the *USA Today* report.

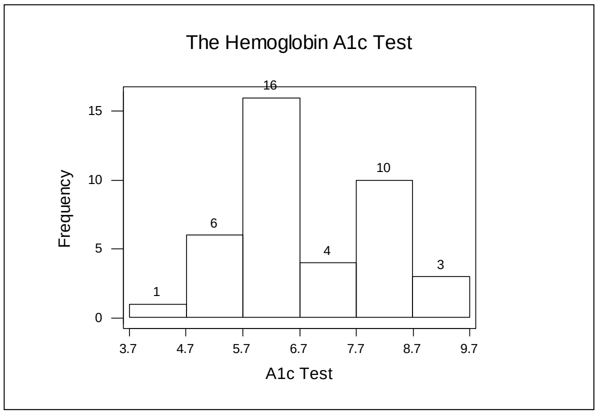
Refer to frequency distribution information before problem 2.39 if necessary. Either class boundaries or class midpoints may be used to determine increments along the horizontal axis for histograms of grouped frequency distributions.

2.73 a.

Class boundaries	frequency
3.7 - 4.7	1
4.7 - 5.7	6
5.7 - 6.7	16
6.7 - 7.7	4
7.7 - 8.7	10

8.7 - 9.7	3
	40

b. 4.2, 5.2, 6.2, 7.2, 8.2, 9.2



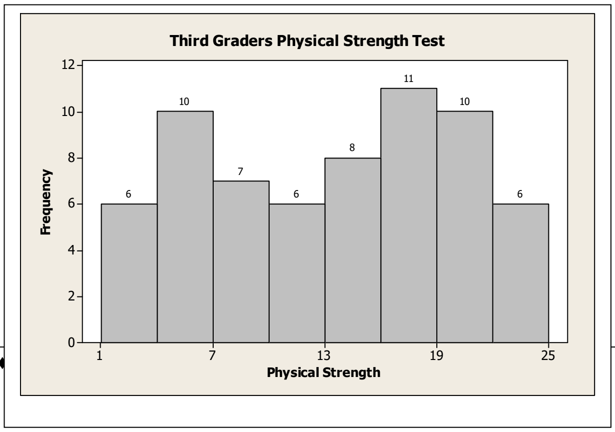
c.

2.74 a. Third Graders at Roth Elementary School

```
                   :
       . . : .     :     .   . : . : : : . . .
   . : : : : :   . : : . : . : : : : : : : : : . :
 +---------+---------+---------+---------+---------+--PhyStren
0.0       5.0      10.0      15.0      20.0      25.0
```

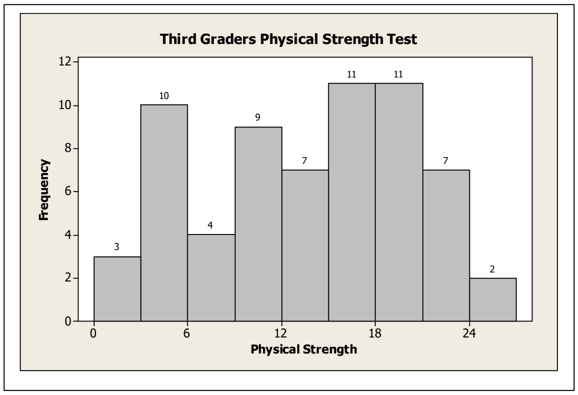
b.

Class boundaries	frequency
1 - 4	6
4 - 7	10
7 - 10	7
10 - 13	6
13 - 16	8
16 - 19	11
19 - 22	10
22 - 25	6
	64



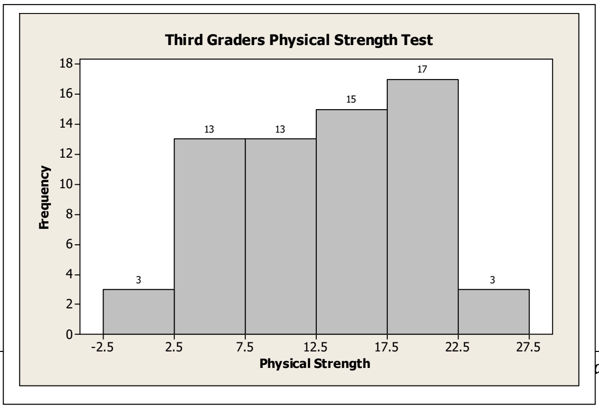
c.

Class boundaries	frequency
0 - 3	3
3 - 6	10
6 - 9	4
9 - 12	9
12 - 15	7
15 - 18	11
18 - 21	11
21 - 24	7
24 - 27	2
	64



d.

Class boundaries	frequency
-2.5 - 2.5	3
2.5 - 7.5	13
7.5 - 12.5	13
12.5 - 17.5	15
17.5 - 22.5	17
22.5 - 27.5	3
	64



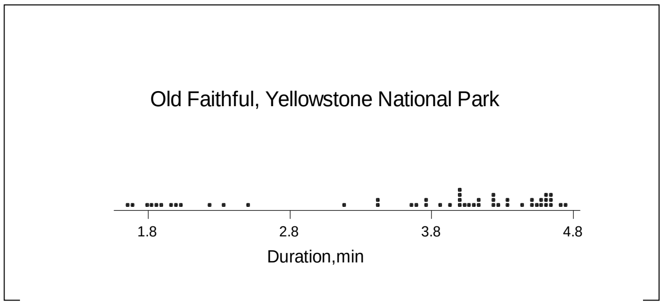
e. Answers will vary.

f. The histograms in parts b and c demonstrate a bimodal distribution, whereas the distribution in part d is skewed left.

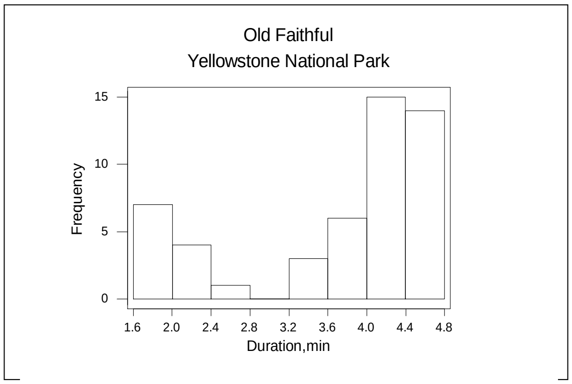
Dotplot shows mode to be 9, which is in the 7-10 class and shows a cluster centered around 17; the histogram shows the two modal classes to be 4-7 and 16-22. The mode is not in either modal class.

g. Answers will vary, but as the number of classes and the choice of class boundaries change, values will fall into various classes, thereby giving different appearances, all for the same set of data.

2.75 a.



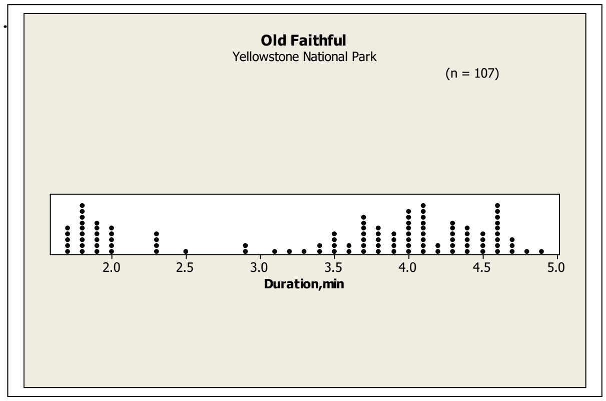
b.

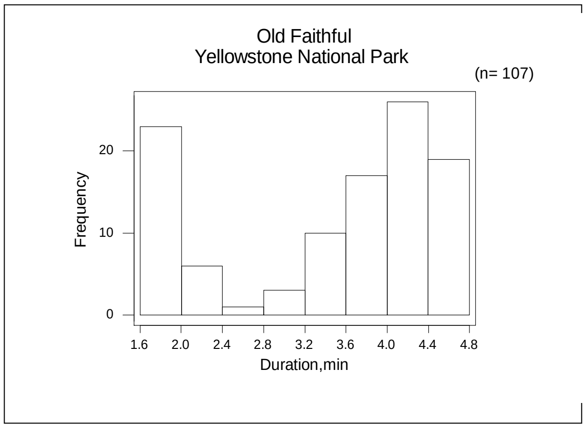


A guideline that can be used for selecting the number of classes is: *n*(classes) ≈

c. and d. The histograms will vary as class boundaries and class widths are changed.

e.



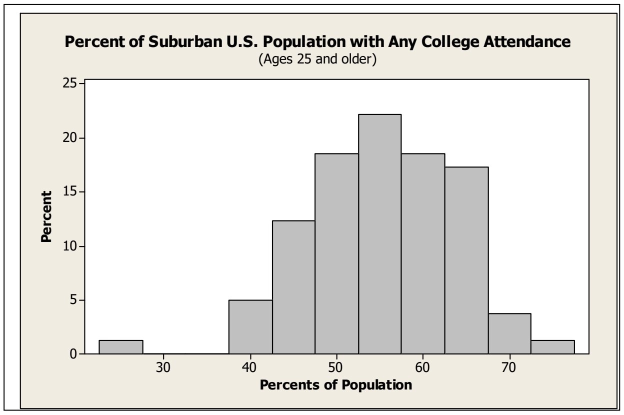


f. Answers will vary. Histogram is easier to read.
g. Answers will vary.

2.76 a. and b.

Class boundaries	Class Midpoint	Frequency
22.5 - 27.5	25	1
27.5 - 32.5	30	0
32.5 - 37.5	35	4
37.5 - 42.5	40	10
42.5 - 47.5	45	15
47.5 - 52.5	50	18
52.5 - 57.5	55	15
57.5 - 62.5	60	14
62.5 - 67.5	65	3
67.5 - 72.5	70	1

c.



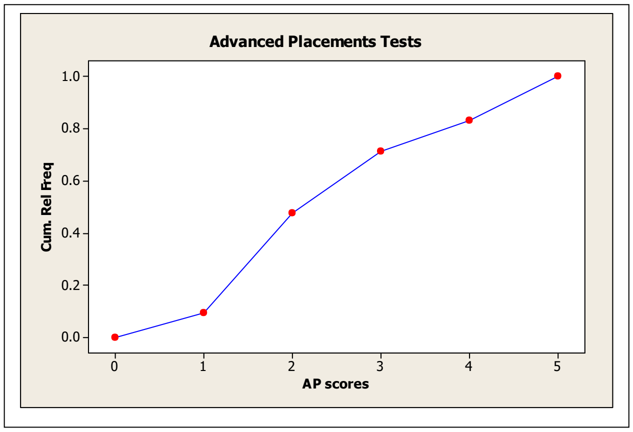
2.77 a. Answers will vary. Some possibilities include: Symmetric - weight of dry cereal per box, breaking strength of certain type of string
b. Answers will vary. One possibility includes: Uniform - result from rolling a die several hundred times
c. Answers will vary. Some possibilities include: Skewed Right - salaries, high school class sizes
d. Answers will vary. One possibility includes: Skewed left - hour exam scores
e. Answers will vary. One possibility includes: Bimodal - heights, weights for groups containing both male and female

2.78 a. uniform
b. J-shaped
c. skewed right

2.79 a.

Test Score	Cum Rel Freq
1	0.095
2	0.476
3	0.714
4	0.833
5	1.000

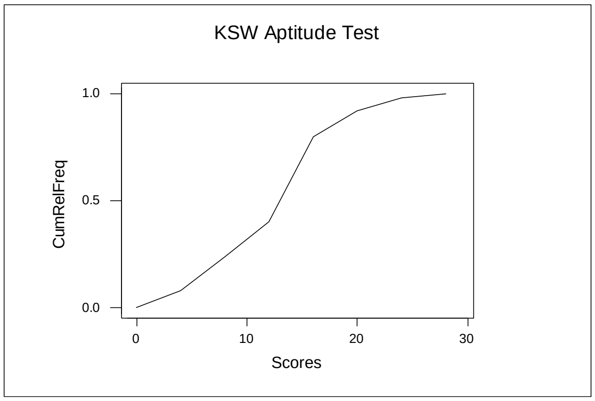
b.



2.80 a. (relative frequencies taken from ex. 2.41)

Class limits	Cum.Rel.Freq.
0 - 4	0.08
4 - 8	0.24
8 - 12	0.40
12 - 16	0.80
16 - 20	0.92
20 - 24	0.98
24 - 28	1.00

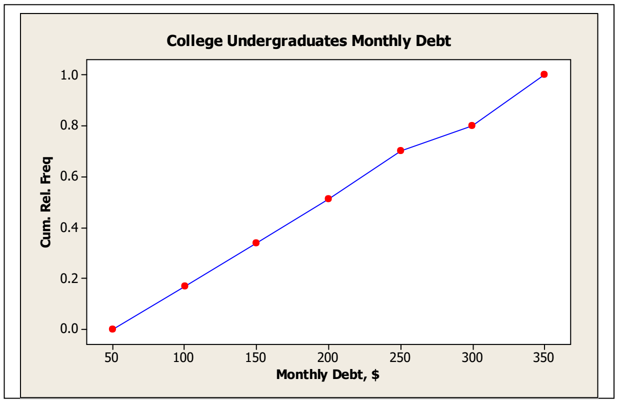
b.



2.81 a.

Class Boundaries	Cum.Rel.Freq.
less than 100	0.17
100 - 150	0.34
150 - 200	0.51
200 - 250	0.70
250 - 300	0.80
300 or more	1.00

b.



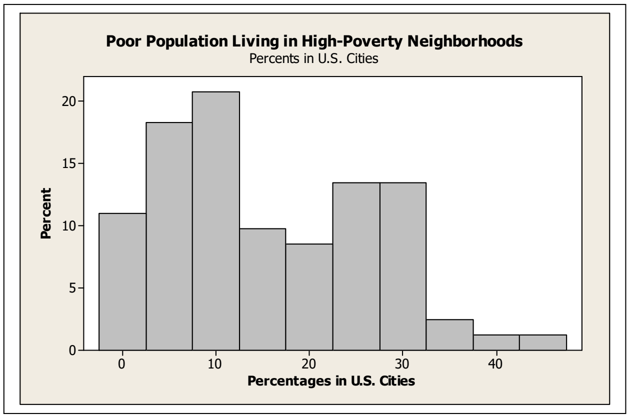
2.82 a.

Class Boundar.	Class Midpoints	Freq.
-2.5 – 2.5	0	9
2.5 – 7.5	5	15
7.5 – 12.5	10	17
12.5 – 17.5	15	8
17.5 – 22.5	20	7
22.5 – 27.5	25	11
27.5 – 32.5	30	11
32.5 – 37.5	35	2
37.5 – 42.5	40	1
42.5 – 47.5	45	1

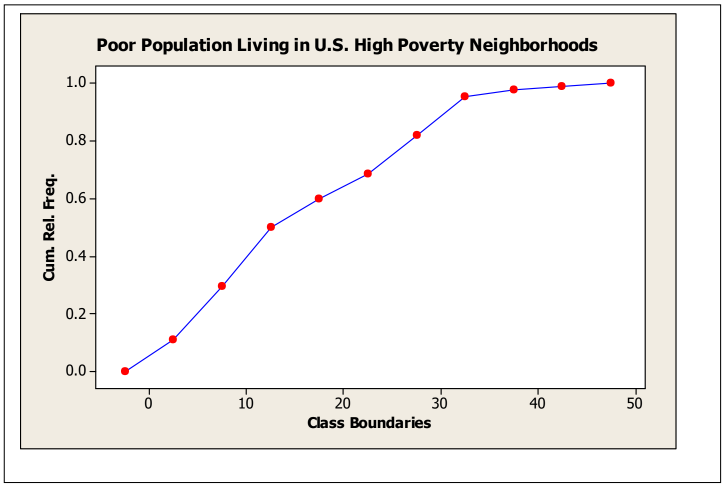
b. and d.

Class Midpoints	Rel.Freq.	Cum. Rel. Freq.
0	0.110	0.110
5	0.183	0.293
10	0.207	0.500
15	0.098	0.598
20	0.085	0.683
25	0.134	0.817
30	0.134	0.951
35	0.024	0.975
40	0.012	0.987
45	0.012	0.999 ≈ 1.000

c.



e.



2.83 a. Largest spread: C   Smallest spread: B
b. Histograms A and D

OBJECTIVE 2.3 EXTRA PROBLEMS

2.84 The data resulting from a quantitative variable are numbers with which arithmetic (addition, subtraction, etc.) can be performed. The data resulting from a qualitative variable are 'category' type values, such as color. It is not possible to add three colors together, and divide by 3, to obtain a value for the mean color.

2.85 a. 9   b. value = 0

2.86 a. $\bar{x} = \sum x/n = (16+132+124+191+183+299)/6 = 945/6 = \underline{157.5}$

Note: For a length of highway with 6 interchanges, there are only 5 sections of highway between them.

b. $\bar{x} = \sum x/n = (16+132+124+191+183+299)/10 = 945/10 = \underline{94.5}$

2.87 a. $\bar{x} = \sum x/n = 123/36 = \underline{3.4}$
b. $\bar{x} = \sum x/n = 161/31 = \underline{5.2}$
c. $\bar{x} = \sum x/n = 217/39 = \underline{5.6}$
d. $\bar{x} = \sum x/n = 252/43 = \underline{5.9}$
e. $\bar{x} = \sum x/n = (123+161+252+217)/152 = 753/152 = \underline{4.95}$
where $152 = (37+32+44+40) - 1 = 153 - 1$
f. $\bar{x} = \sum x/n = (3.4+5.2+5.9+5.6)/4 = 20.1/4 = \underline{5.025}$
g. Answers will vary. There are a different number of interchanges in each state, so the states do not weigh equally in finding the mean.

2.88 a. mean – larger, median - same
b. mean – smaller, median – same;
c. median

2.89 mode = $\underline{2}$

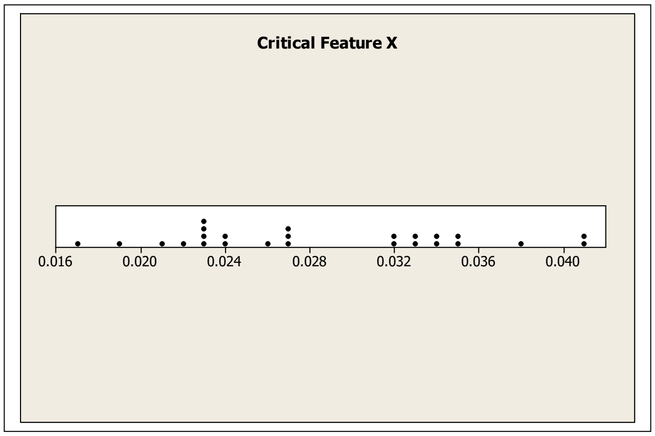
2.90 midrange = $(L+H)/2 = (100+350)/2 = 450/2 = \underline{225}$

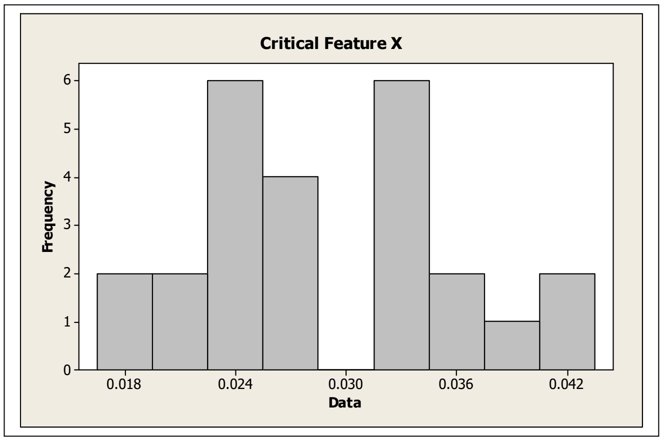
2.91 a. $\bar{x} = \sum x/n = (9+6+7+9+10+8)/6 = 49/6 = 8.166 = \underline{8.2}$
Ranked data: 6, 7, 8, 9, 9, 10
$d(\tilde{x}) = (n+1)/2 = (6+1)/2 = 3.5\text{th}; \quad \tilde{x} = \underline{8.5}$
mode = $\underline{9}$
midrange = $(L+H)/2 = (6+10)/2 = 16/2 = \underline{8.0}$
b. Answers will vary. All show centers.

2.92 a. $\bar{x} = \sum x/n = (3+5+6+7+7+8)/6 = 36/6 = \underline{6.0}$
b. $d(\tilde{x}) = (n+1)/2 = (6+1)/2 = 3.5\text{th}; \quad \tilde{x} = (6+7)/2 = \underline{6.5}$
c. mode = $\underline{7}$
d. midrange = $(H+L)/2 = (8+3)/2 = 11/2 = \underline{5.5}$

2.93 {4, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 11}
a. $\bar{x} = \sum x/n = 104/15 = 6.9333 = \underline{6.9}$
b. $d(\tilde{x}) = (n+1)/2 = (15+1)/2 = 8\text{th}; \quad \tilde{x} = \underline{7}$
c. mode = $\underline{7}$
d. midrange = $(H+L)/2 = (11+4)/2 = \underline{7.5}$

2.94 a.





b. $\bar{x} = \sum x/n = 0.714/25 = 0.02856 = \underline{0.0286}$

c. $d(\tilde{x}) = (n+1)/2 = (25+1)/2 = 13\text{th};\quad \tilde{x} = \underline{0.027}$

d. midrange = $(L+H)/2 = (0.017+0.041)/2 = \underline{0.029}$

e. mode = $\underline{0.023}$

f. Bimodal distribution. The central tendency statistics all fall around the 0.030 center split of the data. This occurs because the two most populous classes are separated by two classes. This suggests that two populations are being sampled.

Formatted: Indent: Hanging: 0.25"

g. Answers will vary but problem could be that there are two populations.

2.95 a. $\bar{x} = \sum x/n = 2205.89/31 = 71.158 = \underline{71.16\%}$

b. $d(\tilde{x}) = (n+1)/2 = (31+1)/2 = 16\text{th};\quad \tilde{x} = \underline{72.66\%}$

c.

```
Stem-and-leaf of Percentage  N = 31
Leaf Unit = 1.0


 2    5  89
 5    6  244
 10   6  57999
(17)  7  00111223333444444
 4    7  578
```

d. The left-tail (or smaller value) tail causes the mean to be less in value than the median. The 2 data, 58.60 and 59.25, are separate from the rest of the pack and have a reducing effect on the mean value, but do not effect the value of the median.

2.96 a. The mean is a very useful statistic in situations where the "total value" of the data is also meaningful. The mean is a very misleading statistic when the distribution is skewed. A good example of that is a distribution of salaries for an organization.

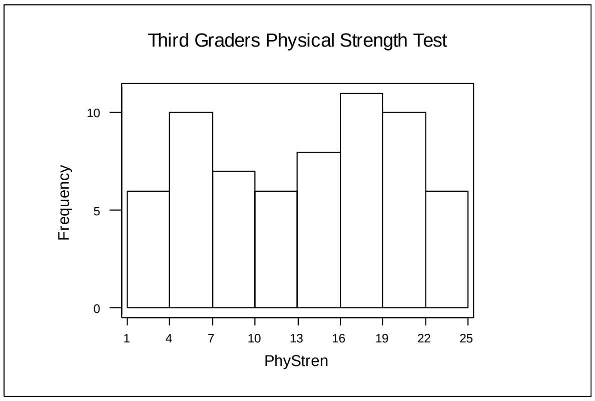
b. The median is useful when representing a skewed distribution, such as salaries. The median is a misleading statistic when the distribution is bimodal.

2.97 a.

```
               Third Graders at Roth Elementary School
                      :
          . . : .     :     .   . : . : : : . . .
      . : : : : :   . : : . : . : : : : : : : : : . :
    +---------+---------+---------+---------+---------+--PhyStren
   0.0       5.0      10.0      15.0      20.0      25.0
```

b. mode = 9

c.



c.

Class boundaries	frequency
1 - 4	6
4 - 7	10
7 - 10	7
10 - 13	6
13 - 16	8
16 - 19	11
19 - 22	10
22 - 25	6
	64

d. The distribution appears to be bimodal. Modal classes are 4-7 and 16-19.

e. Dotplot shows mode to be 9, which is in the 7-10 class, while the histogram shows the two modal classes to be 4-7 and 16-19. The mode is not in either modal class.

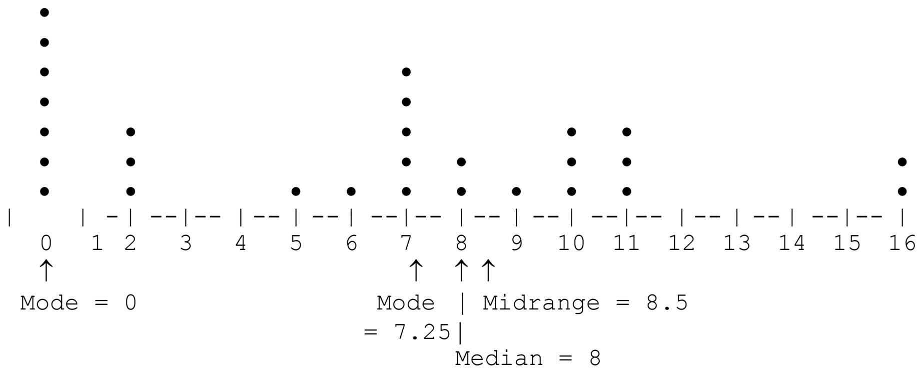
f. No. In an ungrouped distribution, there is only one numerical value per class.

g. The mode is simply the single data value that occurs most often, while a modal class results from data tending to bunch up and form a cluster of data values, not necessarily all of one value.

2.98 a.

	Mean	Median	Mode	Midrange
Calories	91.073	85	50	120
Fat (g)	6.143	7	0	8
Sodium (mg)	530.357	535	490	560

b. The dotplot is shown below:



c. (49)(85) = 4165 calories;(49)(7) = 343 g of fat
(49)(535) = 26,215 mg of sodium. Yes, he would have exceeded it by nearly 11 times the recommended daily intake. Also, most of the hot dogs in the study are likely not to contain as much sodium as the ones the contestant consumed, because they are targeted for people who may be restricting their intake of at least one of the

three nutritional categories.

2.99 a. and b.

	Runs at Home	Runs Away	Difference
Mean	4.828	4.797	0.031
Median	4.870	4.860	0.01
Maximum	6.380	5.570	0.81
Minimum	3.630	3.430	0.20
Midrange	5.005	4.500	0.505

c. All five measures of central tendency are greater for runs scored at home than for the number of runs scored away. In conclusion, they score more runs at home.

2.100 a. Negative values mean that annual percentage rate actually decreased.

Positive values mean that annual percentage rate actually increased.

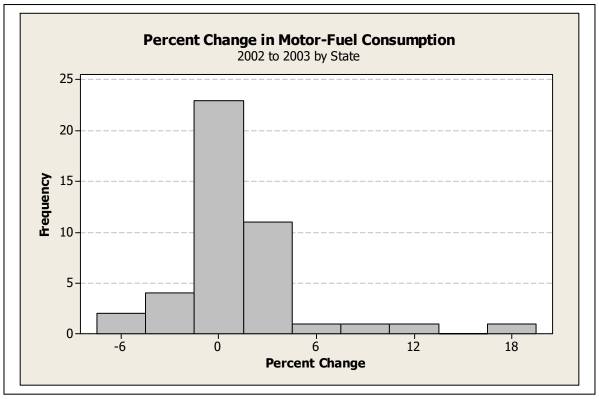
Large values mean the rate actually changed by quite a lot.

Small values mean the rate actually changed very little.

b. Answers will vary. Student responses should indicate an expectation that the distribution will be centered at about 0, and skewed to the right. The student should justify this by identifying the clustering of data around 0, as well as several other values outside the cluster, located to the right.

c. Zero, meaning no change.

d.



e. $\bar{x} = \sum x/n$ = 47.9/44 = 1.08864 = 1.089

f. Part e is the mean of the 44 state values. It is not the mean for the whole country.

Formatted: Indent: Hanging: 0.31"

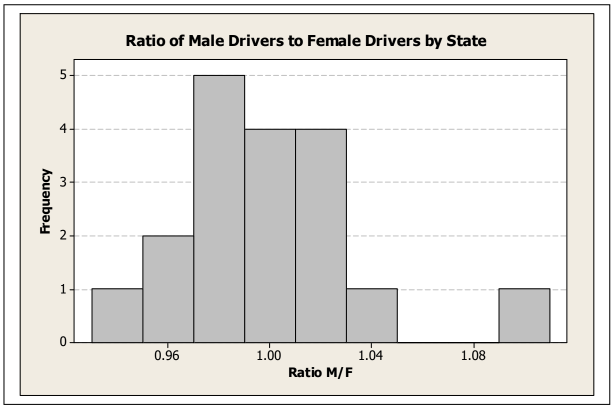
2.101 a. A quick look at the data listed suggests, yes, the number of female licensed drivers is larger for most of the states states listed.

b.

```
Ratio M/F
   0.98520   0.95703   0.96727   1.01151   0.99234   1.10043
   0.93043   1.01435   1.01596   0.98231   0.98466   0.97655
   0.99562   1.03506   0.99592   0.98321   1.01944   0.99244
```

c. Near 1.0 means there is little difference.
Greater than 1.0 means there are more male drivers.
Less than 1.0 means there are more female drivers.

d.



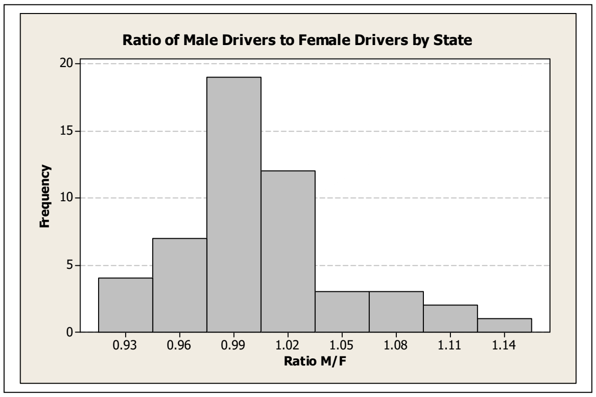
e. The distribution of “Ratio M/F” is mounded and appears to be somewhat normal except for the one value that is considerably larger than the others, thus making the data skewed to the right.

f. $\overline{x} = \sum x/n = 17.9397/18 = 0.996651 = 0.997$

g. The value to the extreme right means that state has considerably more male drivers than female, approximately 10% more. The value to the extreme left means that state has fewer male drivers than female drivers and since the value is approximately 0.94, there are 6% fewer male drivers.

h. Answers will vary.

i.



$\overline{x}$ = $\sum x/n$ = = 51.1858/51 = 1.00364 = 1.004

j. The results are very similar.

k. Answers will vary.

2.102 Many different answers are possible.

a. mean; total number of cars $\sum x = n \cdot \overline{x}$, where n = number of apartments.

b. "1.9" cannot be the median or midrange because the data values (number of vehicles) would all be whole numbers; therefore, the median or midrange would either be a whole number or a '0.5' number. "1.9" cannot be the mode because all the data values would be whole numbers.

c. (1.9)(256) = 486.4; (0.90)(486.4) = 437.76 = 438 spaces

2.103 a. "Taxes per capita" means amount of taxes paid per person, while "percent of personal income" is the percent of personal income that is paid in taxes. A person in North Dakota pays a lesser amount of taxes per person, but apparently they make so much less that the percentage of their income that goes for taxes is actually larger than in New Hampshire. $1283 is 4.8% of $26,729 while $1478 is 4.4% of $33,590.

b. The only "average" that can be found is the midrange.
Midrange = ($2748 + $1283)/2 = $2015.50

c. Midrange = (9.6 + 4.4)/2 = 7.0%

d. The only "average" that is defined by only the extreme values of the variable is the midrange.

2.104 a. "Win bet" means the class mean is greater than or equal to 74.
Mean = 74 is equivalent to sum of all grades is 15×74 or 1110.

The 14 already graded total 14×73.5 or 1029. Therefore you must get 1110 – 1029 or 81 on your exam.

b. In order to "not do community service," the class mean must be greater than or equal to 72. That requires a total of 15×72 or 1080. The class already has 1029; therefore you need 1080 – 1029 or 51 on the exam in order for the class to avoid community service.

2.105 Many different answers are possible.

a. $\sum x$ needs to be 500; therefore, need any three numbers that total 330.

Formatted: Indent: Hanging: 0.31"

100, 100, 130 [70, 100]
$\bar{x} = \sum x/n = 500/5 = 100$, ck

b. Need two numbers smaller than 70 and one larger.
_, __, 70,___, 100: 50, 60, 80 [70, 100]
$d(\tilde{x}) = (n+1)/2 = (5+1)/2 = 3$rd; $\tilde{x} = 70$, ck

c. Need multiple 87s. 87, 87, 87 [70, 100]
mode = 87, ck

d. Need any two numbers that total 140 for the extreme values where one is 100 or larger. _, _, _, 70, 100
40, 50, 60 [70, 100]
midrange = $(L+H)/2 = (40+100)/2 = 70$, ck

e. Need two numbers smaller than 70 and one larger than 70 so that their total is 330. _, _, 70, _, 100;
60, 60, 210 [70, 100]
$\bar{x} = \sum x/n = 500/5 = 100$, ck
$d(\tilde{x}) = (n+1)/2 = (5+1)/2 = 3$rd; $\tilde{x} = 70$, ck

f. Need two numbers of 87 and a third number large enough so that the total of all five is 500.
87, 87, 156 [70, 100]
$\bar{x} = \sum x/n = 500/5 = 100$, ck; mode = 87, ck

g. Mean equal to 100 requires the five data to total 500 and the midrange of 70 requires the total of *L* and *H* to be 140: 40, _, 70, _, 100; that is a sum of 210, meaning the other two data must total 290. One of the last two numbers must be larger than 145, which would then become *H* and change the midrange. Impossible.

h. There must be two 87s in order to have a mode, and there can only be two data larger than 70 in order for 70 to be the median. _, 70, 87, 87, 100; Impossible.

2.106 Answers will vary. One way is to look at the plots and decide which has the smallest mean and which the largest mean, and match those with the values. Then take care of the other two.

OBJECTIVE 2.4 EXTRA PROBLEMS

2.107 a. range = $H - L$ = $2748 - $1283 = $1465

b. range = $H - L$ = 9.6% - 4.4% = 5.2%

2.108 a. First find mean, $\bar{x} = \sum x/n = 25/5 = 5$

	x	$x-\bar{x}$	$(x-\bar{x})^2$
	1	-4	16
	3	-2	4
	5	0	0
	6	1	1
	10	5	25
$\sum$	25	0	46

$s^2 = \sum(x-\bar{x})^2/(n-1)$

$= 46/4 = 11.5$

b.

x	x^2
1	1
3	9
5	25
6	36
10	100
25	171

$SS(x) = \sum x^2 - ((\sum x)^2/n)$

$= 171 - ((25)^2/5)$
$= 171 - 125 = 46$

$s^2 = SS(x)/(n-1) = 46/4 = 11.5$

c. Both results are the same.

2.109 a. range = $H - L$ = 8 - 3 = 5
b. First find mean, $\bar{x} = \sum x/n = 36/6 = 6$

	x	$x-\bar{x}$	$(x-\bar{x})^2$
	3	-3	9
	5	-1	1
	6	0	0
	7	1	1
	7	1	1
	8	2	4
$\sum$	36	0	16

$s^2 = \sum(x-\bar{x})^2/(n-1)$

$= 16/5 = 3.2$

c. s = = = 1.789 = 1.8

2.110 a. First find mean, $\bar{x} = \sum x/n = 72/10 = 7.2$

	x	$x-\bar{x}$	$(x-\bar{x})^2$
	3	-4.2	17.64
	5	-2.2	4.84
	5	-2.2	4.84
	6	-1.2	1.44
	7	-0.2	0.04
	7	-0.2	0.04
	7	-0.2	0.04
	9	1.8	3.24
	10	2.8	7.84
	13	5.8	33.64
$\sum$	72	0	73.60

$s^2 = \sum(x-\bar{x})^2/(n-1)$

$= 73.60/9$

$= 8.1778 = 8.2$

b.

x	x^2
3	9
5	25
5	25
6	36
7	49
7	49
7	49
9	81
10	100
13	169
$\sum$ 72	592

$SS(x) = \sum x^2 - ((\sum x)^2/n)$

$= 592 - ((72)^2/10)$

$= 592 - 518.4 = 73.6$

$s^2 = SS(x)/(n-1)$

$= 73.6/9 = 8.1778 = \underline{8.2}$

c. $s = \ = \ = 2.8597 = \underline{2.9}$

2.111 $n = 10$, $\sum x = 402$, $\sum x^2 = 16704$

a. Range = $H - L = 49 - 28 = 21$

b. $s^2 = 60.4$ $SS(x) = \Sigma x^2 - ((\Sigma x)^2/n) = 16704 - ((402)^2/10) = 543.6$
$s^2 = SS(x)/(n-1) = 543.6 / 9 = 60.4$

c. $s = 7.7717 = 7.8$ $s = \sqrt{60.4} = 7.7717 = 7.8$

2.112 a. Original data: $n = 6$, $\sum x = 37{,}116$, $\sum x^2 = 229{,}710{,}344$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 229{,}710{,}344 - (37{,}116^2/6) = 110{,}768.0$

$s^2 = SS(x)/(n-1) = 110{,}768.0/5 = \underline{22{,}153.6}$

b. Smaller numbers: $n = 6$, $\sum x = 1{,}116$, $\sum x^2 = 318{,}344$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 318{,}344 - (1{,}116^2/6) = 110{,}768.0$

$s^2 = SS(x)/(n-1) = 110{,}768.0/5 = \underline{22{,}153.6}$

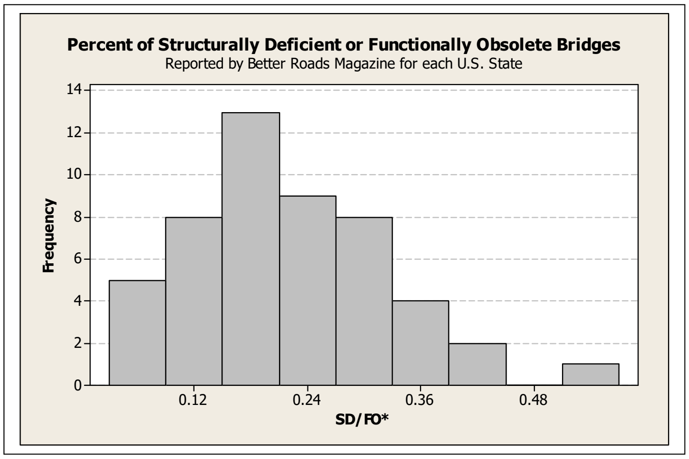
Both sets of data have the same variance.

2.113 $n = 12$, $\sum x = 11250$, $\sum x^2 = 11533488$

a. $\bar{x} = 11250/12 = 937.5$

b. $s^2 = 89692.0909$; $s = 299.5$

2.114 a.



b. The variable “%SD/FO” appears to have a skewed right distribution.

c. $\bar{x}$ = 0.2176 $\bar{x}$ = Σ*x*/*n* = 10.88/50 = 0.2176

d. $\tilde{x}$ = 0.20 *d*($\tilde{x}$) = (*n*+1)/2 = (50+1)/2 = 25.5th; $\tilde{x}$ = ([25th] + [26th])/2= 0.20

e. range = *H* - *L* = 0.55 – 0.03 = 0.52

f. *s* = 0.1038 *SS*(*x*) = 0.528; *s* = √[*SS*(*x*)/(*n* -1)]= √[*SS*(*x*)/(49)]= *s* = 0.1038

2.115 a. Range = *H* – *L* = 80.3 – 58.6 = 21.7; *s* = 5.242

b.

```
Stem-and-leaf of Percentage  N = 31
Leaf Unit = 1.0

 2    5  89       - - - - - - - - - -
 5    6  244
 10   6  57999    - - - - - -  s    range
(17)  7  00111223333444444 - -
 4    7  578
 1    8  0        - - - - - - - - - -
```

c. The distribution of the data is skewed to the left. To accommodate the smaller values, the standard deviation is a bit high compared to the range. For a normal distribution, approximately 6 standard deviations equal the range.

2.116 a. Since *s* = 0, all data must be the same value.
100, 100, 100, 100, [100]

Many different answers are possible for parts b, c, and d. One possible answer is shown for each.

b. 99, 99.5, 100.5, 100 [100]; $s = 0.57$

c. 107, 95, 94, 108, [100]; $s = 6.53$

d. 75, 78, 123, 124, [100]; $s = 23.53$

2.117 Different answers will result, depending on the relative size of the data making up the two sets.

Larger, if the data in one set are larger in value than the data values of the first set; the combined set is more dispersed.

Set I: {4,6,10,14,16} and Set II: {14,16,20,24,26}

Set I: $n = 5$, $\sum x = 50$, $\sum x^2 = 604$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 604 - (50^2/5) = 104$

$s^2 = SS(x)/(n-1) = 104/4 = 26$

$s = \ = \ = 5.099 = 5.1$

Set II: $n = 5$, $\sum x = 100$, $\sum x^2 = 2104$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 2104 - (100^2/5) = 104$

$s^2 = SS(x)/(n-1) = 104/4 = 26$

$s = \ = \ = 5.099 = 5.1$

Together, Set I and Set II;

$n = 10$, $\sum x = 150$, $\sum x^2 = 2708$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 2708 - (150^2/10) = 458$

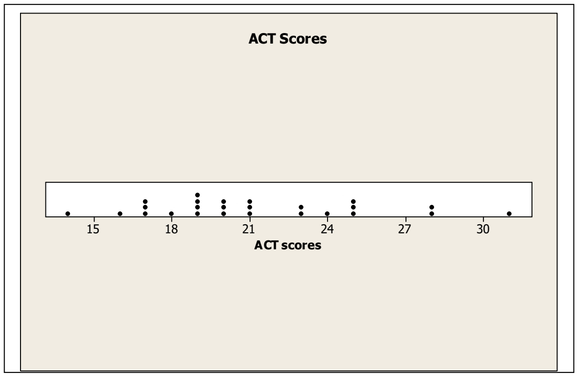
$s^2 = SS(x)/(n-1) = 458/9 = 50.88888$

$s = \ = \ = 7.133644 = 7.13$

2.118 Answers will vary. One method could utilize the spread of standard deviation first to narrow down the possibilities, then determine the center based on the means given.

OBJECTIVE 2.5 EXTRA PROBLEMS

2.119 a.



Ranked data:

14 16 17 17 17 18 19 19 19 19 20 20 20
21 21 21 23 23 24 25 25 25 28 28 31

b. 19th from the *L*, 7th from the *H*

c. $nk/100 = (25)(5)/100 = 1.25$;
therefore $d(P_5) = 2^{nd}$ from *L*
$P_5 = \underline{16}$

$nk/100 = (25)(10)/100 = 2.5$;
therefore $d(P_{10})$ = 3rd from *L*
$P_{10} = \underline{17}$

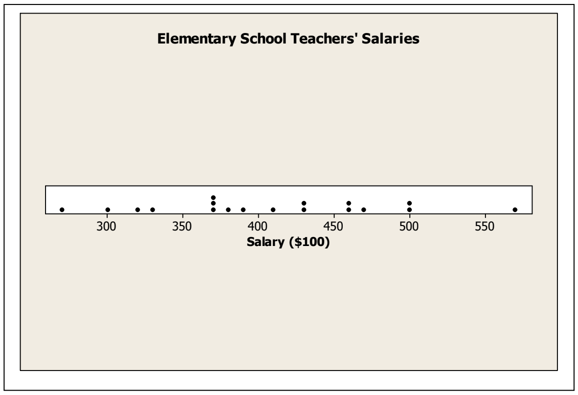
$nk/100 = (25)(20)/100 = 5.0$;
therefore $d(P_{20}) = 5.5^{th}$ from *L*
$P_{20} = (17+18)/2 = \underline{17.5}$

d. $nk/100 = (25)(1)/100 = 0.25$;
therefore $d(P_{99}) = 1^{st}$ from *H*
$P_{99} = \underline{31}$

$nk/100 = (25)(10)/100 = 2.5$;
therefore $d(P_{90})$ = 3rd from *H*
$P_{90} = \underline{28}$

$nk/100 = (25)(20)/100 = 5.0$;
therefore $d(P_{80}) = 5.5^{th}$ from *H*
$P_{80} = (25+25)/2 = \underline{25}$

2.120 a.



Ranked data:

269 295 317 326 367 367 371 376 391 413 433 434 455
458 471 495 501 574

b. 2nd from the *L*, 17th from the *H*

c. $nk/100 = (18)(25)/100 = 4.5$;
therefore $d(Q_1)$ = 5th from *L*
Q_1 = 367 = $36,700

d. $nk/100 = (18)(75)/100 = 13.5$;
therefore $d(Q_3)$ = 14th from *L*
Q_3 = 458 = $45,800

2.121 Ranked data:

3.73 5.85 6.38 8.68 9.48 9.67 11.74
13.43 15.24 21.31 29.64 52.71 64.19 69.18 101.68

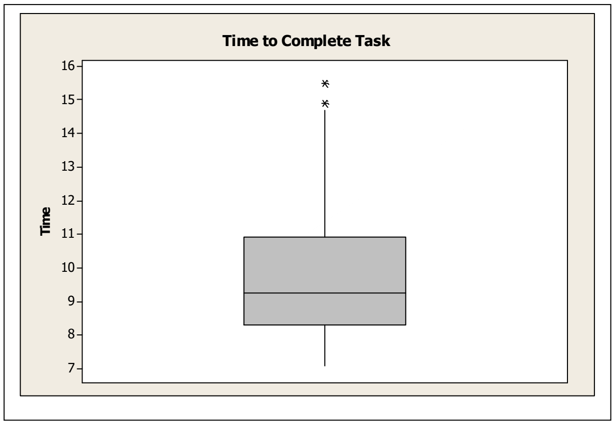
a. $nk/100 = (15)(25)/100 = 3.75$;
therefore $d(Q_1)$ = 4th from *L*
$Q_1 = 8.68$

$nk/100 = (15)(75)/100 = 11.25$;
therefore $d(Q_3)$ = 12th from *L*
$Q_3 = 52.71$

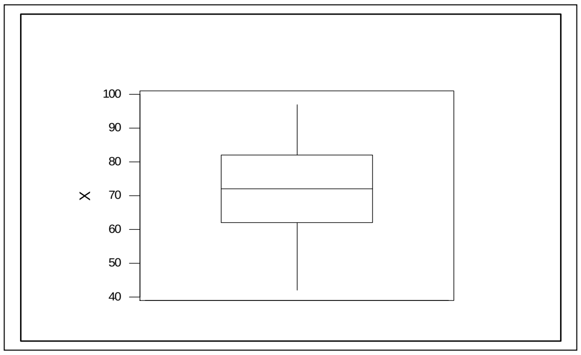
b. midquartile = $(Q_1 + Q_3)/2 = (8.68 + 52.71)/2 = 30.695$

2.122 a. $nk/100 = (40)(25)/100 = 10$; therefore $d(Q_1)$ = 10.5th
$Q_1 = (8.3+8.3)/2 = 8.3$
b. d(median) = $(40+1)/2$ = 20.5th,
Q_2 = median = $(9.1+9.4)/2 = 9.25$
c. $nk/100 = (40)(75)/100 = 30$; therefore $d(Q_1)$ = 30.5th
$Q_3 = (10.7+11.0)/2 = 10.85$
d. $nk/100 = (40)(95)/100 = 38$; therefore $d(P_{95})$ = 38.5th
$P_{95} = (14.7+14.9)/2 = 14.8 = 14.8$
e. 5-number summary: 7.1, 8.3, 9.25, 10.85, 15.5

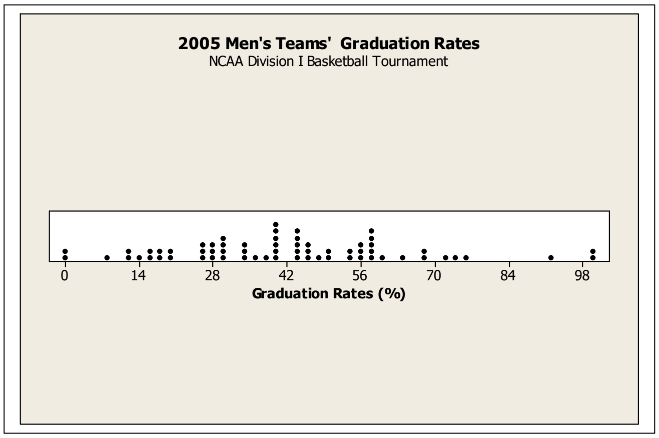
f.


Time to Complete Task
Time
16
15
14
13
12
11
10
9
8
7

2.123


X
100
90
80
70
60
50
40

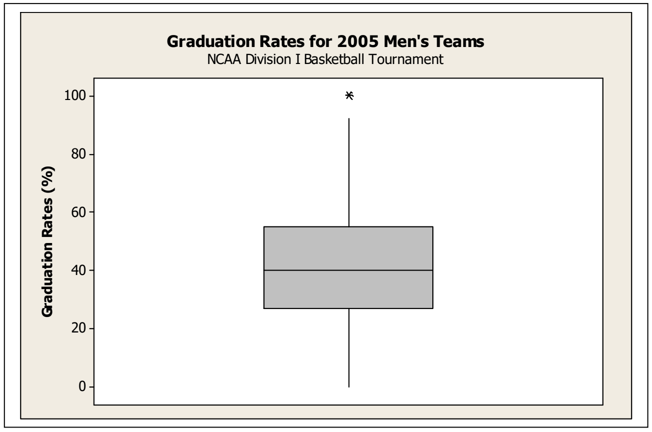
2.124 a.


2005 Men's Teams' Graduation Rates
NCAA Division I Basketball Tournament
0
14
28
42
56
70
84
98
Graduation Rates (%)

b.

```
Stem-and-leaf of Graduation Rates (%)  N = 64
Leaf Unit = 1.0


 3    0   008
 11   1   11455779
 20   2   055577799
 27   3   0033368
(15)  4   000000334445557
 22   5   003455577888
 10   6   0477
 6    7   135
 3    8
 3    9   2
 2    10  00
```

c. 5-number summary: 0, 27, 40, 55, 100



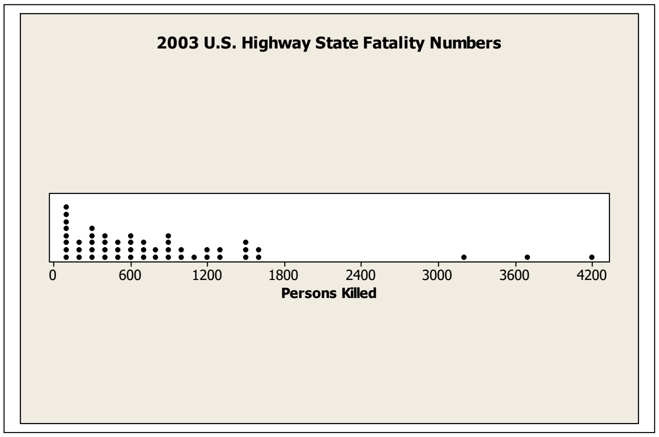
d. $nk/100$ = (64)(5)/100 = 3.2; therefore $d(P_5)$ = 4th
P_5 = 11

$nk/100$ = (64)(95)/100 = 60.8; therefore $d(P_{95})$ = 61st
P_{95} = 75

e. Skewed to the right, centered around 40% graduation rate. The two 100% values are distinctly separate from the other values.

f. The 92% and the two 100% are quite different from the rest. The next closest rate is a 75% graduation rate.

2.125 a.



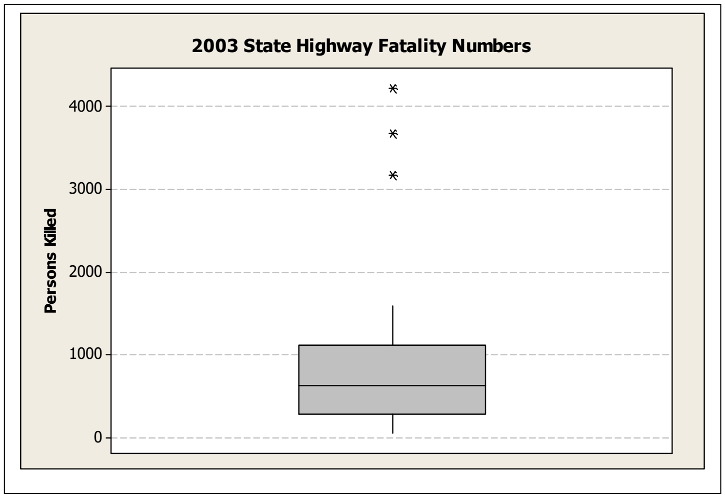
b.

```
Stem-and-leaf of Persons Killed
N = 51   Leaf Unit = 100


 22   0  0001111112222223334444
(15)  0  566666678888999
 14   1  01122244
 6    1  556
 3    2
 3    2
 3    3  1
 2    3  6
 1    4  2
```

Ranked data:

```
67     69     95     104    105   127    135    142    165    203    207    262
293    293    294    309    368    394   439    441    462    471    512     600
627    632    649    657    668    747    834   848    871    894    928     943
968   1001   1120   1193   1232   1277   1283   1453  1491   1531   1577   1603
3169   3675  4215
```

c. 5-number summary: 67, 293, 632, 1120, 4215



Values that are distinctly different than the balance of a sample will show up by forming a separate cluster on the graph. They may be kept or they may be trimmed from the sample. Before such a decision is made, you should look for an explanation as to why these are so different from the rest. This information should allow you to make the appropriate decision before continuing.

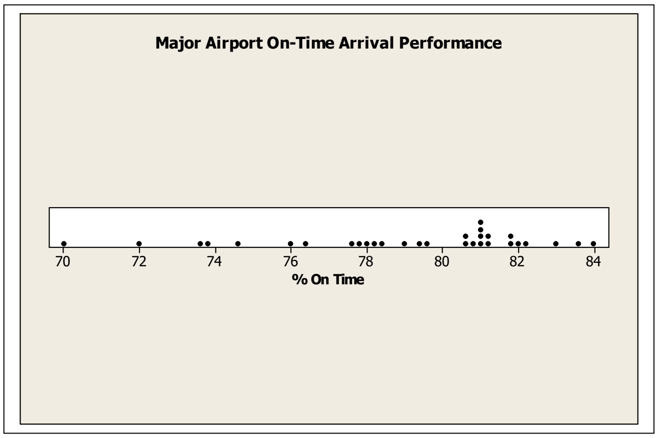
d. $nk/100 = (51)(10)/100 = 5.1$; therefore $d(P_{10}) = 6^{th}$
$P_{10} = 127$

$nk/100 = (51)(90)/100 = 45.9$; therefore $d(P_{90}) = 46^{st}$
$P_{90} = 1531$

e. The distribution is skewed to the right with the 3 states with the highest rate of fatalities distinctly different from the rest.

f. Three of the data values are very large relative to the rest of the sample and therefore distort the sample statistics. Also, since the states and DC are of different sizes and population densities, the relative safety of all of our highways would be more accurately demonstrated with fatalities/1000 drivers. This would allow you to make a more fair comparison.

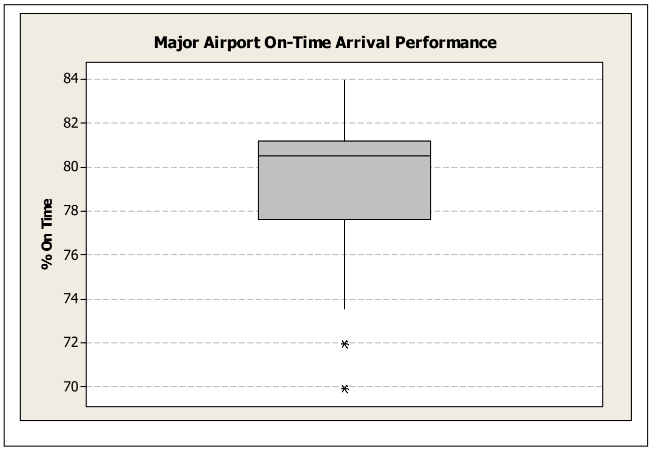
2.126 a.



b.

```
Stem-and-leaf of % On Time  N = 31
Leaf Unit = 0.10


 1   69  9
 1   70
 2   71  9
 2   72
 4   73  58
 5   74  6
 5   75
 7   76  04
 10  77  689
 13  78  239
 15  79  45
(5)  80  55899
 11  81  0012779
 4   82  29
 2   83  69
```

c. 5-number summary: 69.96, 77.62, 80.50, 81.20, 83.93



Ranked Data:

```
69.96   71.96   73.55   73.85   74.60   76.00   76.45   77.62
77.82   77.94   78.26   78.38   78.92   79.42   79.50   80.50
80.51   80.85   80.90   80.91   81.04   81.08   81.10   81.20
81.71   81.73   81.90   82.28   82.99   83.60   83.93
```

d. $nk/100 = (31)(10)/100 = 3.1$; therefore $d(P_{10}) =$ 4th
$P_{10} = 73.85$

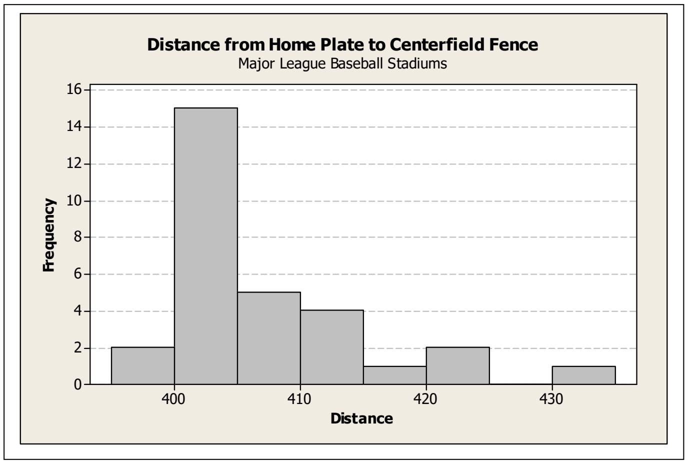
$nk/100 = (31)(20)/100 = 6.2$; therefore $d(P_{20}) =$ 7th
$P_{20} = 76.45$

e. The distribution is skewed to the left with the 2 airports having the lowest rate of on-time performance quite separated from the others.

f. Travelers are more interested in being on time and want the best on-time performance rate.

g. The airports with the lowest percentages are quite different from the rest of the airports. The two airports are EWR (Newark, NJ) and ORD (Chicago-O'Hare). Two airports have percentages that are quite different from the other the airports in the data set. The two airports are EWR (Newark, NJ) and ORD (Chicago-O'Hare). Students should mention that both values are below the lower whisker if they reference a computer-generated box-and-whisker plot. Alternately, students may mention that the bottom whisker is much longer than the other three quartiles, if their box-and-whisker plot is made using the 5-number summary. In either case, these are visual indicators that the data set is skewed.

2.127 a.



b. Q_1 = 400 Q_3 = 410 Interquartile range is from 400 to 410, therefore, midrange = 10.

c. One field is considerably larger at 435 ft.
d. No, all but 6 are between 400 and 410 ft.

2.128 The distribution needs to be symmetric.

2.129 a. n = 29, $\sum x$ = 157.99, $\sum x^2$ = 862.0855

$\bar{x}$ = $\sum x/n$ = 157.99/29 = 5.4479 = $\underline{5.448}$

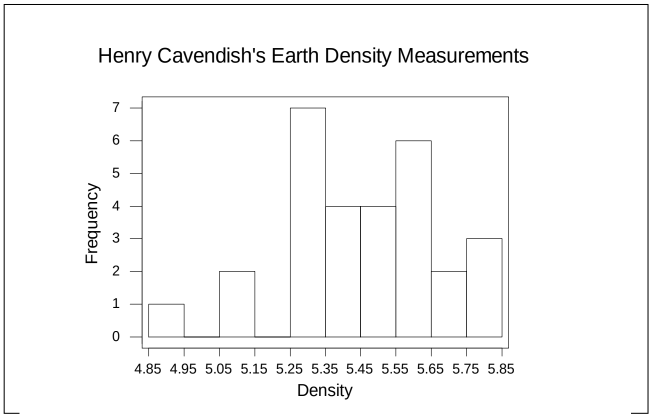
$d(\tilde{x})$ = $(n+1)/2$ = 15th; $\tilde{x}$ = $\underline{5.46}$

$SS(x)$ = $\sum x^2 - ((\sum x)^2/n)$
= 862.0855 - (157.99²/29) = 1.36688

s^2 = $SS(x)/(n-1)$ = 1.36688/28 = 0.048817

s = = $\sqrt{0.0488}$ = $\underline{0.2209}$

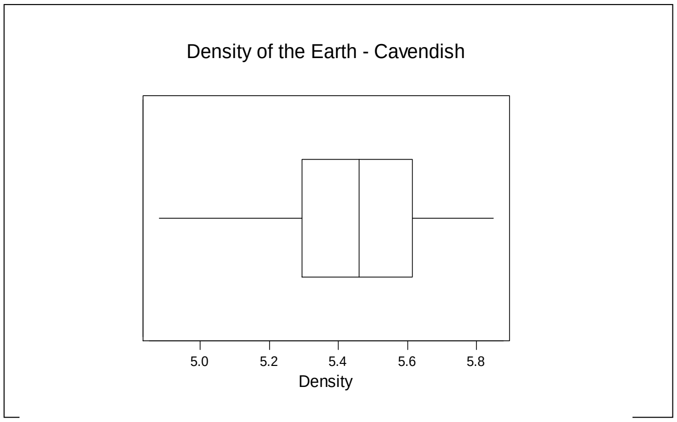
b.



The mean and median are almost equal, which is true in approximately symmetrical distributions. The standard deviation is about 1/6 of the range.

c. 5-number summary: 4.88, 5.295, 5.46, 5.615, 5.85

d.



In the box-and-whisker plot, the left and right sides correspond to the first and third quartiles, which are the second and fourth numbers in the 5-number summary (5.295 and 5.615). The vertical line in the box corresponds to the median, the third number in the 5-number summary (5.46). The left tip of the lower whisker corresponds to the low value (4.88) and the right tip of the upper whisker corresponds to the high value (5.85).

e. approximately normal

f. $\bar{x} \pm 2s = 5.448 \pm 2(0.2213) = \underline{5.0054 \text{ to } 5.8906}$

It is close: 28/29 = 96.6%.

z is a measure of position. It gives the number of standard deviations a piece of data is from the mean. It will be positive if x is to the right of the mean (larger than the mean) and negative if x is to the left of the mean (smaller than the mean). Keep 2 decimal places. (hundredths)

z = (x - mean)/st. dev. $z = (x - \bar{x})/s$

2.130 z = (x - mean)/st.dev.

for x = 92, z = (92 - 72)/12 = 1.67
for x = 63, z = (63 - 72)/12 = -0.75

2.131 z = (x - mean)/st.dev.

a. for x = 54, z = (54 - 50)/4.0 = 1.0
b. for x = 50, z = (50 - 50)/4.0 = 0.0
c. for x = 59, z = (59 - 50)/4.0 = 2.25
d. for x = 45, z = (45 - 50)/4.0 = -1.25

2.132 If z = (x - mean)/st.dev; then x = (z)(st.dev) + mean

for z = 1.8, x = (1.8)(100) + 500 = 680

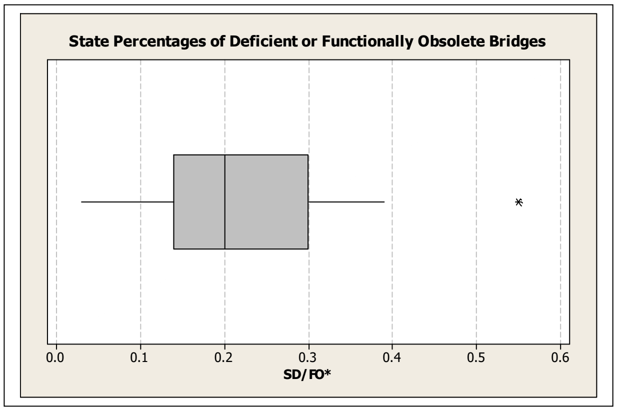
2.133 a. 152 is one and one-half standard deviations above the mean.
b. The score is 2.1 standard deviations below the mean.
c. The standard score is the number of standard deviations from the mean.

2.134 a. Ranked data:

0.03 0.05 0.05 0.06 0.07 0.10 0.13 0.14 0.14 0.14 0.14 0.14
0.14 0.16 0.16 0.16 0.16 0.17 0.17 0.17 0.19 0.20 0.20 0.20
0.20 0.20 0.21 0.21 0.21 0.22 0.22 0.23 0.23 0.24 0.25 0.29
0.29 0.30 0.30 0.31 0.31 0.32 0.32 0.34 0.35 0.36 0.37 0.39
0.39 0.55

b.

Lowest Value	First Quartile	Median	Third Quartile	Highest Value
0.03	0.14	0.20	0.30	0.55



c. Midquartile = (0.14 + 0.30) / 2 = 0.22
Interquartile range = 0.30 – 0.14 = 0.16

d.

	California	Hawaii	Nebraska	Oklahoma	Rhode Is.
z-score	-0.75	1.66	-1.42	0.22	3.20

2.135 English Math
a. (30-20.4)/5.9 = 1.63 (30-20.7)/5.0 = 1.86
b. (23-20.4)/5.9 = 0.44 (23-20.7)/5.0 = 0.46
c. (12-20.4)/5.9 = -1.42 (12-20.7)/5.0 = -1.74
d. The relative size of the difference from the mean and the standard deviation caused the reversal of position.
e. Eng. z = 0.95, Math z = 1.06, Read. z = 0.78, Science reas. z = 1.11, Composite z = 1.06. Therefore, Science Reasoning; it has the highest positive z-score.

2.136 for A: z = (28.1-25.7)/1.8 = 1.33
for B: z = (39.2-34.1)/4.3 = 1.19
Therefore, B has the lower relative position.

OBJECTIVE 2.6 EXTRA PROBLEMS

2.137 From 175 through 225 words, inclusive.

2.138 Nearly all of the data, 99.7%, lies within 3 standard deviations of the mean.

2.139 a. 97.6 is 2 standard deviations above the mean {z = (97.6-84.0)/6.8 = 2.0}, therefore 2.5% of the time more than 97.6 hours will be required.

b. 95% of the time, the time to complete will fall within 2 standard deviations of the mean, that is 84.0 ± 2(6.8) or from 70.4 to 97.6 hours.

2.140 a. 50% b. ≈68% c. ≈84%

2.141 a. Range should be approximately equal to 6 times the standard deviation for data with a normal distribution.

b. An approximation of the standard deviation can be found by dividing the range by 6.

2.142 $1 - (1/k^2) = 1 - (1/4^2) = 1 - (1/16) = 15/16 = 0.9375$;
at least 93.75%

2.143 a. at most 11% b. at most 6.25%

2.144 The interval 11.7 to 30.1 represents the mean plus or minus two standard deviations.

a. According to Chebyshev's theorem, we can be sure that at least 75% of the distribution is contained within the interval.

b. If the distribution is normal, then approximately 95% of the distribution is contained within the interval.

2.145 a.

class limits	*x*	*f*
1 - 4	2.5	6
4 - 7	5.5	9
7 - 10	8.5	8
10 - 13	11.5	10
13 - 16	14.5	6
16 - 19	17.5	4
19 - 22	20.5	4
22 - 25	23.5	2
25 - 28	26.5	1
	∑	50

b. $\bar{x} = \sum x/n = 560/50 = \underline{11.2}$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 8184.5 - (560^2/50) = 1912.5$

$s = \sqrt{SS(x)/(n-1)} = \quad = \quad = 6.247 = \underline{6.2}$

c. $\bar{x} \pm 2s = 11.2 \pm 2(6.2) = 11.2 \pm 12.4$ or $\underline{-1.2}$ $\underline{\text{to}}$ $\underline{23.6}$
96% of the data (48/50) is between -1.2 and 23.6.

2.146 a. $\bar{x} = \sum x/n = 1621.1/100 = 16.211 = \underline{16.21}$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 27988.6 - (1621.1^2/100) = 1708.9479$

$s = \sqrt{SS(x)/(n-1)} = \sqrt{1708.9479/99} = 4.155 = \underline{4.16}$

```
b. 12.0   12.0   12.1   12.1   12.1   12.2   12.2   12.3   12.3
   12.3   12.5   12.5   12.5   12.5   12.6   12.7   12.7   12.8
   12.8   12.9   13.0   13.0   13.0   13.0   13.1   13.1   13.2
   13.4   13.4   13.4   13.5   13.6   13.7   13.7   13.7   13.8
   13.9   14.0   14.2   14.2   14.2   14.3   14.3   14.3   14.4
   14.4   14.5   14.6   14.7   14.8   14.9   15.1   15.1   15.2
   15.2   15.3   15.4   15.4   15.9   15.9   15.9   16.0   16.0
   16.1   16.2   16.3   16.8   16.8   16.8   17.0   17.1   17.3
   17.4   17.6   17.7   17.7   17.9   18.0   18.4   18.7   18.8
   19.3   19.3   20.1   20.7   21.3   21.3   21.4   21.5   21.5
   22.0   22.7   22.8   24.8   25.7   25.8   26.8   27.1   29.9
   30.7
```

c. $\bar{x} \pm s = 16.21 \pm 4.16 = \underline{12.05}\ \underline{\text{to}}\ 20.\underline{37}$ → 82/100 = 0.82

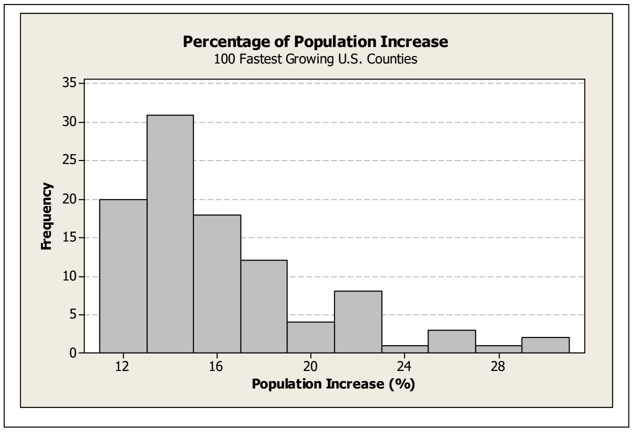
$\bar{x} \pm 2s = 16.21 \pm 2(4.16) = 16.21 \pm 8.32$ or $\underline{7.89}\ \underline{\text{to}}\ \underline{24.53}$
93% of the data (93/100) is between 7.89 and 24.53.

$\bar{x} \pm 3s = 16.21 \pm 3(4.16) = 16.21 \pm 12.48$ or $\underline{3.73}\ \underline{\text{to}}\ \underline{28.69}$
98% of the data (98/100) is between 3.73 and 28.69.

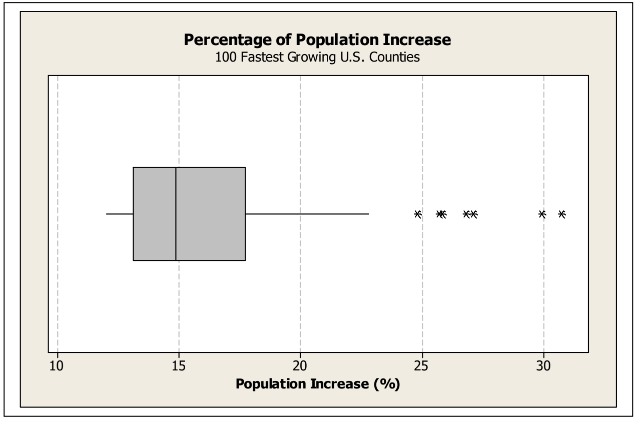
d. The empirical rule says approximately 68%, 95%, and 99.7% of the data are within one, two, and three standard deviations, respectively; the 82%, 93% and 98% do not agree with the rule; the distribution is not normal.

e. Chebyshev's theorem says at least 75%, and 89%, of the data are within two, and three standard deviations, respectively; 93% and 98% both satisfy the theorem.

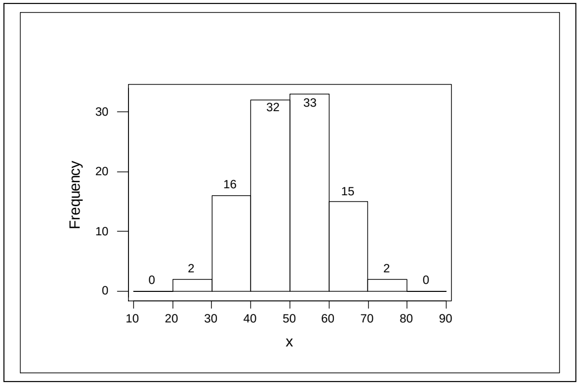
f. Graphs indicate a skewed right distribution, not normal.



Graphs may vary, but the boxplot may be the best choice for showing skewness.



2.147 a. Answers will vary.



Within one standard deviation, 40 to 60, is 33 + 32 or 65 of the 100 data. 65%

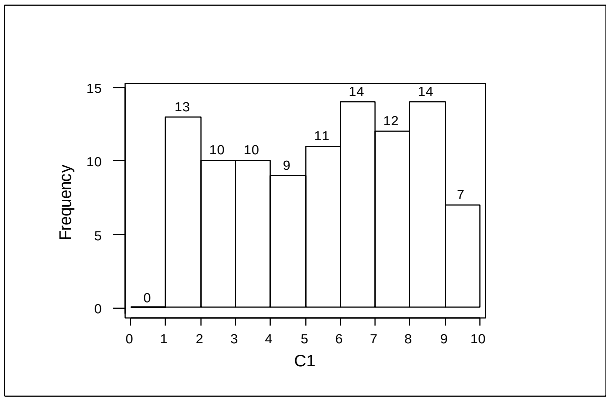
Within two standard deviations, 30 to 70, is 16+32+33+15 or 96 of the 100 data. 96%

Within three standard deviations, 20 to 80, is all 100 of the data, or 100%.

The above results are extremely close to what the empirical rule claims will happen.

b, c, d. Not all sets of 100 data will result in percentages this close. However, expect very similar results to occur most of the time.

2.148 a. Answers will vary



For the data pictured above; mean = 5.4677,
standard deviation = 2.5788

Within one standard deviation: 2.8889 to 8.0465, 56 of the 100 data or 56%
Within two standard deviations: 0.3101 to 10.6253, 100 of the 100 data or 100%
Within three standard deviations: 100%
Within four standard deviations: 100%

Chebyshev's theorem says at least 75%, 89%, and 93.75% of the data are within two, three and four standard deviations; 100%, 100%, and 100% all satisfy the theorem.

The empirical rule says approximately 68%, 95%, and 99.7% of the data are within one, two, and three standard deviations; the 64%, 100% and 100% do not agree with the rule; the distribution is not normal.

NOTE: If the percentage calculated for within one standard deviation is significantly less than 68%, a rectangular distribution is strongly suggested. (see the graph in part a)

b, c, d. The results obtained will vary from sample to sample, but they will all satisfy Chebyshev's theorem.

OBJECTIVE 2.7 EXTRA PROBLEMS

2.149 Yes, if all 8 employees earned $300 each, the mean would be $405.56 and if all 8 employees earned $350 each, the mean would be $450. $430 falls within this interval.

Or, if the mean of 9 employees is 430, then the total is 3870. The 8 employees then would need to make 2620 (3870-1250) and their average earnings would be 2620/8, or 327.50, which is within the interval.

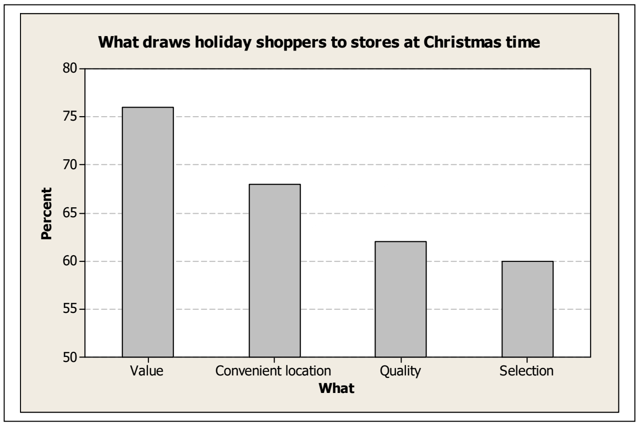
2.150 a. Answers will vary, but following are a few of the more obvious possibilities. If you research and find the original information, you'll find more.

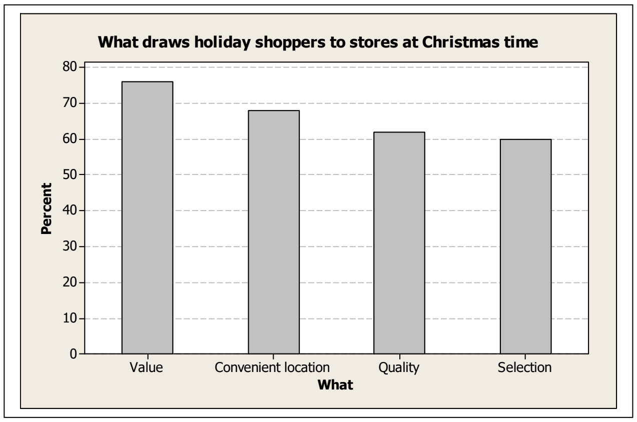
1. The ranking information covers an 11-year period, while the tuition information covers 35 years. Yet they are shown horizontally as being the same.
2. The units for tuition (share of median income) and ranking (rank number) are totally different, yet the vertical axis treats them as having common units.
3. The ranking graph is placed below the tuition graph, creating the impression that cost exceeds quality. Since the vertical scale is meaningless, either line could have been "on top".
4. The sharp "drop" in the ranking graph actually represents an improved ranking. A ranking of "15th best" is not better than a ranking of "6th best," however the vertical scale used makes it look like it is.

b. The caption under the graph suggests that Cornell's rank has been erratic by varying from 6th to 15th on the national ranking over the 12 years reported. With the hundreds of colleges and universities that exist, to consistently hold a rank like this is quite good.

The "upside-down" scale with the best ranking at the bottom is totally misleading.

2.151





Answers will vary but the 50 to 80 scale gives a misleading conclusion, namely that there is a significant difference among the responses. Starting the percent scale at 0 gives the true perspective of the relationship among the percents.

2.152 Answers will depend on the article selected. Answers will vary.

Show the relationships between the mean and standard deviation formulas for ungrouped and grouped data with a "side by side" example.

ungrouped data

	x	x^2
	2	4
	2	4
	3	9
	9	81
	11	121
Σ	27	219

grouped data

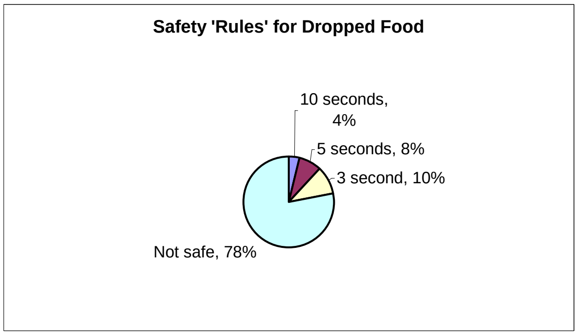
x	f	xf	x^2f
2	2	4	8
3	1	3	9
9	1	9	81
11	1	11	121
Σ	5	27	219

Have the students note that $n = \Sigma f$, $\Sigma x = \Sigma xf$, and $\Sigma x^2 = \Sigma x^2 f$. Work through formulas for $\overline{x}$ and s for each set of data.

CHAPTER REVIEW PROBLEMS

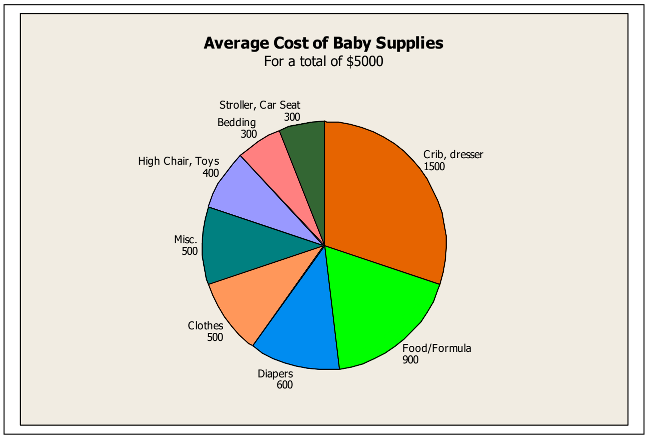
Formatted: Centered

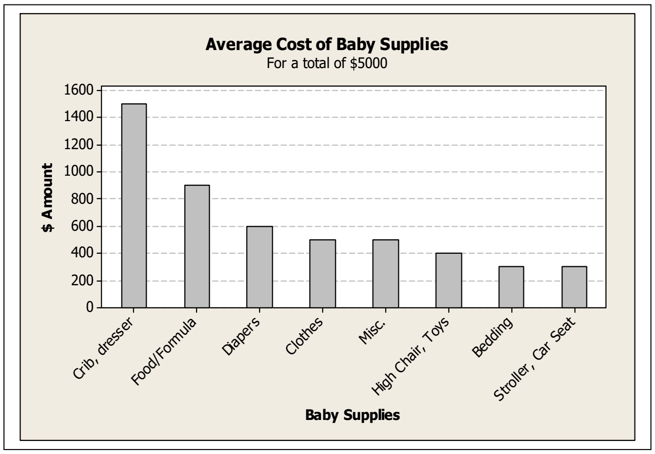
2.153



b. (300)(0.10) = 30 respond "Three-second rule"
(300)(0.08) = 24 respond "Five-second rule"
(300)(0.04) = 12 respond "10-second rule"
(300)(0.78) = 234 respond "Not safe"

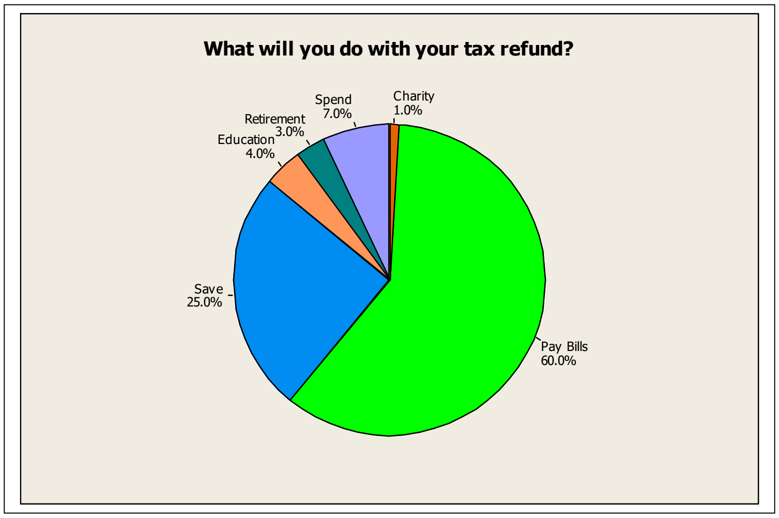
2.154 a.





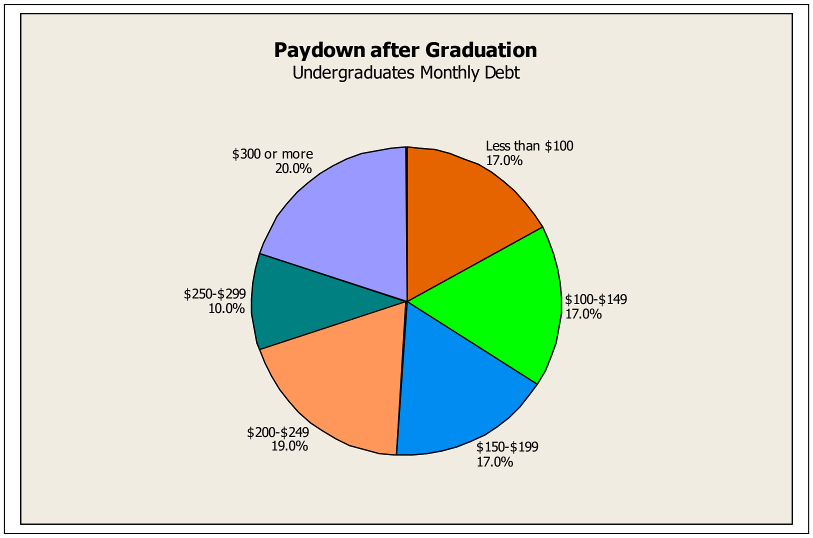
c. Answers will vary. For demonstrating the major cost of $1500, the bar graph works best. For demonstrating the relationship between the costs with respect to the whole amount, the pie chart or a divided bar graph work best.

2.155 a.

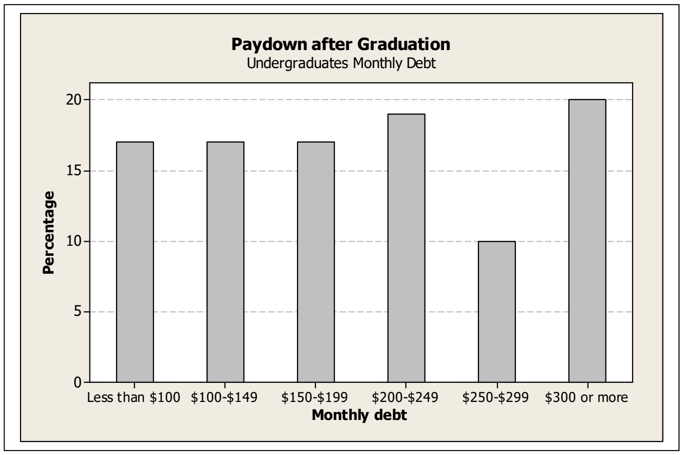


b. Answers may vary. Circle graph demonstrates the smaller percentages differences. It is easier to read.

2.156 a.

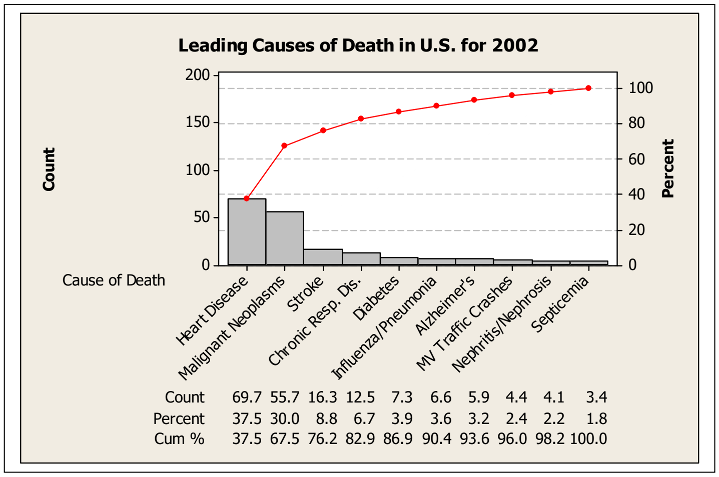


b.



c. Answers will vary. The circle graph gives a more uniform appearance among all of the debt amounts. The bar chart emphasizes the difference between the two highest debts.

2.157 a.



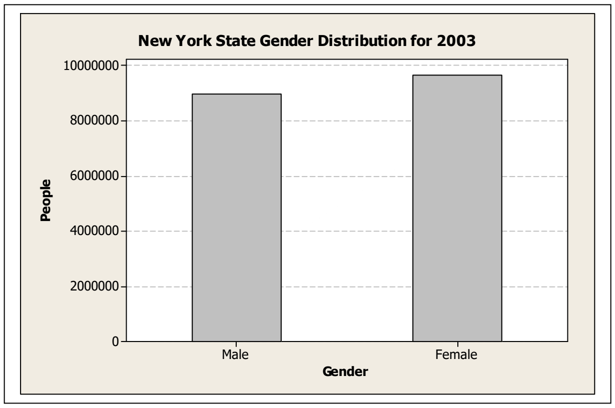
b. Answers will vary but the Pareto shows that the two leading causes of death, heart disease and malignant neoplasms account for nearly 70% of deaths.

2.158 a.

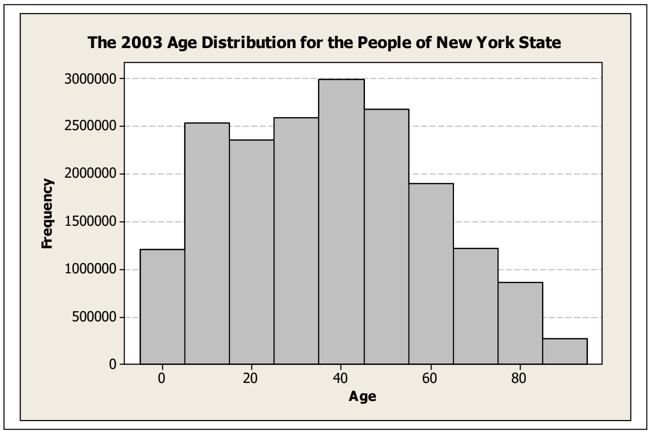
Gender - **x**	People - **f**	Rel. Freq.
Male	8953019	0.481331
Female	9647508	0.518669
Σ	18600527	1.000000

Age - **x**	Number - **f**	Rel. Freq.
Under 5 years	1205816	0.064827
5 to 14 years	2537813	0.136438
15 to 24 years	2353665	0.126538
25 to 34 years	2587995	0.139136
35 to 44 years	2991609	0.160835
45 to 54 years	2682845	0.144235
55 to 64 years	1897521	0.102014
65 to 74 years	1218850	0.065528
75 to 84 years	857177	0.046083
85 years and over	267236	0.014367
Σ	18600527	1.000000

b.



c.



d. In part c, the variable is qualitative and the values can be arranged in any order. In part c, the variable is quantitative and the "bars" have only one meaningful arrangement, numerical order.

2.159 a. numerical b. attribute c. numerical
d. attribute e. numerical

2.160 a. numerical b. numerical c. attribute
d. attribute e. attribute

2.161 a. Mean increased; when one data increases, the sum increases.

b. Median is unchanged; the median is affected only by the middle value(s).

c. Mode is unchanged.

d. Midrange increased; an increase in either extreme value increases the sum *H*+*L*.

e. Range increased; difference between high and low values increased.

f. Variance increased; data are now more spread out.

g. Standard deviation increased; data are now more spread out.

2.162 a. Mean is unchanged; the sum remained the same.

b. Median increased; the median is affected only by the middle value and since it increased, so did the median.

c. Mode increased; the change created a shift in repeated values.

d. Midrange is unchanged; extreme values did not change.

e. Range is unchanged; extreme values did not change.

f. Variance increased; data are now more spread out.

g. Standard deviation increased; data are now more spread out.

2.163 Data summary: $n = 8$, $\sum x = 36.5$, $\sum x^2 = 179.11$

a. $\bar{x} = \sum x/n = 36.5/8 = 4.5625 = \underline{4.56}$

b. $s = \sqrt{[\sum x^2 - ((\sum x)^2/n)]/(n-1)}$

=

= = 1.3405 = $\underline{1.34}$

c. These percentages seem to average very closely to 4%.

2.164 Data summary: $n = 13$, $\sum x = 1133.1$, $\sum x^2 = 98{,}769.35$

a. $\bar{x} = \sum x/n = 1133.1/13 = 87.1615 = \underline{87.16}$

b. $s = \sqrt{[\sum x^2 - ((\sum x)^2/n)]/(n-1)}$

=

= = 0.7422 = $\underline{0.74}$

c. The octane ratings seem to average somewhat less than 87.5 with a surprisingly large variability.

2.165 Data summary: $n = 118$, $\sum x = 2364$

a. $\bar{x} = \sum x/n = 2364/118 = 20.034 = \underline{20.0}$

b. $d(\tilde{x}) = (n+1)/2 = (118+1)/2 = 59.5$th;
$\tilde{x} = (17+17)/2 = \underline{17}$

c. mode = $\underline{16}$

d. $nk/100 = (118)(25)/100 = 29.5$; therefore $d(P_{25}) = 30$th
$Q_1 = P_{25} = \underline{15}$

$nk/100 = (118)(75)/100 = 88.5$; therefore $d(P_{75}) = 89$th

$Q_3 = P_{75} = \underline{21}$

e. $nk/100 = (118)(10)/100 = 11.8$; therefore $d(P_{10}) = 12$th

$P_{10} = \underline{14}$

$nk/100 = (118)(95)/100 = 112.1$; therefore $d(P_{95}) = 113$

$P_{95} = \underline{43}$

2.166 a. Hours of TV watched

```
0.| 00000 00000 0
0.| 55
1.| 0000
1.| 55
2.| 0000
2.| 55555
3.| 0
3.|
4.| 0
4.|
5.| 0
5.|
6.| 0
```

b. $\bar{x} = \sum x/n = 46.5/32$
$= 1.453 = \underline{1.45}$

c. $d(\tilde{x}) = (n+1)/2 = 16.5$th
$\tilde{x} = \underline{1}$

d. mode = $\underline{0}$

e. midrange = $(H+L)/2$
$= (0+6)/2 = \underline{3}$

f. The mode represents the most common amount of TV watched.

g. The mean includes the concept of total amount of television watched by the people in the sample.

h. range = $H - L = 6.0 - 0.0 = \underline{6.0}$

i. $SS(x) = \sum x^2 - ((\sum x)^2/n)$
$= 142.25 - (46.5^2/32) = 74.6797$

$s^2 = SS(x)/(n-1) = 74.6797/31 = 2.409 = \underline{2.41}$

j. $s = \ = \ = 1.552 = \underline{1.55}$

2.167 Data:

```
63  67  66  63  69  74  72  70  71  71
72  70  75  85  84  85  85  86  94  91
90  90  95 105 104
```

Data summary: $n = 25$, $\sum x = 1{,}997$, $\sum x^2 = 163{,}205$

$\bar{x} = \sum x/n = 1997/25 = 79.88 = \underline{79.9}$

$SS(x) = \sum x^2 - ((\sum x)^2/n)$
$= 163{,}205 - (1{,}997^2/25) = 3684.64$

$s^2 = SS(x)/(n-1) = 3684.64/24 = 153.5267$

$s = \ = \ = 12.3906 = \underline{12.4}$

2.168 Summary: $n = 17$, $\sum x = 56.2$, $\sum x^2 = 206.924$

a. $\bar{x} = \sum x/n = 56.2/17 = 3.306 = \$3.31$

b. depth = (17+1)/2 = 9th; median = $3.67

ranked:

1.36 1.71 1.83 2.20 2.61 2.86 2.92 3.13 3.67
3.68 3.71 3.75 3.97 4.00 4.36 4.70 5.74

c. midrange = (1.36 + 5.74)/2 = $3.55

d. Answers will vary. Sample student answer: The mean, median, and mode are close together. This is often found in normal distributions. Because the mean is less than the median, the distribution is slightly skewed to the left.

e. s^2 = 1.320201, s = 1.149 = $1.15

f. 3.31 ± 1.15 is $2.16 to $4.46 -> 12/17 = 70.6%

g. 3.31 ± 2(1.15) is $1.01 to $5.61 -> 16/17 = 94.1%

h. Data are approximately normally distributed. Because 70.6% of the data is within 1 standard deviation of the mean, and because 94.1% of the data is within 2 standard deviations of the mean, this data has an approximately normal distribution since 68% of the data is within 1 standard deviation of the mean, and 95% of the data is within 2 standard deviations of the mean in a normally distributed data set.

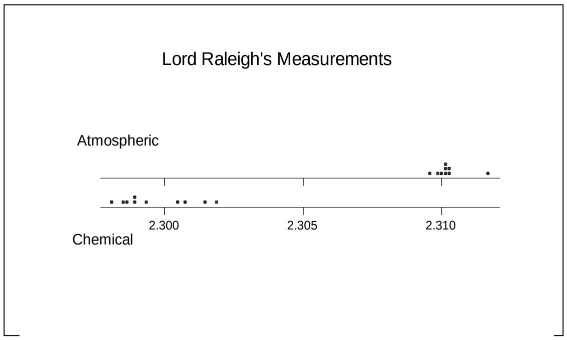
2.169 a. The population is the U.S. commercial airline industry. three variables are involved: number of reports, numbers of passengers, number of reports per 1000 airline passengers.

b. Data; they are values of the variable, number of reports per 1000 airline passengers.

c. Statistic; it summarizes the data for one month. It is used to estimate the parameter, or the value for the whole population.

d. No. The 19 airline values are a sample of the airline industry; not all are included here.

2.170 a.



b. **'Atmospheric'**

2.30956 2.30986 2.31001 2.31010 2.31010 2.31017 2.31024
2.31028 2.31163

$n = 9$, $\sum x = 20.79195$, $\sum x^2 = 48.03391206$

$\bar{x} = \sum x/n = 20.79195/9 = 2.31022 = \underline{2.3102}$

$d(\tilde{x}) = (n+1)/2 = 5^{th}$ $\tilde{x} = \underline{2.3101}$

$SS(x) = \sum x^2 - ((\sum x)^2/n)$
$= 48.03391206 - (20.79195^2/9) = 0.0000026375$

$s^2 = SS(x)/(n-1) = 0.0000026375/8 = 0.0000003296875$

$s = \ = \sqrt{0.00008} = 0.00057 = \underline{0.0006}$

$nk/100 = (9 \times 25)/100 = 2.25$ >> depth = 3rd >>
$Q_1 = 2.31001$ and $Q_3 = 2.31024$

'Chemical'

2.29816 2.29849 2.29869 2.29889 2.29890 2.29940
2.30054 2.30074 2.30143 2.30182

$n = 10$, $\sum x = 22.99706$, $\sum x^2 = 52.88649238$

$\bar{x} = \sum x/n = 22.99706/10 = 2.299706 = \underline{2.2997}$

$d(\tilde{x}) = (n+1)/2 = 5.5^{th}$ $\tilde{x} = \underline{2.29915}$

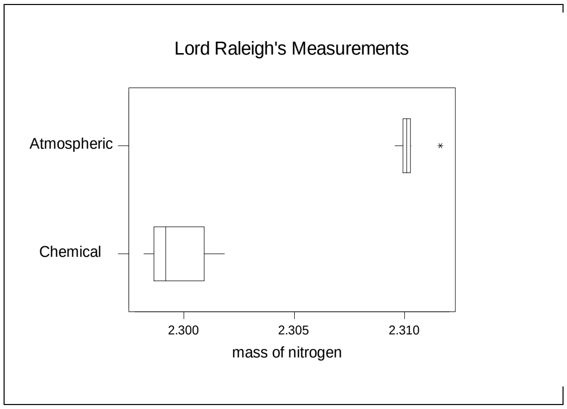
$SS(x) = \sum x^2 - ((\sum x)^2/n)$
$= 52.88649238 - (22.99706^2/10) = 0.00001551604$

$s^2 = SS(x)/(n-1) = 0.00001551604/9 = 0.00000172396$

$s = \ = \sqrt{0.00000} = 0.00131 = \underline{0.0013}$

$nk/100 = (10 \times 25)/100 = 2.5$ >> depth = 3rd >>
$Q_1 = 2.29869$ and $Q_3 = 2.30074$

c.



d. The two sets of data appear to be unrelated. The graphs both show two totally disjointed sets of data.

2.171 a. $\bar{x}$ = \$196,861 s = \$62,819

b. $\bar{x}$ − s = 196,861 − 62,819 = \$134,042 and
$\bar{x}$ + s = 196,861 + 62,819 = \$259,680

c. 34, 34/50 = 0.68 = 68%

d. $\bar{x}$ − 2s = 196,861 − 2(62,819) = \$71,223 and
$\bar{x}$ + 2s = 196,861 + 2(62,819) = \$322,499

e. 48, 48/50 = 0.96 = 96%

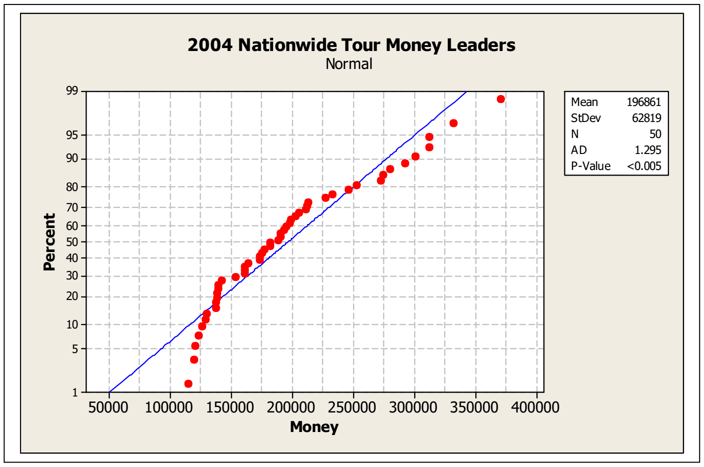
f. $\bar{x}$ − 3s = 196,861 − 3(62,819) = \$8,404 and
$\bar{x}$ + 3s = 196,861 − 3(62,819) = \$385,318

g. 50/50 = 1.00 = 100%

h. They agree with Chebyshev's theorem; both percentages exceed the values cited.

i. 68%, 96% and 100% as a set are very close to the 68%, 95% and 99.7% cited by the empirical rule. The results suggest the distribution might be approximately normal, but we need to look at an appropriate graph before making that decision.

j.



k. The normality test graph suggests that the distribution is not normal, both by means of the points not following a straight line and the *p*-value being less than 0.05. By itself, the empirical rule is not enough to go by; one must also graph the distribution and/or use a standardized normality test. You might also construct a histogram.

2.172 a. through d. Answers will vary.

2.173 Data summary: $n = 100$, $\sum x = 1315$

a. $\bar{x} = \sum x/n = 1315/100 = 13.15$

b. $d(\tilde{x}) = (n+1)/2 = 50.5\text{th}$; $\tilde{x} = 13.85$

c. mode = 15.0

d. midrange = $(H+L)/2 = (15.8+10.1)/2 = 12.95$

e. range = $H - L = 15.8 - 10.1 = 5.7$

f. $nk/100 = (100)(25)/100 = 25$; therefore $d(P_{25}) = 25.5\text{th}$

$Q_1 = P_{25} = 10.95$

$nk/100 = (100)(75)/100 = 75$; therefore $d(P_{75}) = 75.5\text{th}$

$Q_3 = P_{75} = 14.9$

g. midquartile = $(Q_1 + Q_3)/2 = (10.95+14.9)/2 = 12.925$

h. $nk/100 = (100)(35)/100 = 35$; therefore $d(P_{35}) = 35.5\text{th}$

$P_{35} = 12.05$

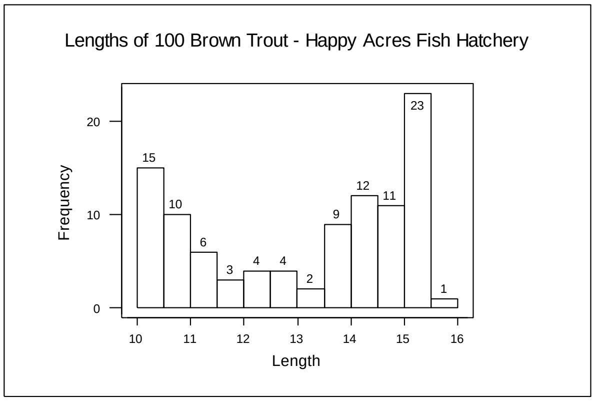
nk/100 = (100)(64)/100 = 64; therefore $d(P_{64}) = 64.5\text{th}$

$P_{64} = 14.5$

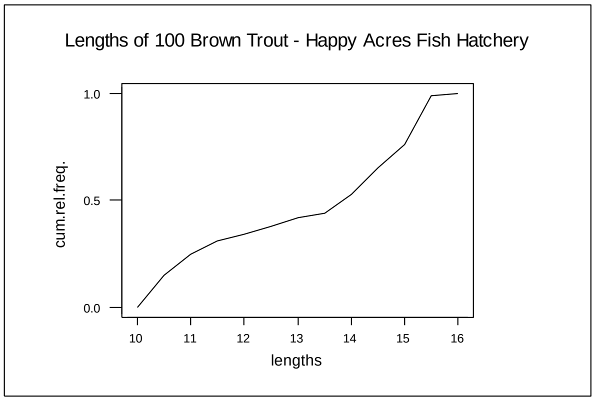
i.

Class limits	x	f	xf	x^2f	rel.fr	cum.r
10.0 - 10.5	10.25	15	153.75	1575.94	0.15	0.15
10.5 - 11.0	10.75	10	107.50	1155.62	0.10	0.25
11.0 - 11.5	11.25	6	67.50	759.38	0.06	0.31
11.5 - 12.0	11.75	3	35.25	414.19	0.03	0.34
12.0 - 12.5	12.25	4	49.00	600.25	0.04	0.38
12.5 - 13.0	12.75	4	51.00	650.25	0.04	0.42
13.0 - 13.5	13.25	2	26.50	351.13	0.02	0.44
13.5 - 14.0	13.75	9	123.75	1701.56	0.09	0.53
14.0 - 14.5	14.25	12	171.00	2436.75	0.12	0.65
14.5 - 15.0	14.75	11	162.25	2393.19	0.11	0.76
15.0 - 15.5	15.25	23	350.75	5348.94	0.23	0.99
15.5 - 16.0	15.75	1	15.75	248.06	0.01	1.00
	$\sum$	100	1314.00	17635.26		

j.



k. Shown in part i above.



l.

m. Summary: $n = 100$; $\Sigma xf = 1314$, $\Sigma x^2 f = 17635.26$

$\bar{x} = \sum xf/\sum f = 1314/100 = 13.14$

n. $SS(x) = \sum x^2 f - ((\sum xf)^2/\sum f)$
$= 17635.26 - (1314^2/100) = 369.3$

$s^2 = SS(x)/(\sum f - 1) = 369.3/99 = 3.73030$

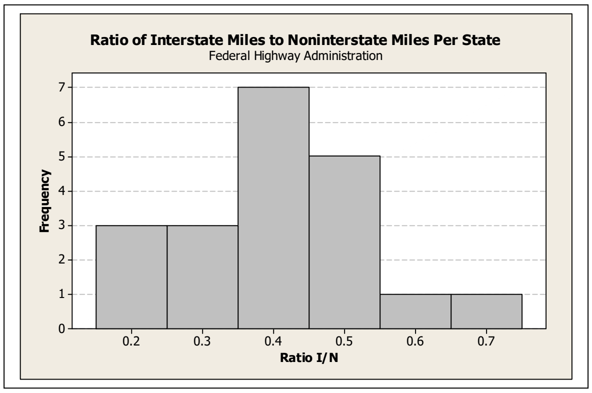
$s = \;=\; = 1.931399 = 1.93$

2.174 a. Answers will vary. They seem to be near 0.50.

b. Ratio I/N

0.398981	0.507942	0.398048	0.188356	0.444279
0.298039	0.367908	0.382559	0.371627	0.360406
0.494015	0.516796	0.531452	0.318303	0.559744
0.321414	0.481588	0.193109	0.745687	0.185714

c.



d. Data summary: $n = 20$, $\sum x = 8.06597$

$\bar{x} = \sum x/n = 8.06597/20 = 0.4033$

e. sum of interstate miles = 19,399

sum of noninterstate miles = 48,013

ratio I/N = 19,399/48,020 = 0.404036 = 0.404

f. The mean in part (d) treats all states equally and each does not have an equal number of highway miles. Part (e) uses the totals from all states, so each state influences the mean portionally to the number of miles of highway it has.

g. $\sum x^2 = 3.61829$

$SS(x) = \sum x^2 - ((\sum x)^2/n)$
$= 3.61829 - (8.06597^2/20) = 0.36529$

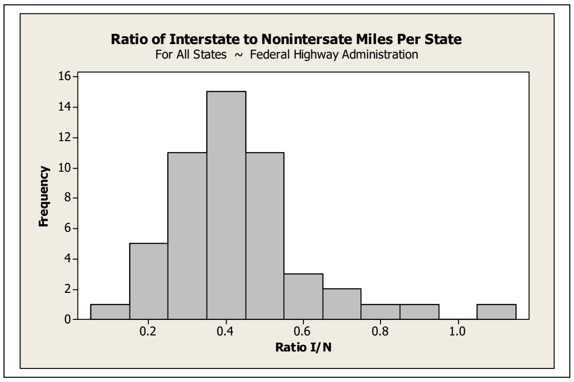
$s^2 = SS(\text{x})/(n-1) = 0.36529/19 = 0.019226126$

$s = \; = \sqrt{0.01922} = 0.138658 = 0.139$

h~~(b)~~.

State-All	Interstate-All	Non-Interstate-All	Ratio I/N
AL	905	2715	0.333333
AK	1082	1030	1.050485
AZ	1167	1565	0.745687
AR	656	2040	0.321569
CA	2458	5166	0.475803
CO	956	2624	0.364329
CT	346	617	0.560778
DE	41	281	0.145907
DC	13	70	0.185714
FL	1471	2896	0.507942
GA	1245	3384	0.367908
HI	55	292	0.188356
ID	611	1760	0.347159
IL	2170	3511	0.618058
IN	1169	1713	0.682428
IA	782	2433	0.321414
KS	874	2874	0.304106
KY	763	2130	0.358216
LA	904	1701	0.531452
ME	367	922	0.398048
MD	481	976	0.492828
MA	569	1392	0.408764
MI	1243	3501	0.355041
MN	912	3060	0.298039
MS	685	1881	0.364168
MO	1181	3215	0.367341
MT	1192	2683	0.444279
NE	482	2496	0.193109
NV	560	1573	0.356008
NH	235	589	0.398981
NJ	431	1641	0.262645
NM	1000	1935	0.516796
NY	1674	3476	0.481588
NC	1019	2742	0.371627
ND	572	2156	0.265306
OH	1574	2812	0.559744
OK	930	2431	0.382559
OR	728	3026	0.240582
PA	1758	3729	0.47144
RI	71	197	0.360406
SC	842	1780	0.473034
SD	679	2260	0.300442
TN	1073	2172	0.494015
TX	3233	10157	0.318303
UT	940	1253	0.7502
VT	320	373	0.857909
VA	1118	2365	0.472727
WA	764	2654	0.287867
WV	549	1195	0.459414
WI	745	3404	0.21886
WY	913	2037	0.448208

h~~(c)~~.



(d). Data summary: n = 51, $\sum x$ = 21.4809

$\bar{x}$ = $\sum x/n$ = 21.4809/51 = <u>0.421</u>

(e). sum of interstate miles = 46508

sum of noninterstate miles = 114885

ratio I/N = 46508/114885 = 0.404822 = <u>0.405</u>

(f). The mean in part $h(d)$ treats all states equally and each does not have an equal number of highway miles. Part $h(e)$ uses the totals from all states, so each state influences the mean portionally to the number of miles of highway it has.

(g). Σx^2 = 10.5491

$SS(x) = \sum x^2 - ((\sum x)^2/n)$ = 10.5491- (21.4809²/51) = 1.50147

$s^2 = SS(x)/(n-1)$ = 1.50147/50 = 0.030029425

s = = $\sqrt{0.0300}$ = 0.17329 = <u>0.173</u>

2.175 a. through c. Answers will vary.

d. Sum of Area = 3,022,316

Sum of population = 279,583,437

Overall density = 279,583,437/3,022,316

= 92.506 = 93 people/sq. mile

```
Densities
     5.04      6.13      9.09      9.80     14.95     15.49     18.08
    22.16     26.30     32.68     35.28     38.38     41.26     45.04
    49.35     50.34     51.99     58.52     59.62     63.35     74.79
    78.00     80.29     86.17     86.43     92.10     95.52    100.06
   129.19    132.82    134.68    139.05    152.70    167.55    170.72
   173.43    213.44    220.20    270.90    272.92    275.41    380.94
   382.78    554.79    679.89    768.94    863.52   1073.81
```

e. Data summary: $n = 48$, $\sum x = 8503.88$

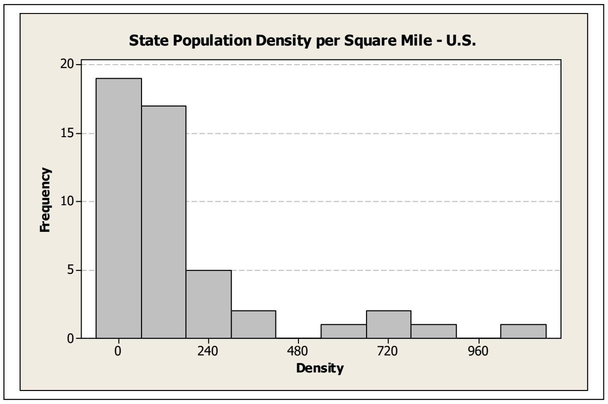
$\bar{x} = \sum x/n = 8503.88/48 = 177.164 = \underline{177.2}$

$d(\tilde{x}) = (n+1)/2 = 24.5\text{th};\quad \tilde{x} = (86.17+86.43)/2 = \underline{86.3}$

no mode – no value repeats

midrange = $(H+L)/2 = (1073.81+5.04)/2 = \underline{539.425}$

f.



g. Highest: New Jersey, Rhode Island, Massachusetts, Connecticut, and Maryland

Lowest: Wyoming, Montana, North Dakota, South Dakota, New Mexico

h. Answers will vary depending on answers in parts a through c.

2.176 a. Data summary: $n = 20$, $\sum x = 1122.4$

$\bar{x} = \sum x/n = 1122.4/20 = \underline{56.12}$

$d(\tilde{x}) = (n+1)/2 = 10.5\text{th};\quad \tilde{x} = (25.6+36.5)/2 = 31.05 = \underline{31.1}$

midrange = $(H+L)/2 = (259.0+15.0)/2 = \underline{137.0}$

b. $\Sigma x^2 = 129465$

$SS(x) = \sum x^2 - ((\sum x)^2/n)$

$= 129465 - (1122.4^2/20) = 66475.912$

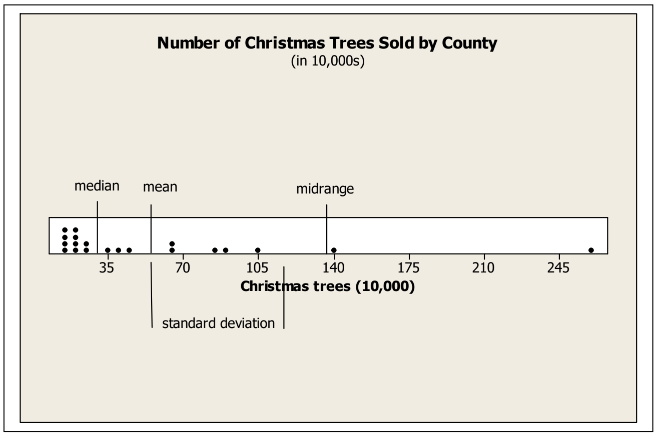
$s^2 = SS(x)/(n-1) = 66475.912/19 = 3498.732211$

$s = \ = \sqrt{3\,4\,9783\,2} = \underline{59.15}$

c. All of the measures of central tendency are quite different, indicating a skewed distribution.

d. The large standard deviation is due to the wide range of values.

e.

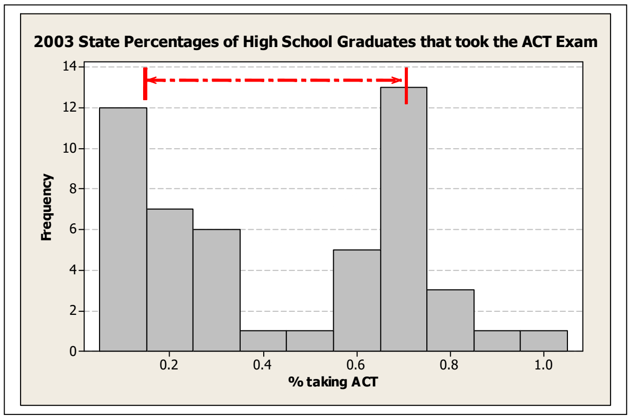


f. see above

g. Data is skewed right, J-shaped.

2.177 a. Answers will vary.

b.



c. Distribution is bimodal. Large concentrations between 10% and 30% and then 60% and 80%.

d. Answers will vary.

e. $\bar{x}$ = 0.4308

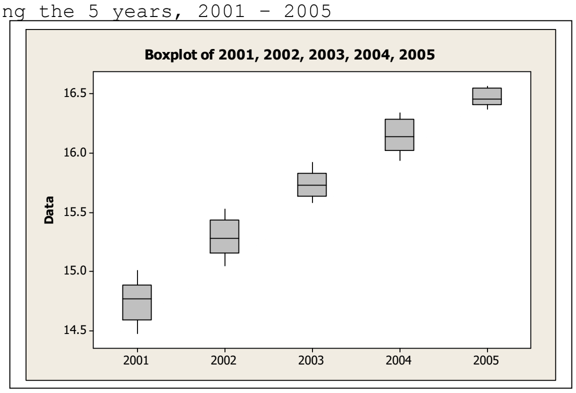
f. Mean falls between the two high concentration areas. It is not representative of these data.

g. *s* = 0.2874

h. Percentage between:
Looking at histogram: [7+6+1+1+5+(1/2)(13)]/50 = 53%
Looking at data: 28/50 = 56%

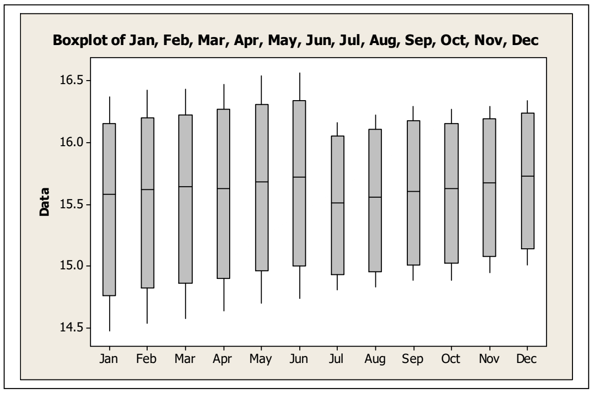
i. The standard deviation is so large due to there being so few data near the mean, resulting in the data being quite wide spread. The lowest value is 0.05 and the highest is 1.0, that is from 5% to 100%, with very few data between 0.35 and 0.55.

2.178 a.
Using the 5 years, 2001 - 2005



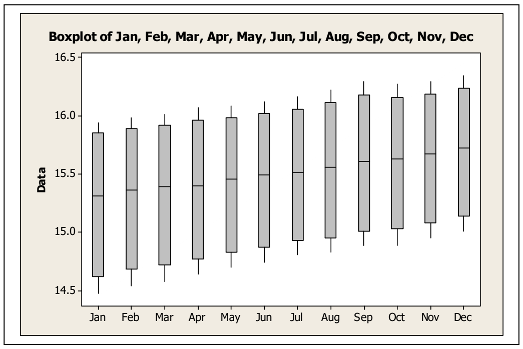
```
Year    N    Mean    StDev
2001   12   14.755   0.170
2002   12   15.286   0.152
2003   12   15.736   0.110
2004   12   16.148   0.135
2005    6   16.465   0.0734
```

Using 12 months, 2001 - 2005



```
Month  N    Mean     StDev
Jan    5   15.484    0.741
Feb    5   15.534    0.735
Mar    5   15.562    0.724
Apr    5   15.596    0.722
May    5   15.646    0.717
Jun    5   15.682    0.711
Jul    4   15.500    0.581
Aug    4   15.540    0.597
Sep    4   15.598    0.600
Oct    4   15.605    0.583
Nov    4   15.648    0.573
Dec    4   15.700    0.567
```

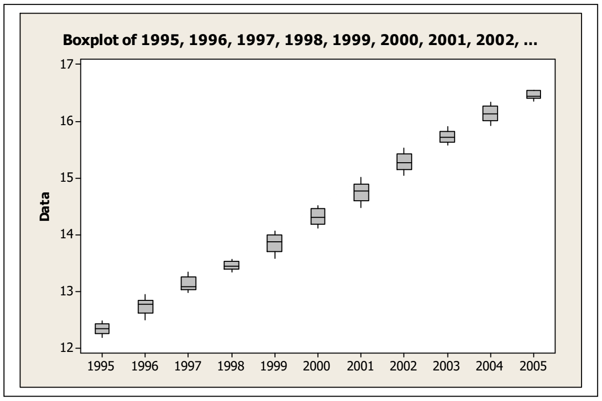
Notice the pattern "restarted" with July, so the 2005 January to June data was removed temporarily.



Notice that the pattern extends throughout the entire year now. All three boxplots show a consistent increasing pattern, so why the "hitch" in the second boxplot?

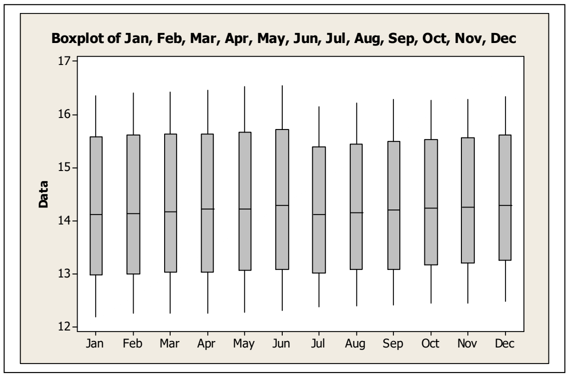
Hourly earnings are definitely increasing annually as seen by the regular increase in the annual mean hourly earnings. The standard deviation seems quite stable for both the monthly and yearly calculations. However on closer inspection of the data, it can be seen that the hourly earnings are increasing at a very steady rate, approximately $0.04, every month over the 5-year period.

b. Using 11 years of data, 1995 to 2005



```
Year      N    Mean   StDev
1995     12  12.341  0.0977
1996     12  12.745   0.138
1997     12  13.133   0.125
1998     12  13.450  0.0750
1999     12  13.855   0.161
2000     12  14.314   0.139
2001     12  14.755   0.170
2002     12  15.286   0.152
2003     12  15.736   0.110
2004     12  16.148   0.135
2005      6  16.465  0.0734
```

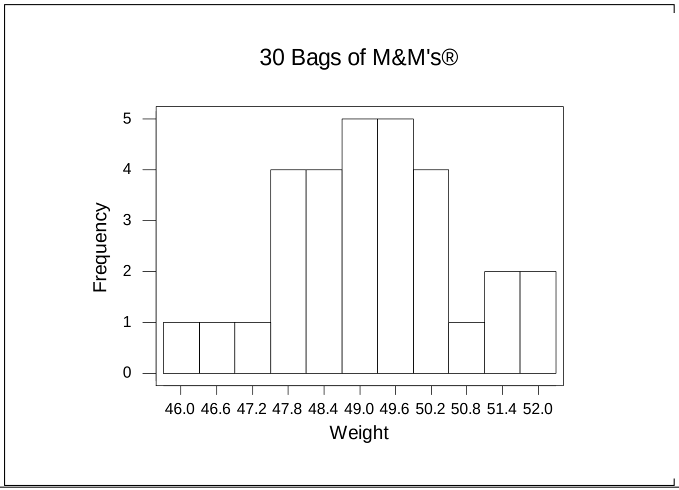
Using 12 months of data, 1995 - 2005

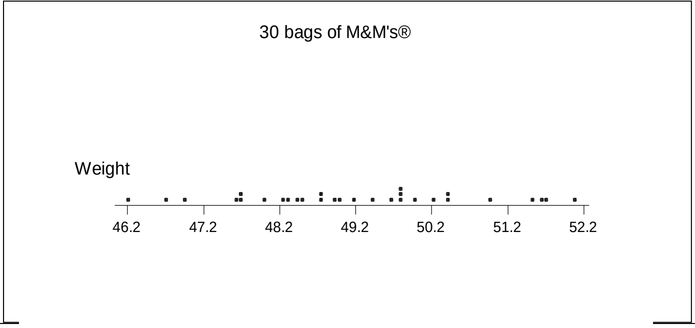


```
Month    N   Mean  StDev
Jan     11  14.205  1.401
Feb     11  14.240  1.413
Mar     11  14.264  1.422
Apr     11  14.305  1.413
May     11  14.342  1.423
Jun     11  14.376  1.423
Jul     10  14.183  1.297
Aug     10  14.232  1.296
Sep     10  14.276  1.310
Oct     10  14.298  1.299
Nov     10  14.329  1.306
Dec     10  14.372  1.308
```

The same observations are made for the 11-years of information as were made in part (a).

2.179 a. Weight





~~b. $\bar{x}$ = 49.215, median = 49.07, s = 1.522, min = 46.22, max = 52.06~~

b.

$$\mu = \frac{\sum x}{n} = \frac{1476.44}{30} = 49.215$$

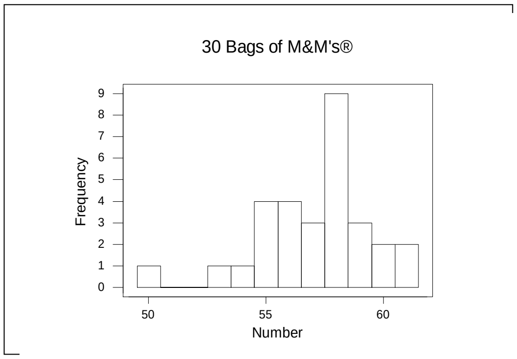
$d(\tilde{x})$ = (n+1)/2 = (30+1)/2 = 15.5th; $\tilde{x}$ = (48.98+49.16)/2 = 49.07

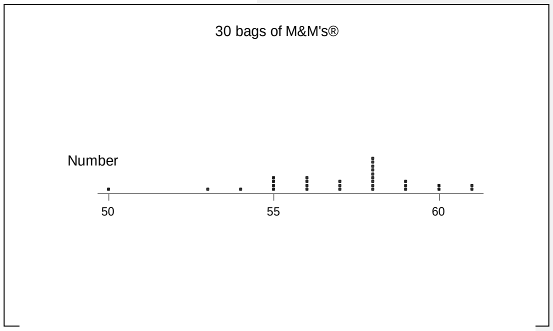
$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{72729.68 - \frac{(1476.44)^2}{30}}{30-1}} = 1.522$$

$\bar{x}$ = 49.215, median = 49.07, s = 1.522, min = 46.22, max = 52.06

c. No, there do not seem to be any inconsistencies in the weight data.

d. Find number per bag: 58, 62, 59, etc. e.





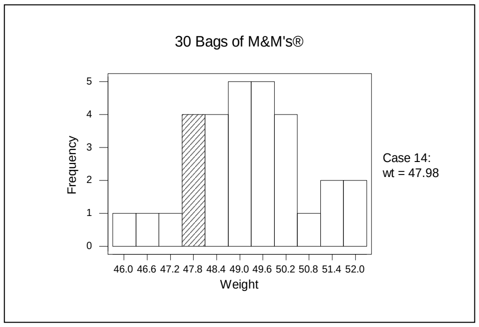
f. $\bar{x}$ = 57.1, median = 58, s = 2.383, min = 50, max = 61

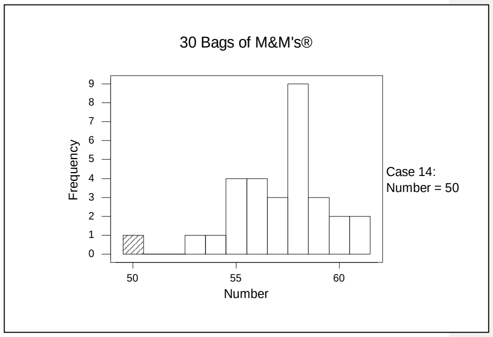
$$\mu = \frac{\sum x}{n} = \frac{1713}{30} = 57.1$$

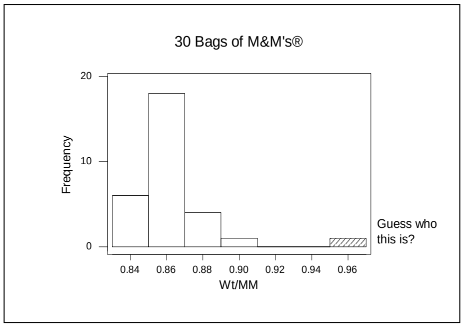
$d(\tilde{x})$ = (n+1)/2 = (30+1)/2 = 15.5th; $\tilde{x}$ = (58+58)/2 = 58

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{97977 - \frac{(1713)^2}{30}}{30-1}} = 2.383$$

g. One bag has "only 50" M&M's in it. This one value appears to be quite different from the rest of the values. Case 14 seems to have a total bag weight that is "typical" (see histogram below). However, its count number is approximately 10% smaller than the "typical" count (see histogram below). This means that the individual M&M's would have to be 10% larger to make up the weight (see histogram below). Very suspicious!







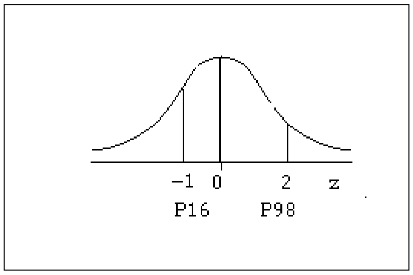
h. After bag 14 was weighed and before the M&M's were counted, someone ate a few of them, approximately 5 of them.

Draw a diagram of a normal curve with its corresponding percentages for standard deviations away from the mean. Add the percentages from the left to the right, until the desired *z*-value is reached. The sum equals the percentile.

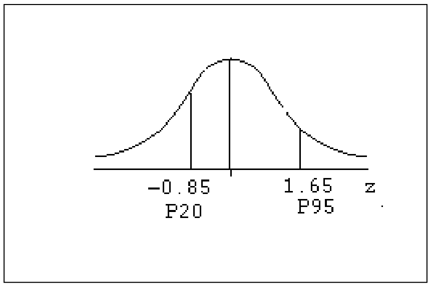
2.180 a. 2.5+13.5+34+34+13.5 = 97.5%; therefore, P_{98}

b. 2.5+13.5 = 16%; therefore, P_{16}

c.



2.181 a. 0.50-0.20 = 0.30; therefore, $z \approx -0.8$ or -0.9
b. 0.95-0.50 = 0.45; therefore, $z \approx +1.6$ or +1.7
c.



2.182 x = (z)(st.dev) + mean

Sit-ups: $x = (-1)(12) + 70 = \underline{58}$
Pull-ups: $x = (-1.3)(6) + 8 = 0.2 = \underline{0}$
Shuttle Run: $x = (0)(0.6) + 9.8 = \underline{9.8}$
50 yd. dash: $x = (1)(.3) + 6.6 = \underline{6.9}$
Softball: $x = (0.5)(16) + 173 = \underline{181}$

2.183 The z-scores must be changed to percentiles in order to make the comparisons. Percentages are obtained from the empirical rule.

$z = 2$ corresponds to P_{97}
$z = 1$ corresponds to P_{84}
$z = -1$ corresponds to P_{16}
$z = 0$ corresponds to P_{50}

Therefore, Joan has the higher relative score for fitness, agility, and flexibility. Jean scored highest in posture, while they scored the same in strength.

2.184 a. and b. Answers will vary. Certainly not much changed over the years.

Standard deviation for English has increased from 5.6 to 5.9, that is approximately a 5% increase. That is about the only change.

2.185 Data summary: $n = 8$, $\sum x = 31{,}825$, $\sum x^2 = 126{,}894{,}839$

a. $\overline{x} = \sum x/n = 31,825/8 = 3978.1$

b. $SS(x) = \sum x^2 - ((\sum x)^2/n)$
$= 126,894,839 - (31,825^2/8) = 291,010.88$

$s^2 = SS(x)/(n-1) = 291,010.88/7 = 41,572.982$

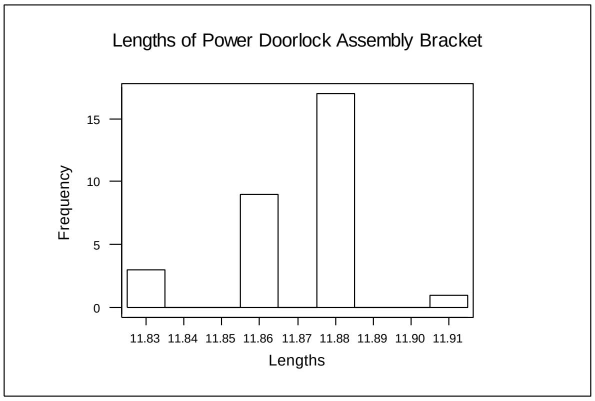
$s = \ = \ = 203.9$

c. $\overline{x} \pm 2s = 3978.1 \pm 2(203.9)$
$= 3978.1 \pm 407.8$ or 3570.3 to 4385.9

2.186 a. 11.87 or 11.88 are reasonable estimates
b. $n = 30$, $\sum xf = 356.10$, $\sum x^2 = 4226.9160$

x	f	xf	x^2f
11.83	3	35.49	419.8467
11.86	9	106.74	1265.9364
11.88	17	201.96	2399.2848
11.91	1	11.91	141.8481
$\sum$	30	356.10	4226.9160

c.



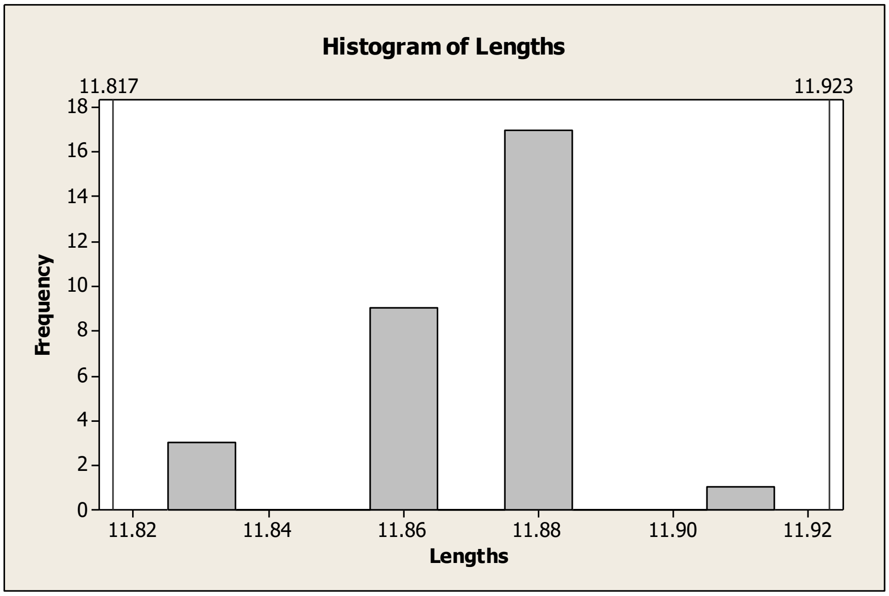
~~c.~~

d. $\bar{x} = \sum xf / \sum f = 356.10/30 = \underline{11.87}$

$SS(x) = \sum x^2 f - ((\sum xf)^2/\sum f)$
$= 4226.9160 - (356.10^2/30) = 0.009$
$s^2 = SS(x)/(\sum f-1) = 0.009/29 = 0.0003$

$s = \ = \ = \underline{0.0176}$

e. $\bar{x} \pm 3s = 11.87 \pm 0.053$ or $\underline{11.817}$ $\underline{\text{to}}$ $\underline{11.923}$



f. Yes. The interval in part e is within 11.7 to 12.3.

2.187 a.

```
Variable     N     Mean  StDev   Variance   Sum   Sum of Squares
Calories     40   111.88  44.92   2017.55  4475.00  579325.00
Sodium(mg)   40    566.3  238.4   56824.0  22650.0  15041700.0
```

b. Calories: 111.88 ± 2(44.92) or between 22.04 and 201.72
Sodium: 566.3 ± 2(238.4) or between 89.5 and 1043.1 mg
Only 2 of the brands of soups fall outside this calorie interval, so 95% are included within the interval. Only 1 brand falls outside the sodium content interval, so 98% are included within the interval.
Yes, Chebyshev's theorem is satisfied; both percentages are at least 75%.

c. Sodium: 566.3 ± 238.4 or between 327.9 and 804.7 mg
The empirical rule predicts that 68% of the brands of soups' sodium content will fall between 327.9 and 804.7, provided the distribution is normally distributed. In fact, these limits include 27 of the brands out of 40, or 67.5%. Therefore, these

results suggest the sodium content of the soups does satisfy that part of the empirical rule.

2.188 a. Stem-and-Leaf plot:

50 Service Times

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

b.

Summary of data: n = 50, Σx = 1810, Σx^2 = 69518

18	19	21	23	24	25	27	27	27	28
28	29	31	31	31	32	32	32	32	33
35	35	35	35	35	35	36	36	38	38
38	38	38	38	40	43	43	43	43	45
46	46	46	48	49	50	51	52	52	53

Mean: $\bar{x} = \sum x/n = 1810/50 = \underline{36.2}$

Median: $d(\tilde{x}) = (n+1)/2 = (50+1)/2 = 25.5^{th}$

$\tilde{x} = (35+35)/2 = 35$

Mode: mode = $\underline{35}$

Range: range = $H-L = 53 - 18 = \underline{35}$

Midrange: midrange = $(H+L)/2 = (53+18)/2 = \underline{35.5}$

Variance:

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 69518 - (1810^2/50) = 3996$

$s^2 = SS(x)/(n-1) = 3996/49 = \underline{81.551}$

Standard deviation:

$s = \sqrt{s^2} = \sqrt{81.551} = \underline{9.03}$

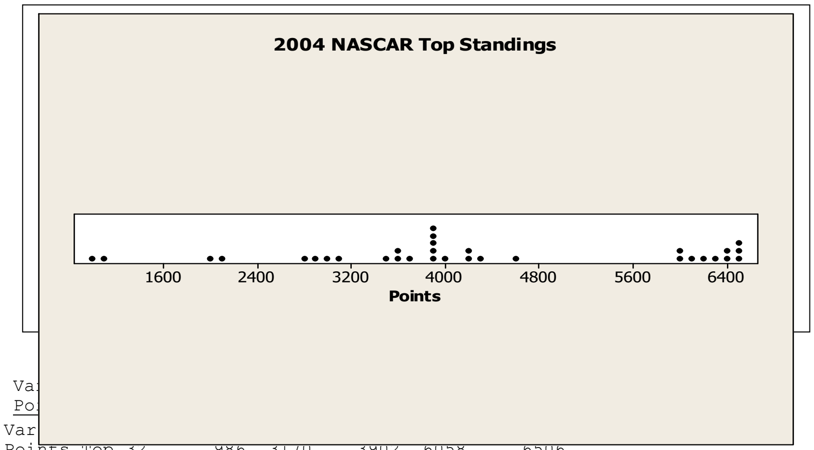
c.

Lowest Value	First Quartile	Median	Third Quartile	Highest Value
18	31	35	43	53

d. Chebyshev's theorem predicts that no less than 75% of the service times will fall between ±2 standard deviations from the mean. This interval is 36.2 ± 2(9.03), or between 18.14 and 54.08 minutes. In fact, 49 out of 50 heads of hair, or 98%, were cut within this interval.

e. Forty minutes should be about right, although this is a judgment call. Less than 30 minutes will make things hectic in the shop and back up customers. Quality would suffer. Over 40 minutes will start to generate excessive slack time between appointments and not keep the barbers busy.

2.189 a.



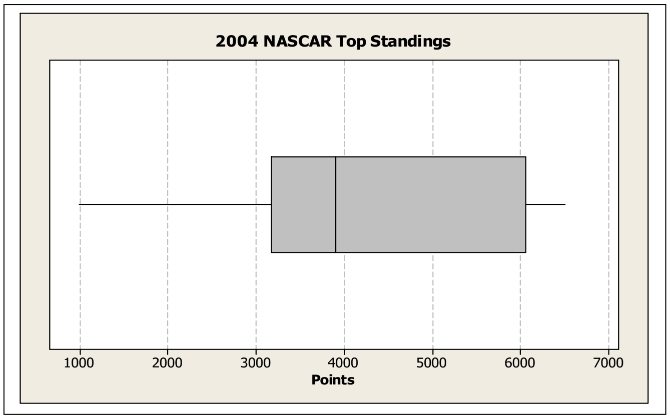
b.

Va

Po

c. Var

```
   Points Top 32      986  3170    3902  6058      6506
```

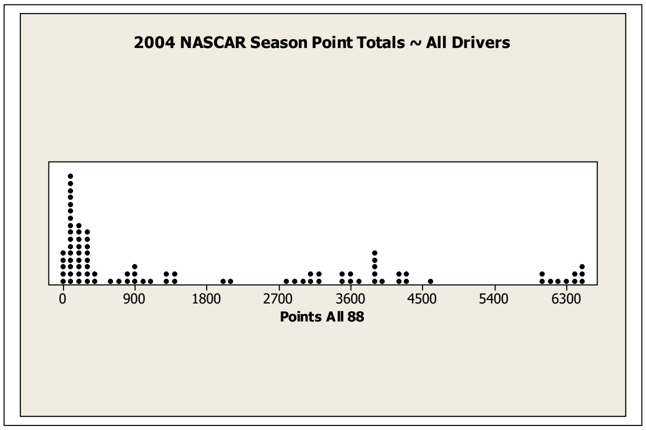


d. 4248 ± 2(1624), or between 1000 and 7496. 31 of the 32, or 96.95%, fall within this interval. This satisfies Chebyshev's requirement of at least 75%.

e. 4248 ± 1624, or between 2624 and 5872. 18 of the 32, or 56.3%, fall within this interval. This does not satify the empirical rule's claim of 68%.

f. The percentages found, 56.3% and 96.95%, are not consistent with the empirical rule. The points distribution for the top 32 is part of a skewed left distribution. There is a "wide" cluster near the top and then the distribution tails out to the left. By using the top 32, much of the left tail has been chopped off. The distribution is not normal.

g(a).

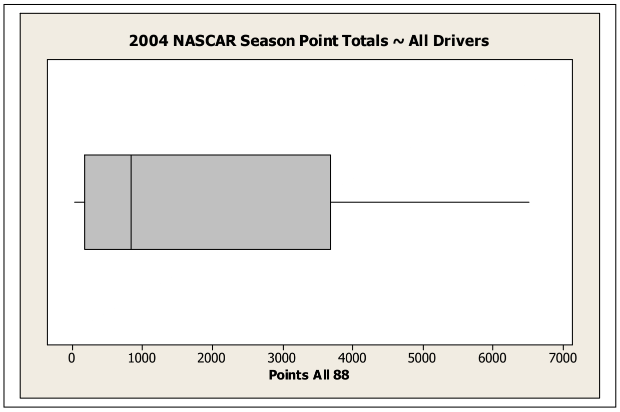


g(b).

```
                                                      Sum of
Variable         N  Mean  StDev      Sum     Squares
Points All 88   88  1946   2145   171209   733484657
```

g(c).

```
Variable       Minimum   Q1  Median    Q3  Maximum
Points All 88     34.0  176     835  3679     6506
```



g(d). 1946 ± 2(2145), or between -2344 and 6236. Yes, 82 of the 88, or 93.2%, fall within this interval. This satisfies Chebyshev's requirement of at least 75%.

g(e). 1946 ± 2145, or between -199 and 4091. No, 73 of the 88, or 83.0%, fall within this interval. This does not satify the empirical rule's claim of 68%.

g(f). The distribution is at least trimodal. There are at least three clusters of points on the dotpot that are distinctly separate from the others: there is a cluster below 400, one between 2700 and 4500, and one above 6000 points. This distribution is definitely different than the distribution of the top 32, although the "top 32" are visable on the dotplot. The distribution is definitely not a normal or an approximately normal distribution.

2.190 a. Including A and B, the distribution seems to be skewed to the right.

b. Excluding A and B, the distribution seems to be approximately normal; that is, mounded and approximately symmetrical about the middle.

```
           Data with  A  and  B  removed :
                              .
.           :    :    :       : .    :  . :    .    :  .
-------+---------+---------+---------+---------+---------
  24.0       28.0      32.0      36.0      40.0      44.0
```

c. "A" does not appear to represent an event that is typical of the distribution represented as being normal in (b).

2.191 There are many possible answers for this question; only one of those possibilities is shown.

a. 70, 77.5, 77.5, 77.5, 85 yields *s* = 5.30, which is the smallest standard deviation for a sample of 5 data with 70 and 85.
b. 70, 76, 85, 89, 95 yields *s* = 10.02.
c. 70, 85, 90, 99,110 yields *s* = 15.02.
d. In order to increase the standard deviation, the data had to become more dispersed.

2.192 There are many possible answers for this question; only one of those possibilities is shown.

a. 95 75 40 78 74 75 70 75 73 75 78 80 75 77 79 77 75 78
$\overline{x}$ = 74.94, *s* = 10.09

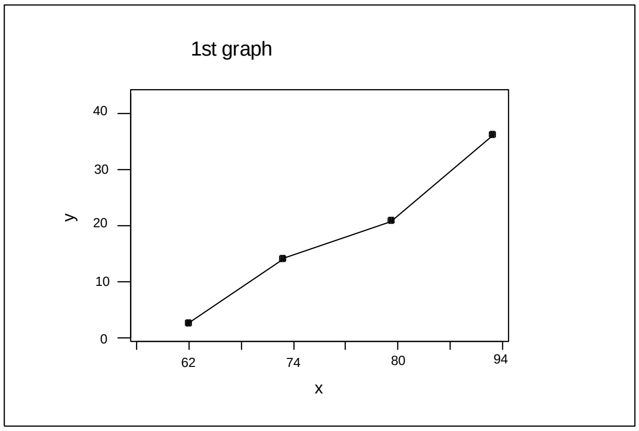
b. 98 75 40 78 74 75 70 75 73 75 78 80 75 77 79 77 75 78
$\overline{x}$ = 75.11, *s* = 10.46

c. 98 60 40 96 74 81 65 75 68 63 78 83 60 77 79 77 98 78
$\overline{x}$ = 75.00, *s* = 14.57

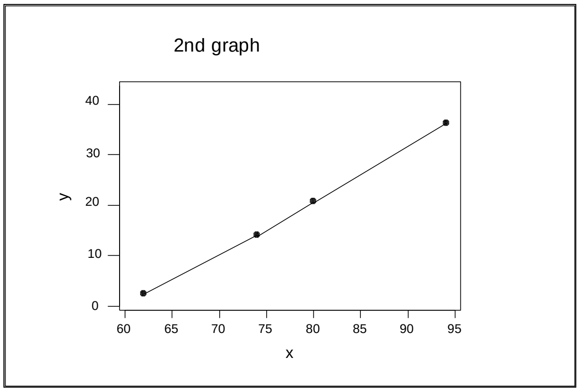
d. The majority of the data in sample (b) is located around the mean of 75, whereas the data in sample (c) is more evenly spread between the 40 and 98.

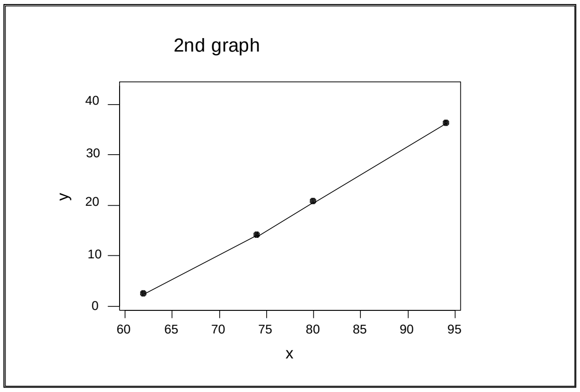
2.193

a.


1st graph
y
40
30
20
10
0
62
74
80
94
x

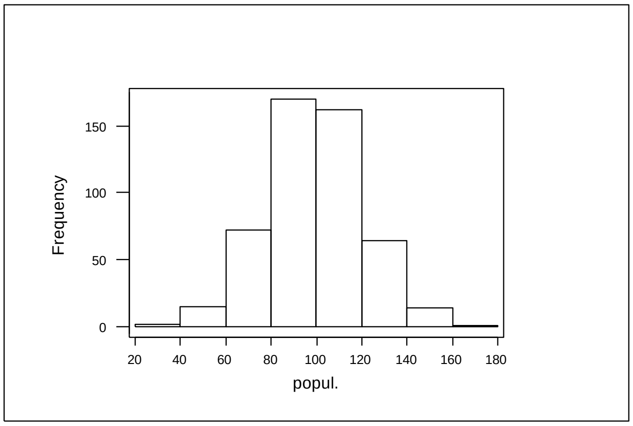
b.~~a.~~


2nd graph
y
40
30
20
10
0
60
65
70
75
80
85
90
95
x

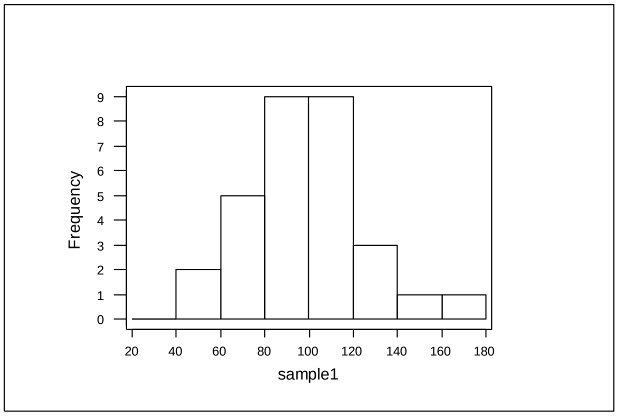


~~b.~~

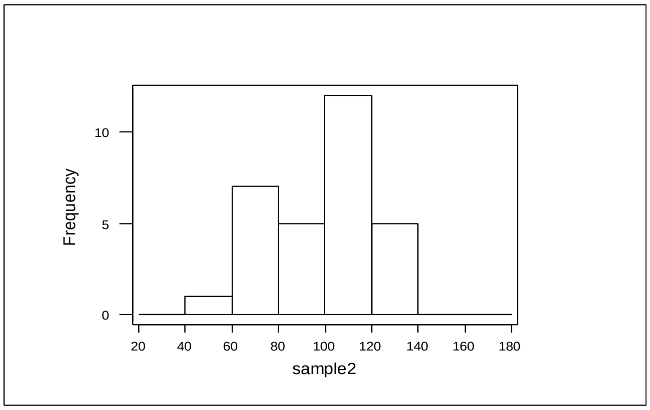
c. The line graph in (a) suggests an accelerated rate of increase from 1980 to 1995, while the line graph in (b) suggests that the rate of increase has been constant from 1962 to 1995.

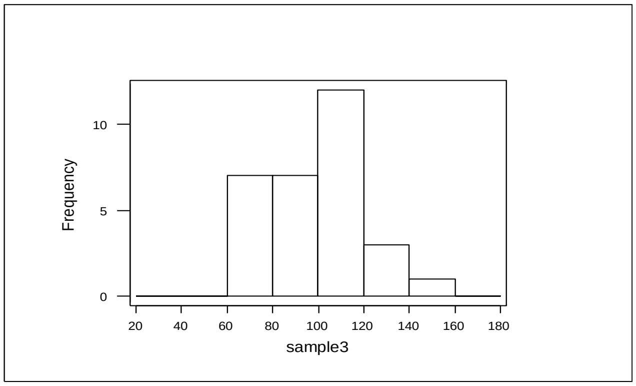

Frequency
150
100
50
0
20
40
60
80
100
120
140
160
180
popul.

2.194 a.


Frequency
9
8
7
6
5
4
3
2
1
0
20
40
60
80
100
120
140
160
180
sample1

b. and c.





```
     d. Variable                N        Mean      Median       StDev
```

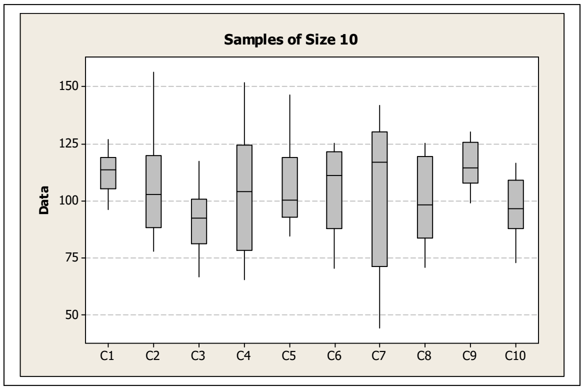
```
   popul.             500      98.932     99.190     20.915
   sample1             30      98.35      96.19      25.53
   sample2             30      96.84     101.26      21.12
   sample3             30      99.25     100.75      20.01
   sample4             30      97.45      95.28      18.83

  Variable       Minimum    Maximum         Q1         Q3
  popul.         31.792    162.786     84.358    113.016
  sample1         53.65     162.79      83.17     114.72
  sample2         53.18     128.83      77.83     112.62
  sample3         65.26     141.59      80.76     115.00
  sample4         57.90     151.98      88.52     107.45
```

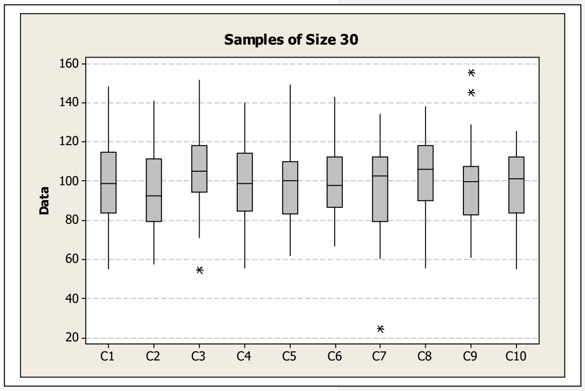
e. Yes, the sample statistics calculated closely resemble the population parameters.

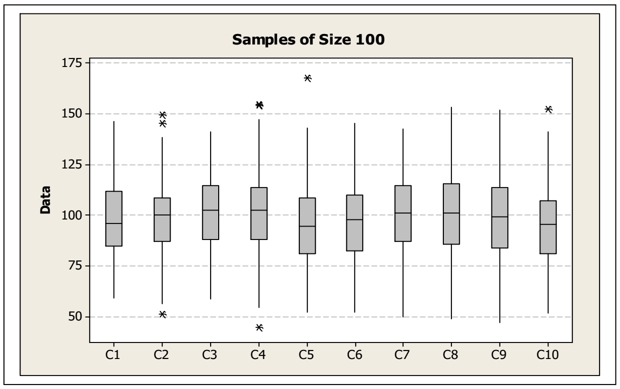
2.195 As the sample size increases, the distribution of the sample looks more like a normal distribution.

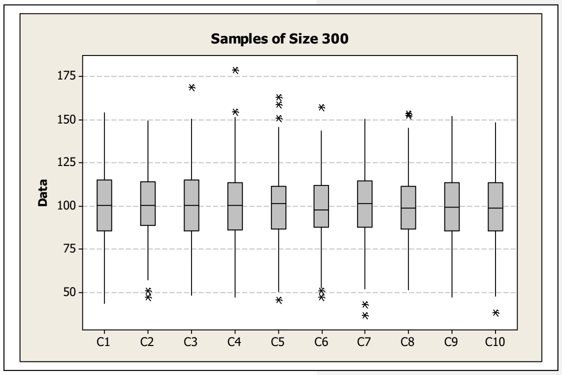
2.196 Samples of size 30 usually demonstrated some of the properties of the population. As the sample size was increased, more of the properties of the population were shown. The suggested distributions in this problem seem to require sample sizes greater than 30 for a closer match to the population.

2.197 a.



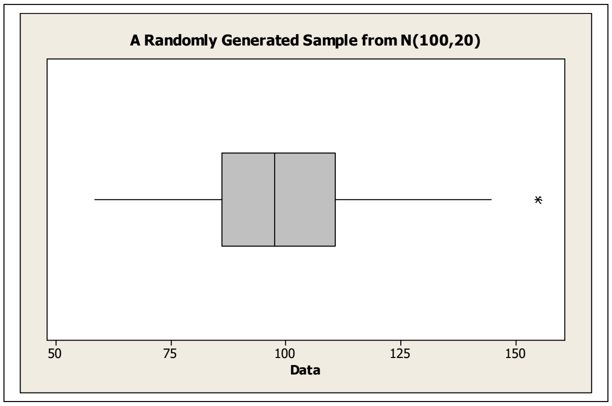




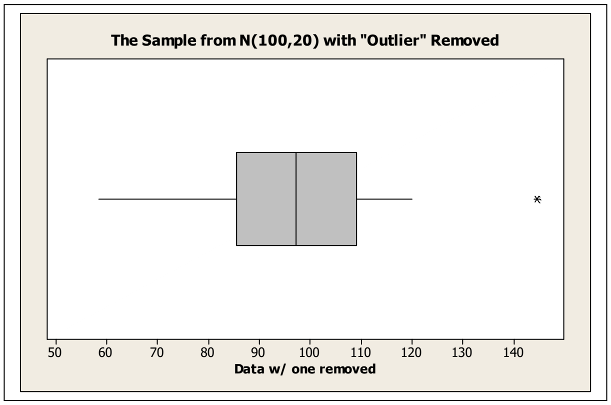


Answers will vary. There is greater variation in the medians and ranges of the data in the smaller sample sizes. There are more outliers in the larger samples sizes.

b.



c.

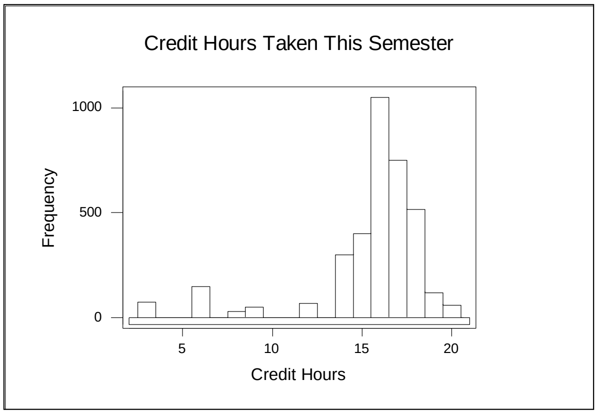


d. The boxplot in part (b) indicated that the largest value was an outlier, then the boxplot in part (c) said the next to the largest data value was now an outlier for the remaining sample. How far could this continue? It doesn't make sense, at this level, to be making decisions about outliers without additional information.

2.198

CredHrs	Freq	*xf*	x^2f	
	3	75	225	675
	6	150	900	5400
	8	30	240	1920
	9	50	450	4050
	12	70	840	10080
	14	300	4200	58800
	15	400	6000	90000
	16	1050	16800	268800
	17	750	12750	216750
	18	515	9270	166860
	19	120	2280	43320
	20	60	1200	24000
	Σ	3570	55155	890655

Summary: $n = 3570$; $\Sigma xf = 55{,}155$, $\Sigma x^2 f = 890{,}655$



a.

b. mean: $\bar{x} = \sum xf/\sum f = 55,155/3570= 15.449 = \underline{15.4}$

median: $d(\tilde{x}) = (\sum f+1)/2 = (3570+1)/2 = 1785.5^{th}$; $\tilde{x} = \underline{16}$

mode: mode = $\underline{16}$

midrange: midrange = $(H+L)/2 = (3+20)/2 = \underline{11.5}$

midquartile: midquartile = $(Q_1+Q_3)/2 = (15+17)/2 = \underline{16}$

c. $nk/100 = (3570)(25)/100 = 892.5$; $d(Q_1) = 893^{rd}$;

$Q_1 = \underline{15}$ and $Q_3 = \underline{17}$

d. $nk/100 = (3570)(15)/100 = 535.5$; $d(P_{15}) = 536$th

$P_{15} = \underline{14}$

$nk/100 = (3570)(12)/100 = 428.4$; $d(P_{12}) = 429$th

$P_{12} = \underline{14}$

e. range: range = $H - L = 20 - 3 = \underline{17}$

variance: $SS(x) = \sum x^2 - ((\sum x)^2/n)$

$= 890655 - (55155^2/3570) = 38533.42437$

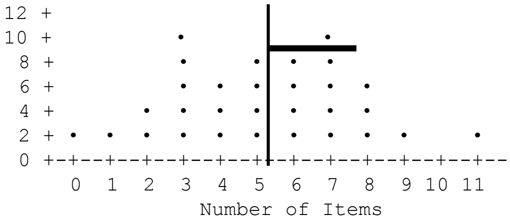
$s^2 = SS(x)/(n-1) = 38533.42437 /3569 = \underline{10.7967}$

standard deviation:

$s = \; = \sqrt{10.796} = 3.2858 = \underline{3.3}$

2.199 Summary: $n = 66$; $\Sigma xf = 353$; $\Sigma x^2 f = 2305$

a. Persistent Disagreements



This dot plot may be slightly misleading in that it does not show the odd frequencies. Since only multiples of 2 are shown, each odd frequency is rounded down to its previous even number. Subsequently, the frequency at 10 does not show. However, the dot plot does display the basic distribution adequately enough to exhibit its statistical characteristics.

b. $d(\tilde{x})$ = $(n+1)/2$ = (66+1)/2 = 33.5th; median = 5

x	f	xf	x^2f
0	2	0	0
1	2	2	2
2	4	8	16
3	10	30	90
4	7	28	112
5	9	45	225
6	8	48	288
7	11	77	539
8	7	56	448
9	3	27	243
10	1	10	100
11	2	22	242
Σ	66	353	2305

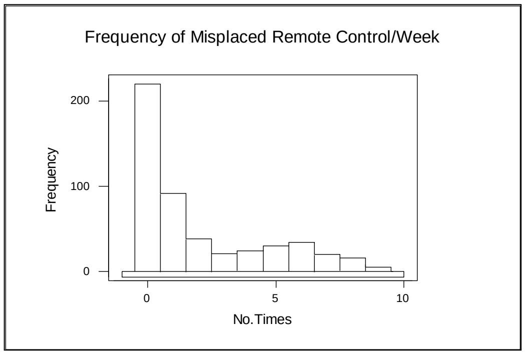
c. $\bar{x}$ = $\sum xf/\sum f$ = 353/66
= 5.34848 = 5.3

d. $SS(x) = \sum x^2f - ((\sum xf)^2/\sum f)$
= 2305 - (353²/66)
= 416.98485

$s = \sqrt{SS(x)/(\sum f-1)}$
=
= = 2.5328 = 2.5

e. and f. On graph in (a).

2.200 a.



b. $n = \Sigma f$ = 500, Σxf = 994, $\sum x^2 f$ = 5200
mean = 1.988, median = 1, mode = 0, midrange = 4.5

c. $SS(x) = \sum x^2f - ((\sum xf)^2/n)$
= 5200 - (994²/500) = 3223.928

$s^2 = SS(x)/(\Sigma f-1)$ = 3223.928/499 = 6.46078 = 6.46

s = = $\sqrt{6.4607}$ = 2.5418 = 2.5

d. Q_1 = 0, Q_3 = 4, P_{90} = 6

$nk/100$ = (500)(25)/100 = 125; $d(Q_1)$ = 125th;

$nk/100$ = (500)(75)/100 = 125; $d(Q_3)$ = 375th;

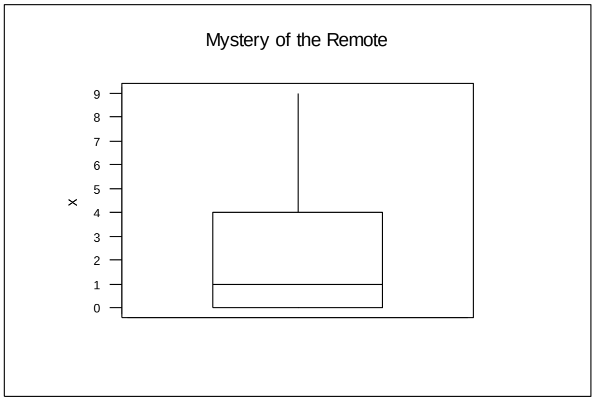
$nk/100$ = (500)(90)/100 = 450; $d(P_{90})$ = 450th;

$Q_1 = 0$, $Q_3 = 4$, and $P_{90} = 6$

e. midquartile = 2

midquartile = $(Q_1 + Q_3)/2 = (0 + 4)/2 = 2$

f. 5-number summary: 0, 0, 1, 4, 9



2.201 Summary: $n = 450$; $\sum xf = 21{,}055$ $\sum x^2 f = 1{,}082{,}162$

a. $\bar{x} = \sum xf / \sum f = 21{,}055/450 = 46.78888 = \underline{46.8}$

b. $SS(x) = \sum x^2 f - ((\sum xf)^2 / \sum f)$

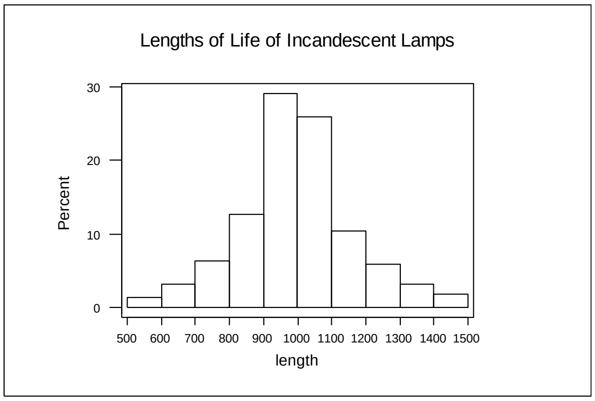
$= 1{,}082{,}162 - (21{,}055^2/450) = 97021.9444$

$s^2 = SS(x)/(\sum f - 1) = 97{,}021.9444/449 = 216.084508$

$s = \ = \ = 14.69981 = \underline{14.7}$

2.202 Summary: $n = 220$; $\sum xf = 219{,}100$, $\sum x^2 f = 224{,}470{,}000$

a.



b. $\bar{x} = \sum xf/\sum f$ = 219,100/220 = 995.90909 = <u>995.9</u>

c. $SS(x) = \sum x^2 f - ((\sum xf)^2/\sum f)$
= 224,470,000 - (219,100²/220) = 6,266,318.182

$s^2 = SS(x)/(\sum f - 1)$ = 6,266,318.182/219 = 28,613.32503

s = = = 169.15473 = <u>169.2</u>

2.203

x	f	xf	x^2f
150	320	48000	7200000
400	430	172000	68800000
600	170	102000	61200000
800	80	64000	51200000
$\sum$	1000	386000	188400000

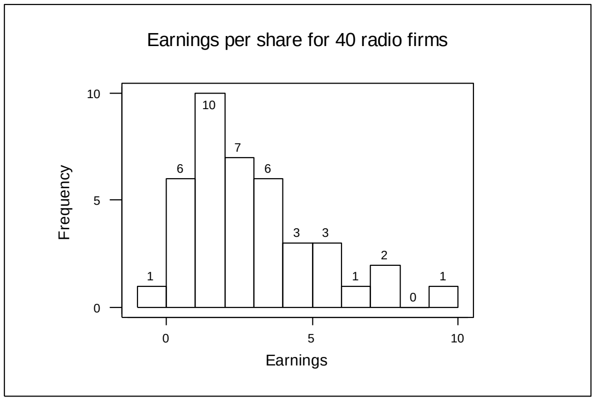
$\bar{x} = \sum xf/\sum f$ = 386000/1000 = $386

$SS(x) = \sum x^2 f - ((\sum xf)^2/\sum f)$ = 188400000 - (386000²/1000)
= 39404000

$s = \sqrt{SS(x)/(\sum f - 1)} = \sqrt{39404000/999}$ = $198.61

2.204 a.

Class limits	f
-1.00-0.00	1
0.00-1.00	6
1.00-2.00	10
2.00-3.00	7
3.00-4.00	6
4.00-5.00	3
5.00-6.00	3
6.00-7.00	1
7.00-8.00	2
8.00-9.00	0



9.0.10.0	1
Σ	40

b. $d(\tilde{x}) = (n+1)/2 = 20.5$th; median is in the class $2.00-$3.00.

CHAPTER PROJECT

Part 1 a. Ranked data, frequency distribution, relative frequency distribution, circle graph; dotplot; histogram; ogive, ogive on probability paper, box-and-whisker

b. Pareto – since data is number of different activities, typically would be organized numerically rather than by most to least; Stem-and-leaf – best used when the data has multiple digits, whereas this data is mostly single digit data, thus the stem-and-leaf would be the same as a dotplot.

c. List of ranked data:

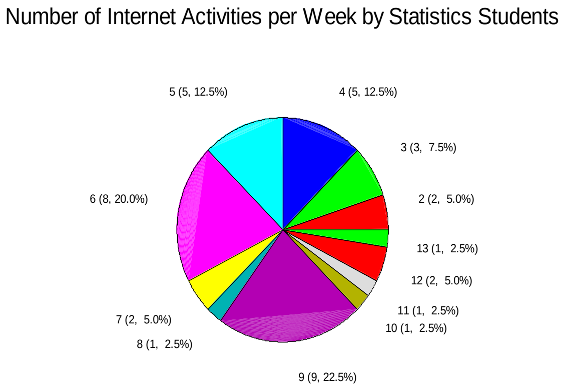
2	2	3	3	3	4	4	4	4	4
5	5	5	5	5	6	6	6	6	6
6	6	6	7	7	8	9	9	9	9
9	9	9	9	9	10	11	12	12	13

Frequency distribution:

Activities	Frequency
2	2
3	3
4	5
5	5
6	8
7	2
8	1
9	9
10	1
11	1
12	2
13	1
	40

```
Relative frequency distribution:
           Activities  Relative frequency
             2    .050
             3    .075
             4    .125
             5    .125
             6    .200
             7    .050
             8    .025
             9    .225
            10    .025
            11    .025
            12    .050
            13    .025
```

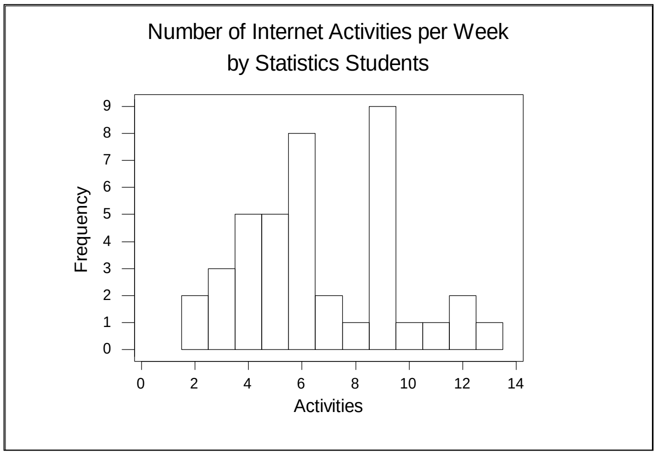
Circle graph:



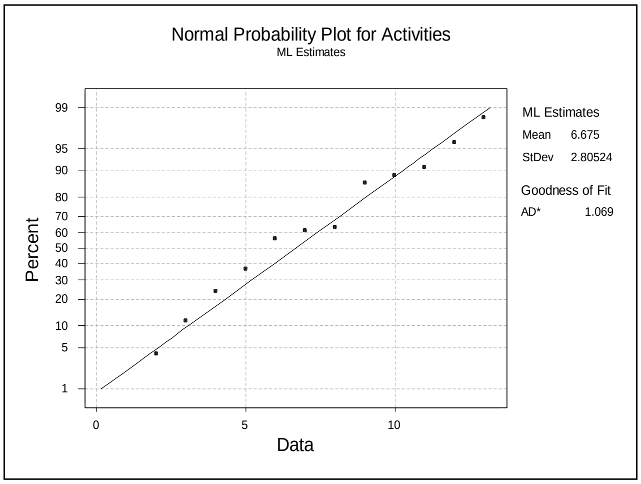
Dotplot: **Activities**

```
                                    .
                     :              :
           .    .    :              :
      .    :    :    :              :
 :    :    :    :    :    :    .    :    .    .    :    .
 +---------+---------+---------+---------+---------+------
2.0       4.0       6.0       8.0      10.0      12.0
```

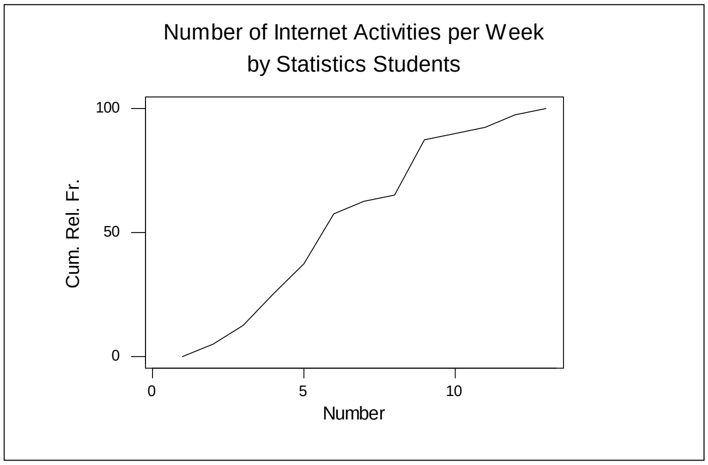
Histogram:



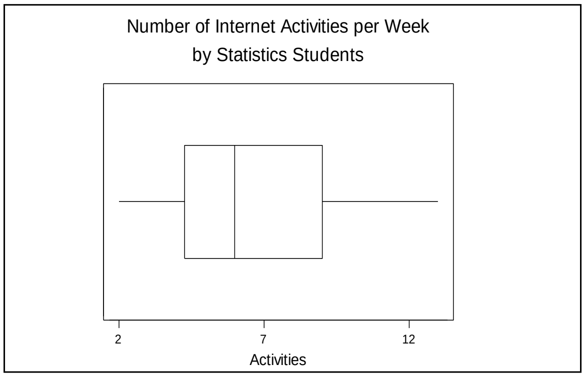
Ogive on Probability Paper


Normal Probability Plot for Activities
ML Estimates
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
0
5
10
Data
ML Estimates
Mean 6.675
StDev 2.80524
Goodness of Fit
AD* 1.069

```
Ogive:
```


Number of Internet Activities per Week
by Statistics Students
Cum. Rel. Fr.
100
50
0
0
5
10
Number

Box-and-whisker:



d. Answer will vary.

e. Mean = 6.675; median = 6; midrange = 7.5; mode = 9; midquartile = (4.5 + 9)/2 = 6.75 [Q_1=4.5; Q_3 = 9]

f. Range = 11; variance = 8.0712; standard deviation = 2.841

g. P_5 = (2+3)/2 = 2.5; P_{10}, = 3; Q_1 = 4.5; Q_3 = 9; P_{90} = (10+11)/2 = 10.5; P_{98} = 13

h. Answers will vary.

i. Answers will vary.

j. 6.675 – 1(2.841) = 3.834; 6.675 + 1(2.841) = 9.516; 30/40 = 0.75 or 75%, 75% is more than the expected 68%

k. 6.675 – 2(2.841) = 0.993; 6.675 + 2(2.841) = 12.357; 39/40 = 0.975 or 98%, 98% is at least 75%

l. Objective 2.1 data listed the different ways each person uses the Internet; for example: e-mail, surfing, etc. The data above is "the number of uses by each person."

Part 2 a. Answers will vary.

b. Answers will vary.

TECHNOLOGY CARD PRACTICE PROBLEMS

2.1. 2.1. a. Answers will vary since each random sample will yield different results.

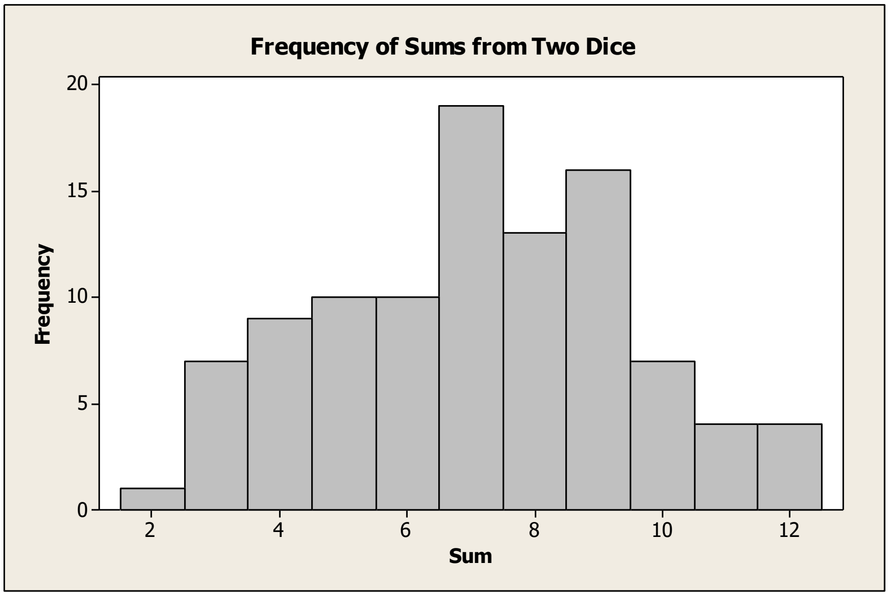
Sample answer:

(6, 2); 8	(6, 6); 12	(2, 1); 3	(6, 6); 12
(2, 1); 3	(3, 2); 5	(6, 4); 10	(4, 3); 7
(1, 6); 7	(2, 6); 8	(2, 5); 7	(1, 6); 7
(2, 1); 3	(3, 3); 6	(3, 4); 7	(6, 1); 7
(4, 5); 9	(3, 2); 5	(6, 3); 9	(2, 6); 8

(6, 5); 11	(6, 5); 11	(5, 4); 9	(4, 3); 7
(5, 4); 7	(6, 6); 12	(6, 4); 10	(1, 5); 6
(4, 2); 6	(2, 5); 7	(5, 6); 11	(1, 3); 4
(4, 6); 10	(6, 3); 9	(1, 3); 4	(3, 1); 4
(3, 1); 4	(4, 2); 6	(5, 4); 9	(5, 4); 9
(2, 3); 5	(5, 1); 6	(1, 2); 3	(4, 5); 9
(2, 4); 6	(3, 5); 8	(3, 5); 8	(2, 5); 7
(1, 2); 3	(1, 1); 2	(5, 3); 8	(3, 5); 8
(6, 2); 8	(1, 5); 6	(5, 1); 6	(5, 4); 9
(1, 4); 5	(6, 1); 7	(2, 6); 8	(5, 4); 9
(3, 2); 5	(1, 3); 4	(3, 6); 9	(5, 5); 10
(3, 5); 8	(1, 2); 3	(4, 3); 7	(3, 6); 9
(2, 2); 4	(5, 2); 7	(3, 2); 5	(5, 4); 9
(3, 4); 7	(3, 2); 5	(5, 2); 7	(5, 1); 6
(5, 4); 9	(6, 2); 8	(3, 2); 5	(4, 3); 7
(6, 1); 7	(3, 5); 8	(2, 5); 7	(6, 5); 11
(3, 1); 4	(4, 5); 9	(3, 4); 7	(6, 2); 8
(6, 6); 12	(6, 4); 10	(1, 4); 5	(1, 6); 7
(1, 5); 6	(6, 3); 9	(3, 2); 5	(1, 2); 3
(3, 1); 4	(5, 5); 10	(2, 3); 4	(5, 5); 10

b. Answers will vary. For above data set:

Sum	frequency
2	1
3	7
4	9
5	10
6	10
7	19
8	13
9	16
10	7
11	4
12	4



c. Answers will vary. Students should state that they expected this graph to be mound-shaped, as there are more outcomes that lead to sums of 7 than 6 or 8; of 6 or 8 than 5 or 9; and so on.