CHAPTER 2

Descriptive Statistics

2.1 a. The graph used is a bar chart.

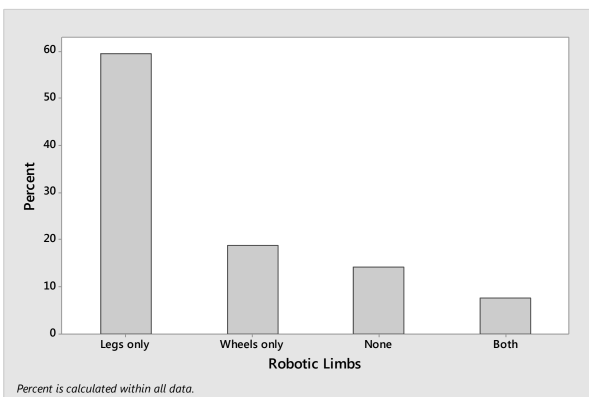
b. The variable measured is the type of robotic limbs on social robots.

c. The social robot design that is currently used the most is legs only.

d. The relative frequencies are found by dividing the frequencies by the sample size, $n = 106$.

Robotic Limbs	Frequency	Relative Frequency
None	15	$15/106 = 0.1415$
Both	8	$8/106 = 0.0755$
Legs only	63	$63/106 = 0.5943$
Wheels only	20	$20/106 = 0.1887$

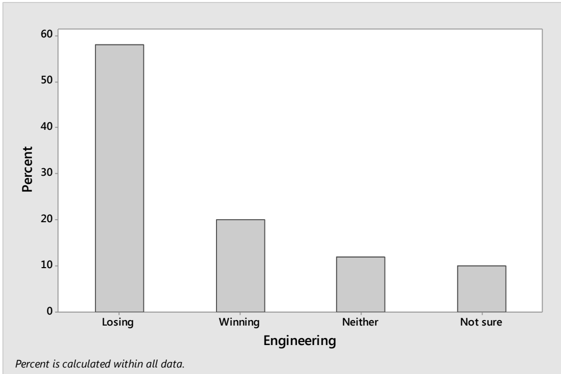
e. Using MINITAB, the Pareto chart is:



2.2 a. The variable described in the pie chart is the belief on whether the field of engineering is winning or losing young people. The classes are “Not sure”, “Losing young people”, “Winning young people”, and “Neither winning nor losing young people”.

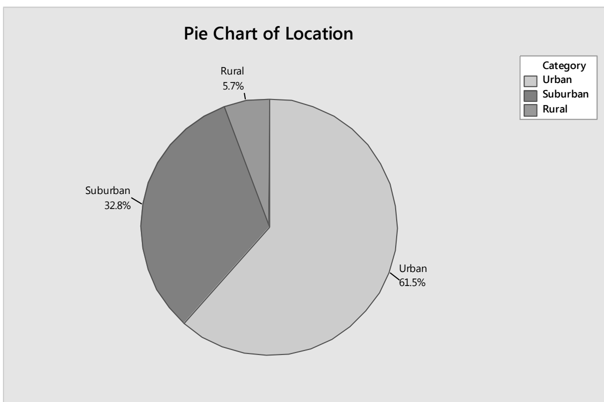
b. The 20% represents the percentage of the 808 American adults surveyed in January, 2009, who believe that the field of engineering is “Winning young people”.

c. Using MINITAB, the Pareto chart is:



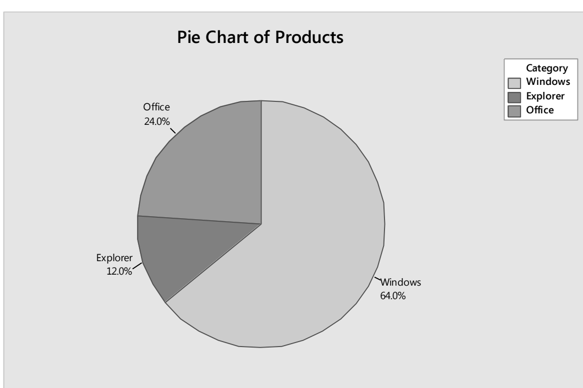
d. The majority opinion of American adults responding to the survey question is that the field of engineering is losing young people.

2.3 Using MINITAB, the pie chart is:



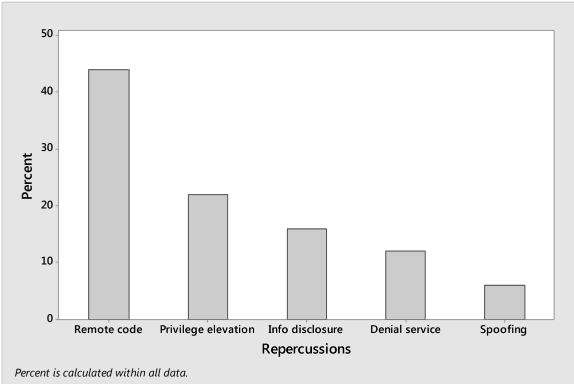
The majority of young women who recently participated in a STEM program are from urban areas (61.5%) and very few are from rural areas (5.7%).

2.4 a. Using MINITAB, a pie chart for the Microsoft products with security issues is:



Explorer had the lowest proportion of security issues in 2012.

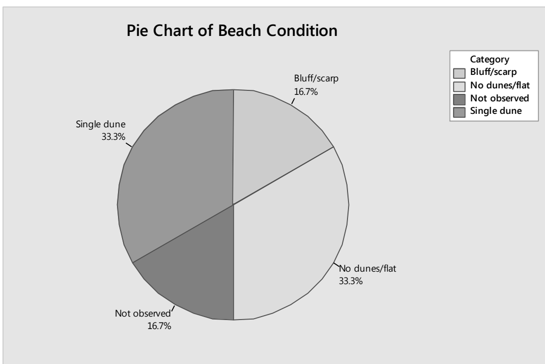
b. Using MINITAB, the Pareto chart for expected repercussions from security issues is:



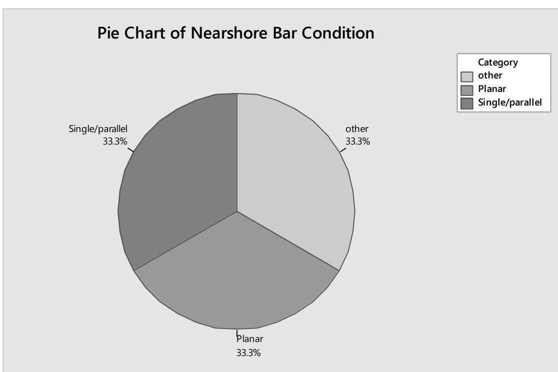
Because remote code execution has the highest proportion of observations, we would advise Microsoft to focus on remote code execution.

2.5 a. The variable beach condition is qualitative, nearshore bar condition is qualitative, and long-term erosion rate is quantitative.

b. Using MINITAB, the pie chart for beach condition is:

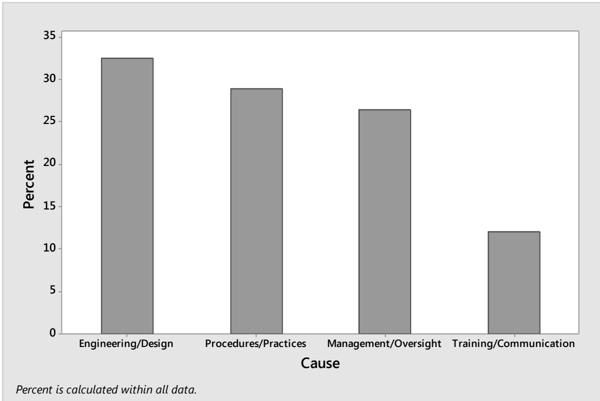


c. Using MINITAB, the pie chart of nearshore bar condition is:



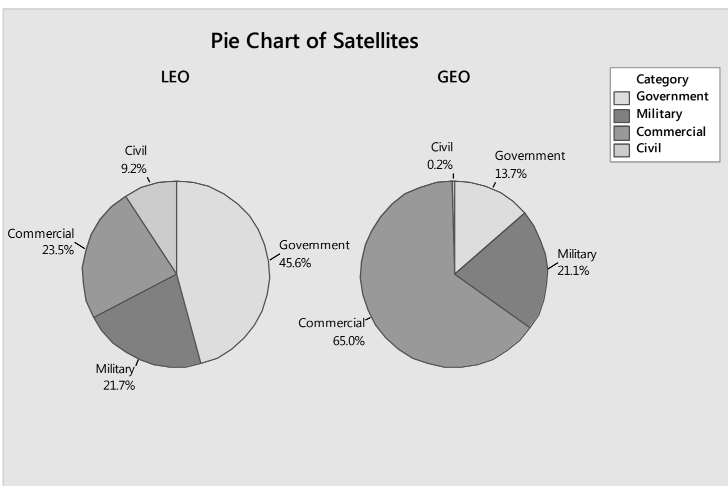
d. The sample size for this study is only 6. It would be very risky to use the information form this sample to make inferences about all beach hotspots. The data were collected using an online questionnaire. It is very doubtful that this sample is representative of the population of all beach hot spots.

2.6 Using MINITAB, the Pareto chart is:



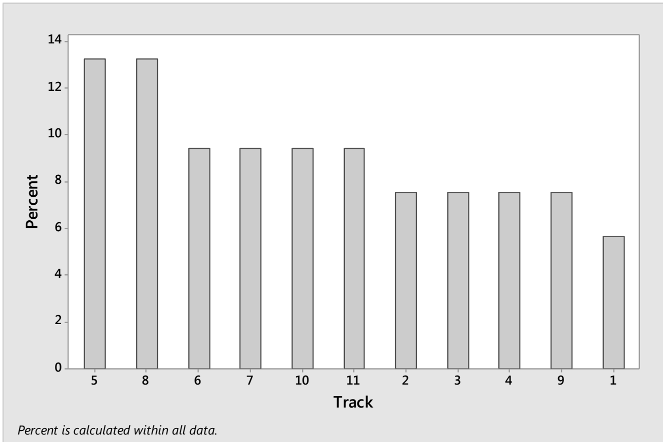
The three top categories (Engineering & Design, Procedures & practices, and Management & Oversight) all have similar relative frequencies and much higher than Training & Communication.

2.7 Using MINITAB, pie charts to compare the two ownership sectors of LEO and GEO satellites are:



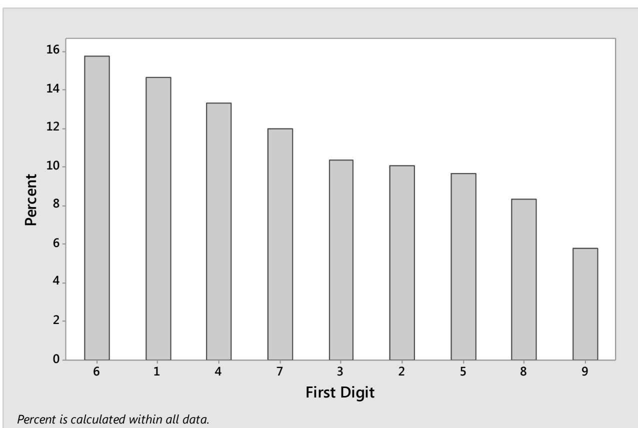
Most LEO satellites are owned by entities in the government (45.6%) while most GEO satellites are owned by entities in the commercial sector (65.0%). The fewest percentage of LEO satellites are owned by entities in the civil sector (9.2%). The fewest percentage of GEO satellites are also owned by entities in the civil sector (0.2%), but the percentage is much smaller than that for the LEO satellites.

2.8 Using MINITAB, the Pareto chart is:



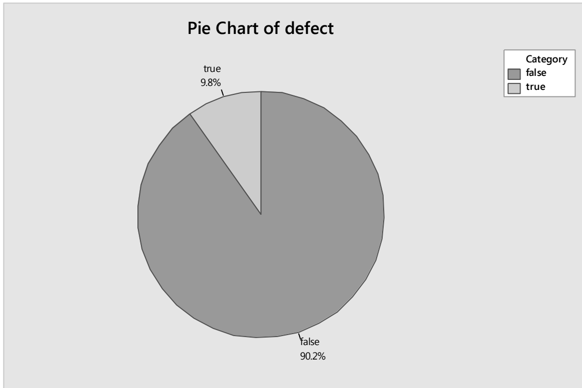
It appears that tracks 5 and 8 could be over-utilized and track 1 could be under-utilized.

2.9 a. Using MINITAB, the Pareto chart is:



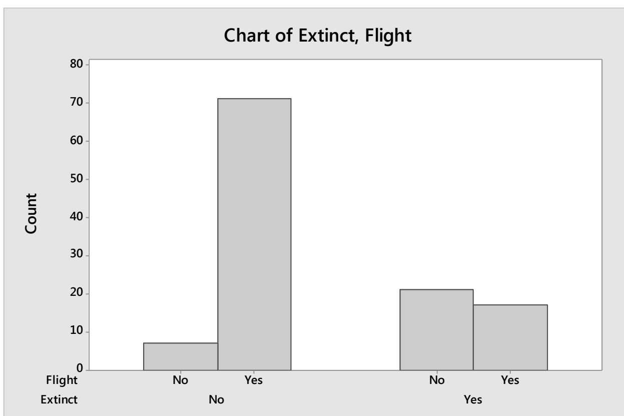
b. Yes and no. The graph does support Benford's Law in that certain digits are more likely to occur than others. In this set of data, the number 6 occurs first 15.7% of the time while the number 9 occurs first only 5.8% of the time. However, Benford's Law also states that the number 1 is the most likely to occur at 30% of the time. In this set of data, the number 1 is not the most frequent number to occur first, and it also only occurs as the first significant digit 14.7% of the time, not the 30% specified by Benford's Law.

2.10 Using MINITAB, the pie chart is:



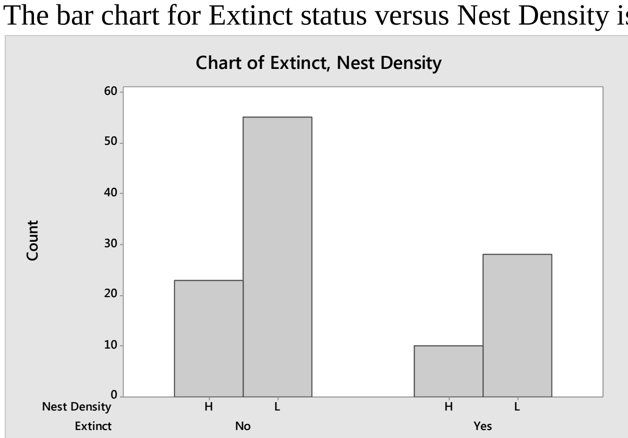
Only 9.8% of the software code is defective, while 90.2% of the software code is not defective.

2.11 Using MINITAB, a bar chart for the Extinct status versus flight capability is:



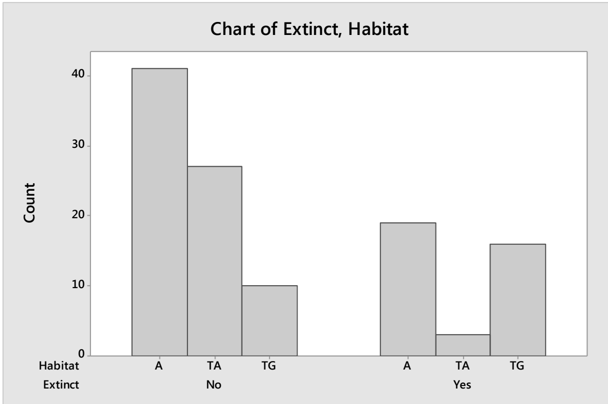
It appears that extinct status is related to flight capability. For birds that do have flight capability, most of them are present. For those birds that do not have flight capability, most are extinct.

The bar chart for Extinct status versus Nest Density is:



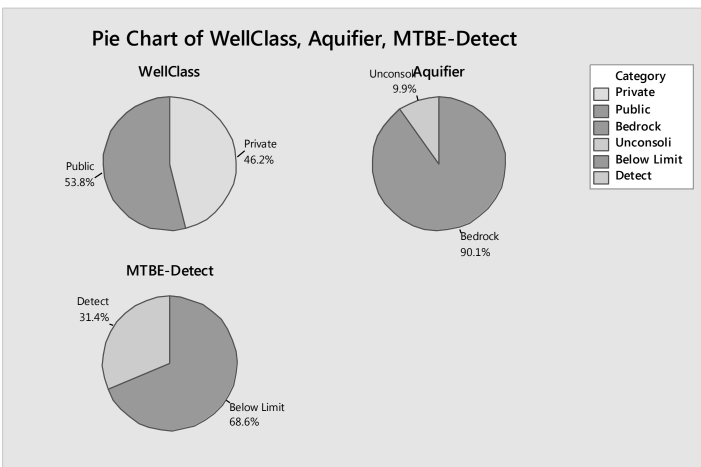
It appears that extinct status is not related to nest density. The proportion of birds present and extinct appears to be very similar for nest density high and nest density low.

The bar chart for Extinct status versus Habitat is:



It appears that the extinct status is related to habitat. For those in aerial terrestrial (TA), most species are present. For those in ground terrestrial (TG), most species are extinct. For those in aquatic, most species are present.

2.12 Using MINITAB, the pie charts for the three variables are:



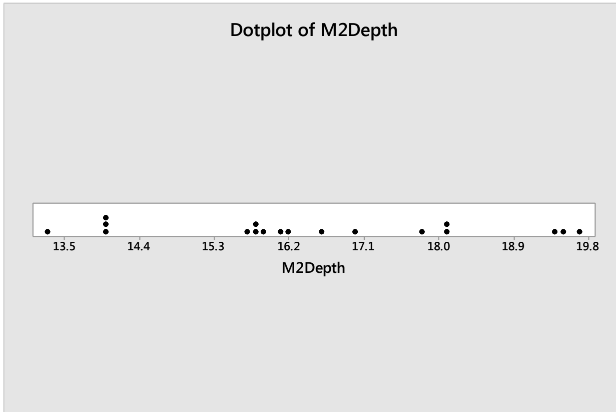
2.13 a. The measurement class 10-20 contains the highest proportion of respondents.

b. The approximate proportion of organizations that reported a percentage monetary loss from malicious insider actions less than 20% is $0.30 + 0.38 = 0.68$.

c. The approximate proportion of organizations that reported a percentage monetary loss from malicious insider actions greater than 60% is $0.07 + 0.025 + .035 + .045 = 0.175$.

d. The approximate proportion of organizations that reported a percentage monetary loss from malicious insider actions between 20% and 30% is 0.12. The actual number is approximately $0.12(144) = 17.28$ or approximately 17.

2.14 a. Using MINITAB, the dotplot is



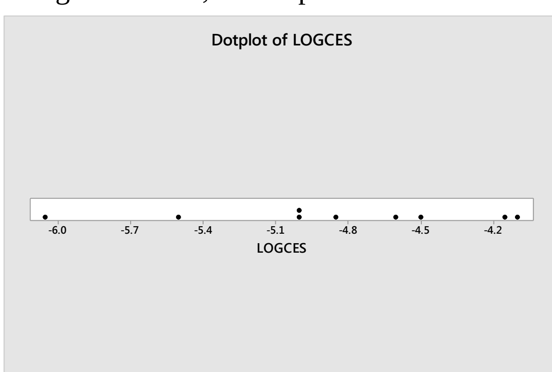
b. Using MINITAB, the stem-and-leaf display is:

Stem-and-Leaf Display: M2Depth

```
Stem-and-leaf of M2Depth  N  = 18
Leaf Unit = 0.10

 2   13  29
 4   14  00
 8   15  7789
(3)  16  125
 7   17  08
 5   18  11
 3   19  347
```

c. There is no actual depth that occurs more than once. However, there are 3 values that are close in value at 13.96, 14.02, and 14.04.

2.15 a. Using MINITAB, the dotplot is:

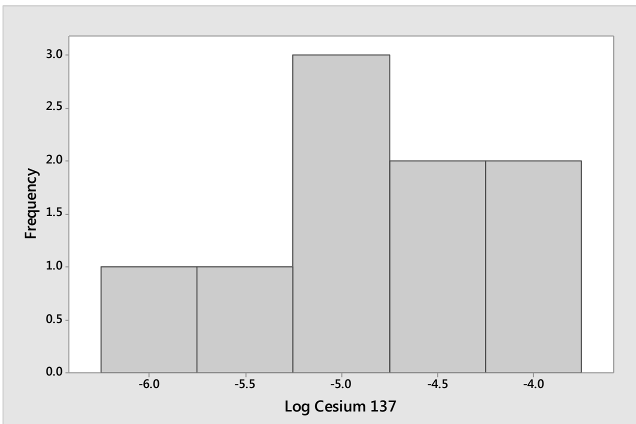


b. Using MINITAB, the stem-and-leaf display is:

Stem-and-Leaf Display: LOGCES

```
Stem-and-leaf of LOGCES  N  = 9
Leaf Unit = 0.10

 1   -6  0
 2   -5  5
 4   -5  00
(3)  -4  865
 2   -4  11
```

c. Using MINITAB, the histogram is:



d. Answers may vary. It appears that the histogram is more informative.

e. Four of the nine measurements are -5.00 or less. The proportion is $4/9 = 0.444$.

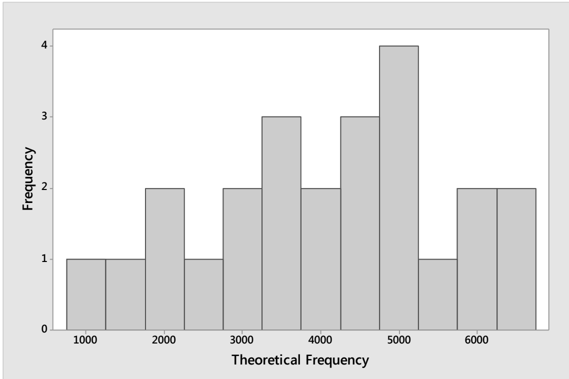
2.16 a. The graph displayed is a histogram.

b. The quantitative variable is the *fup/fumic* ratio.

c. The proportion of *fup/fumic* ratios that fall above 1 is $14/416 = 0.034$.

d. The proportion of *fup/fumic* ratios that fall below 0.4 is $289/416 = 0.695$.

2.17 Using MINITAB, a histogram of the sound frequencies is:



2.18 Using MINITAB, the stem-and-leaf display is:

Stem-and-Leaf Display: DIOXIDE

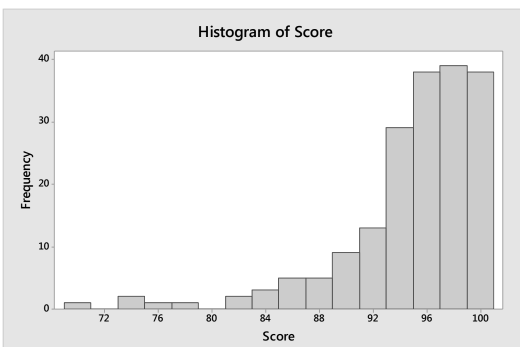
```
Stem-and-leaf of DIOXIDE  N  = 16
Leaf Unit = 0.10

 5   0  12234
 7   0  55
(2)  1  34
 7   1
 7   2  44
 5   2
 5   3  3
 4   3
 4   4  0000
```

The observations highlighted in gray are water specimens that contain oil. Since most of these values are low, there is a tendency for crude oil to be present in water with lower levels of dioxide.

2.19 a. Using MINITAB, the stem-and-leaf display and histogram are:

Stem-and-Leaf Display: Score

```
Stem-and-leaf of Score  N  = 186
Leaf Unit = 1.0

  1    6   9
  1    7
  2    7   3
  3    7   4
  4    7   6
  5    7   8
  7    8   11
  8    8   3
 11    8   445
 20    8   666677777
 25    8   99999
 36    9   00001111111
 54    9   222222333333333333
 89    9   44444444444444444555555555555555555
(42)   9   666666666666666666667777777777777777777777
 55    9   888888888888888889999999999999999999
 19   10   0000000000000000000
```



b. Of the 186 scores, only 11 are less than 86. Thus, there are 175 scores that are 86 or higher. The proportion of ships that have an accepted sanitation standard $175/186 = 0.941$. The stem-and-leaf display was used because one can identify the actual values.

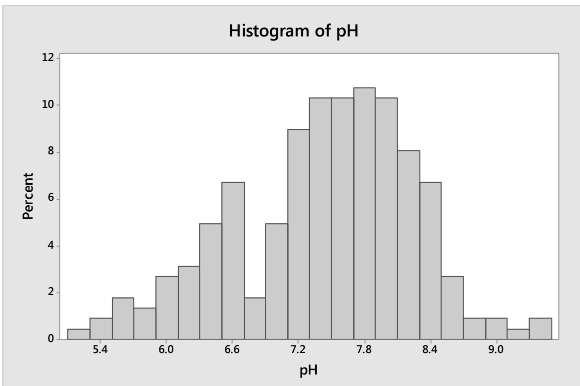
c. A score of 72 would be located above the number 72 on the histogram and in the 3rd row of the stem-and-leaf display.

2.20 Using MINITAB, the stem-and-leaf display is:

Stem-and-Leaf Display: ROUGH

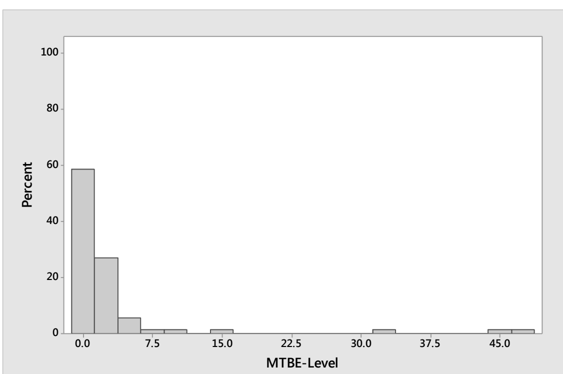
```
Stem-and-leaf of ROUGH  N  = 20
Leaf Unit = 0.10

 3   1  001
 5   1  22
 7   1  45
 8   1  7
 9   1  9
(5)  2  00111
 6   2  23
 4   2  455
 1   2  6
```

2.21 a. Using MINITAB, the histogram is:



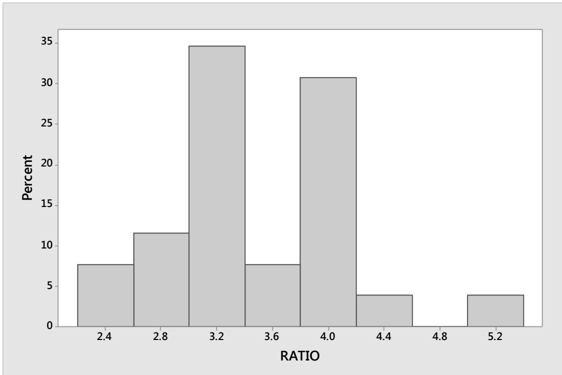
From the histogram, approximately 0.25 of the wells have pH values less than 7.0.

b. Using MINITAB, the histogram of the MTBE values for contaminated wells is:



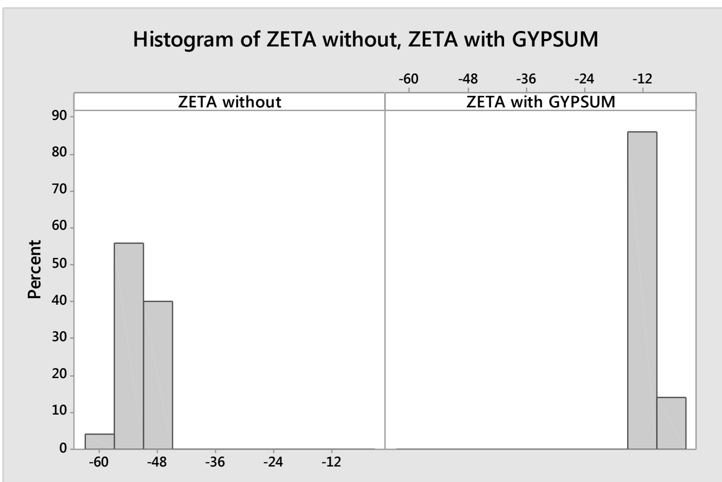
From the histogram, approximately 9% of the MTBE values exceed 5 micrograms per liter.

2.22 Using MINITAB, the histogram is:



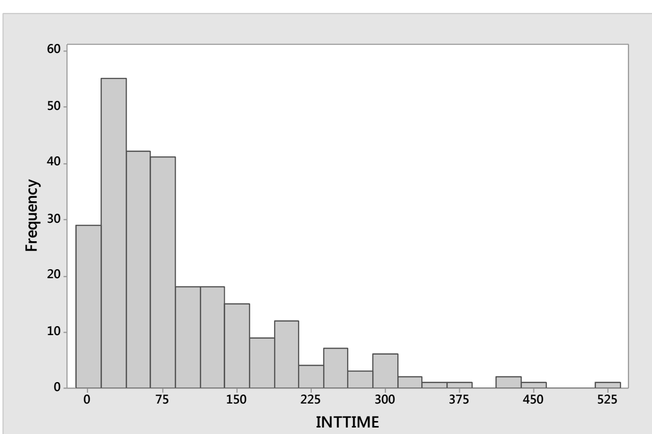
From the graph, about 0.05 of the till specimens have an Al/Be ratio that exceeds 4.5.

2.23 Using MINITAB, the histograms are:



The addition of calcium/gypsum increases the values of the zeta potential of silica. All of the values of zeta potential for the specimens containing calcium/gypsum are greater than all of the values of zeta potential for the specimens without calcium/gypsum.

2.24 Using MINTAB, the histogram is:



This histogram looks very similar to the one shown in the problem. Thus, it appears that there was minimal or no collaboration or collusion from within the company. We could conclude that the phishing attack against the organization was probably not an "inside job".

2.25 a. Assume the data are a sample. The mode is the observation that occurs most frequently. For this sample, there is no mode or all are modes.

The sample mean is:

$$\bar{y} = \frac{\sum y}{n} = \frac{4+3+10+8+5}{5} = \frac{30}{5} = 6$$

The median is the middle number when the data are arranged in order. The data arranged in order are: 3, 4, 5, 8, 10. The middle number is the 3rd number, which is $m = 5$.

b. Assume the data are a sample. The mode is the observation that occurs most frequently. For this sample, there are 2 modes, 4 and 6.

The sample mean is:

$$\bar{y} = \frac{\sum y}{n} = \frac{9+6+12+4+4+2+5+6}{8} = \frac{48}{8} = 6$$

The median is the middle number when the data are arranged in order. The data arranged in order are: 2, 4, 4, 5, 6, 6, 9, 12. The average of the middle 2 numbers is $m = \frac{5+6}{2} = 5.5$.

2.26 a. False. The mean base salary of all software engineering managers is $126,417. Some software engineering managers have salaries less than $126,417 and some have salaries greater than $126,417.

b. True or false. If the data are mound-shaped, then this statement is true because the mean and the median are the same for mound-shaped data. The mean base salary of manufacturing/production engineers is $92,360. If the data are mound-shaped, then the median is also $92,360 and half of the engineers would make less than $92,360. If the data are not mound-shaped or symmetric, then this statement is false.

c. False. It is possible that the lowest earning software engineering managers could make less than the highest earning manufacturing/production engineers.

2.27 a. The sample mean is $\bar{y} = \frac{\sum y}{n} = \frac{18.12+19.48+\cdots+16.20}{18} = \frac{296.99}{18} = 16.499$. The average dentary depth of molars is 16.499mm.

If the largest depth measurement were doubled, then the mean would increase.

b. The data arranged in order are:
13.25, 13.96, 14.02, 14.04, 15.70, 15.76, 15.83, 15.94, 16.12, 16.20, 16.55, 17.00, 17.83, 18.12, 18.13, 19.36, 19.48, 19.70
There is an even number of observations, so the median is the average of the middle two numbers, $m = \frac{16.12+16.20}{2} = 16.16$. Half of the observations are less than 16.16 and half are greater than 16.16.

If the largest depth measurement were doubled, then the median would not change.

c. Since no observation occurs more than once, there is either no mode or all of the observations are considered modes.

2.28 a. The sample mean is $\bar{y} = \frac{\sum y}{n} = \frac{(-5.50)+(-5.00)+\cdots+(-4.60)}{9} = \frac{-43.75}{9} = -4.861$.

The data arranged in order are: -6.05, -5.50, -5.00, -5.00, -4.85, -4.60, -4.50, -4.15, -4.10
There is an odd number of observations, so the median is the middle number or $m = -4.85$.

The mode is -5.00 since it occurs 2 times.

b. The average amount of radioactive element cesium-137 is -4.861.
The median amount of radioactive element cesium-137 is -4.85. Half of the amounts are less than -4.85 and half are greater than -4.85.
The mode amount of radioactive element cesium-137 is -5.00. This amount occurs more than any other.

2.29 The sample mean is $\bar{y} = \frac{\sum y}{n} = \frac{10.94+13.71+\cdots+6.77}{13} = \frac{126.32}{13} = 9.717$. The average rebound length is 9.717 meters.

The data arranged in order are: 4.90, 5.10, 5.44, 5.85, 6.77, 7.26, 10.94, 11.38, 11.87, 11.92, 13.35, 13.71, 17.83
There is an odd number of observations so the median is the middle number or $m = 10.94$.
Half of the rebound lengths are less than 10.94 and half are greater than 10.94.

2.30 a. The sample mean is $\bar{y} = \frac{\sum y}{n} = \frac{1.53+1.50+\cdots+1.48}{8} = \frac{11.77}{8} = 1.471$.

b. The data arranged in order are: 1.37, 1.41, 1.42, 1.48, 1.50, 1.51, 1.53, 1.55. There is an even number of observations so the median is the average of the middle two numbers or $m = \frac{1.48+1.50}{2} = 1.49$.

c. The average daily ammonia level in air in the tunnel is 1.471 parts per million.
The median daily ammonia level in air in the tunnel is 1.49. Half of all observations are less than 1.49 and half are greater than 1.49.

2.31 a. The sample mean is $\bar{y} = \frac{\sum y}{n} = \frac{3.3+0.5+\cdots+4.0}{16} = \frac{29}{16} = 1.813$.

b. The data arranged in order are: 0.1, 0.2, 0.2, 0.3, 0.4, 0.5, 0.5, 1.3, 1.4, 2.4, 2.4, 3.3, 4.0, 4.0, 4.0, 4.0. There is an even number of observations so the median is the average of the middle two numbers or $m = \frac{1.3+1.4}{2} = 1.35$.

c. The mode is the number that occurs the most which is 4.0.

d. The data arranged in order for the no crude oil present are: 0.1, 0.3, 1.4, 2.4, 2.4, 3.3, 4.0, 4.0, 4.0, 4.0. There is an even number of observations so the median is the average of the middle two numbers or $m = \frac{2.4 + 3.3}{2} = 2.85$.

e. The data arranged in order for the crude oil present are: 0.2, 0.2, 0.4, 0.5, 0.5, 1.3. There is an even number of observations so the median is the average of the middle two numbers or $m = \frac{0.4 + 0.5}{2} = 0.45$.

d. The median dioxide amount for no crude oil present is 2.85, while the median dioxide amount for crude oil present is 0.45. It appears that dioxide amount is less when crude oil is present.

2.32 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: Score

```
                                         N for
Variable    N    Mean  Median  Mode   Mode
Score     186  94.441  96.000    97     22
```

The mean sanitation score is 94.441. The median sanitation score is 96.00. Half of the sanitation scores are less than 96.00 and half are greater than 96.00. The mode is 97. More ships had a sanitation score of 97 than any other number.

2.33 a. The average permeability measurement for Group A sandstone is 73.62. The median permeability for Group A is 70.45. Half of the permeability measurements for Group A are less than 70.45 and half are greater than 70.45.

b. The average permeability measurement for Group B sandstone is 128.54. The median permeability for Group B is 139.30. Half of the permeability measurements for Group B are less than 139.30 and half are greater than 139.30.

c. The average permeability measurement for Group C sandstone is 83.07. The median permeability for Group C is 78.65. Half of the permeability measurements for Group C are less than 78.65 and half are greater than 78.65.

d. The mode for Group C is 70.9. Three observations were 70.9. The permeability measurement that occurred the most often from Group C is 70.9.

e. Group B appears to result in faster decay because all three measures of central tendency for Group B are larger than the corresponding measures for Group C.

2.34 a. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: ZETA without

Variable	N	Mean	Median	Mode	N for Mode
ZETA without	50	-52.070	-52.250	-50.2	3

The average zeta potential measurement for liquid solutions prepared without calcium/gypsum is -52.07. The median zeta potential measurement for liquid solutions prepared without calcium/gypsum is -52.25. Half of the zeta potential measurements for liquid solutions prepared without calcium/gypsum are below -52.25 and half are above -52.25. The mode zeta potential measurement for liquid solutions prepared without calcium/gypsum is -50.2. The zeta potential measurements for liquid solutions prepared without calcium/gypsum that occurred the most is -50.2.

b. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: ZETA with GYPSUM

Variable	N	Mean	Median	Mode	N for Mode
ZETA with GYPSUM	50	-10.958	-11.300	-11.3	5

The average zeta potential measurement for liquid solutions prepared with calcium/gypsum is -10.958. The median zeta potential measurement for liquid solutions prepared with calcium/gypsum is -11.30. Half of the zeta potential measurements for liquid solutions prepared with calcium/gypsum are below -11.30 and half are above -11.30. The mode zeta potential measurement for liquid solutions prepared with calcium/gypsum is -11.3. The zeta potential measurements for liquid solutions prepared with calcium/gypsum that occurred the most is -11.3.

c. By adding calcium/gypsum to the solution, the zeta potential measurements increase. All measures of central tendency for the zeta potential measurements are greater when calcium/gypsum are added than when calcium/gypsum is not added.

2.35 a. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: PRDiff_outlier

Variable	N	Mean	Median	Mode	N for Mode
PRDiff_outlier	14	-1.091	-0.655	*	0

The average difference is -1.091. The median difference is -0.655. Half of the differences are less than -0.655 and half of the differences are greater than -0.655. No difference occurs more than once, so there is no mode.

b. The one large difference is -8.11.

c. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: PRDiff

Variable	N	Mean	Median	Mode	N for Mode
PRDiff	14	-0.519	-0.520	*	0

The mean increases from -1.091 to -0.519 or increases by 0.572. The median increases from -0.655 to -0.520 or increases by 0.135. The mean is much more affected by correcting the outlier than the median.

2.36 a. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: PLANTS

```
                                           N for
Variable   N   Mean  Median  Mode   Mode
PLANTS    20  3.950   3.500     1      5
```

b. Using MINITAB and deleting the largest observations, the descriptive statistics are:

Descriptive Statistics: PLANTS_minus

```
                                               N for
Variable       N   Mean  Median  Mode   Mode
PLANTS_minus  19  3.579   3.000     1      5
```

By dropping the largest observation, the mean has been reduced from 3.950 to 3.579. The median has dropped from 3.50 to 3.0. The mode did not change.

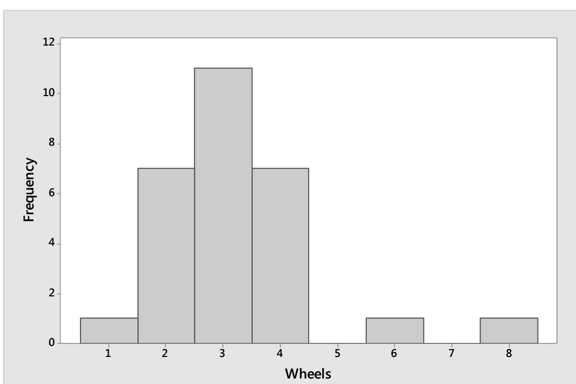
c. Using MINITAB and deleting the 2 largest and smallest observations, the descriptive statistics are:

Descriptive Statistics: PLANTS_trim

```
                                              N for
Variable      N   Mean  Median  Mode   Mode
PLANTS_trim  16  3.563   3.500     3      4
```

The trimmed mean is not affected by very large or very small observations and gives a mean of "most" of the observations without the outliers.

2.37 a. Using MINITAB, the histogram of the data is:



No, the distribution is somewhat mound-shaped but it is not symmetric. The distribution is skewed to the right.

b. The sample mean is $\bar{y} = \frac{\sum y}{n} = \frac{4+4+\cdots+2}{28} = \frac{90}{28} = 3.214$.

The sample variance is $s^2 = \dfrac{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}{n-1} = \dfrac{340 - \dfrac{90^2}{28}}{28-1} = \dfrac{50.7143}{27} = 1.8783$.

The sample standard deviation is $s = \sqrt{1.8783} = 1.371$.

c. $\bar{y} \pm 2s \Rightarrow 3.214 \pm 2(1.371) \Rightarrow 3.214 \pm 2.742 \Rightarrow (0.472,\ 5.956)$

d. According to Chebyshev's rule, at least ¾ or 75% of the observations will fall in this interval.

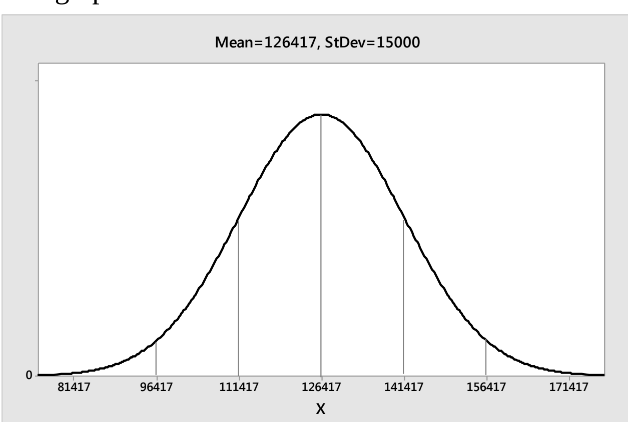
e. According to the Empirical Rule, approximately 95% of the observations will fall in this interval.

f. The actual proportion of observations that fall in the interval is $26/28 = 0.929$ or 92.9%. Yes, the Empirical Rule provides a good estimate of the proportion even though the distribution is not perfectly symmetric.

2.38 The standard deviation is $s = \sqrt{225{,}000{,}000} = 15{,}000$. The intervals are:

$$\bar{y} \pm s \Rightarrow \$126{,}417 \pm \$15{,}000 \Rightarrow (\$111{,}417,\ \ \$141{,}417)$$
$$\bar{y} \pm 2s \Rightarrow \$126{,}417 \pm 2(\$15{,}000) \Rightarrow \$126{,}417 \pm \$30{,}000 \Rightarrow (\$96{,}417,\ \ \$156{,}417)$$
$$\bar{y} \pm 3s \Rightarrow \$126{,}417 \pm 3(\$15{,}000) \Rightarrow \$126{,}417 \pm \$45{,}000 \Rightarrow (\$81{,}417,\ \ \$171{,}417)$$

The graph of the distribution is:



Approximately 68% of software engineering managers have salaries between \$111,417 and \$141,417.
Approximately 95% of software engineering managers have salaries between \$96,417 and \$156,417.
Approximately 100% of software engineering managers have salaries between \$811,417 and \$171,417.

2.39 a. The range is $R = 1.55 - 1.37 = 0.18$.

b. The sample variance is $s^2 = \dfrac{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}{n-1} = \dfrac{17.3453 - \dfrac{11.77^2}{8}}{8-1} = \dfrac{0.0286875}{7} = 0.00410$.

c. The sample standard deviation is $s = \sqrt{0.00410} = 0.0640$.

d. The standard deviation for the morning is 1.45 ppm, while the standard deviation for the afternoon is 0.0640. The morning drive-time has more variable ammonia levels.

2.40 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: ZETA without, ZETA with GYPSUM

```
Variable           N     Mean  StDev
ZETA without      50  -52.070  2.721
ZETA with GYPSUM  50  -10.958  1.559
```

a. The standard deviation for the zeta potential measurements of the liquid solutions without calcium/gypsum is 2.721. The interval that would contain approximately 95% of the measurements is $\bar{y} \pm 2s \Rightarrow -52.07 \pm 2(2.721) \Rightarrow -52.07 \pm 5.442 \Rightarrow (-57.512, -46.628)$.

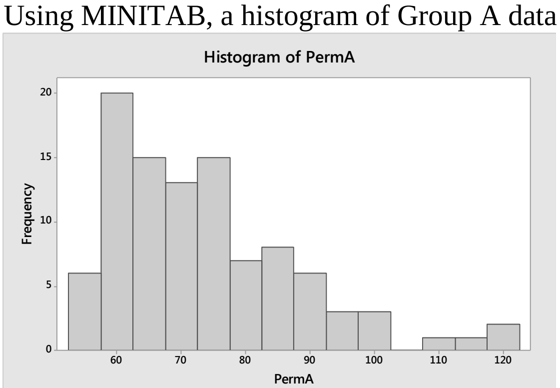
b. The standard deviation for the zeta potential measurements of the liquid solutions with calcium/gypsum is 1.559. The interval that would contain approximately 95% of the measurements is $\bar{y} \pm 2s \Rightarrow -10.958 \pm 2(1.559) \Rightarrow -10.958 \pm 3.118 \Rightarrow (-14.076, -7.840)$.

c. Because the intervals do not overlap, there is evidence to indicate that adding calcium/gypsum to the liquid solution impacts the flotation property of silica.

2.41 a. The range for Group A is 67.20. $R = 122.40 - 55.20 = 67.20$

b. The standard deviation for Group A is 14.48. $s = \sqrt{209.53} = 14.48$

c. Using MINITAB, a histogram of Group A data is:



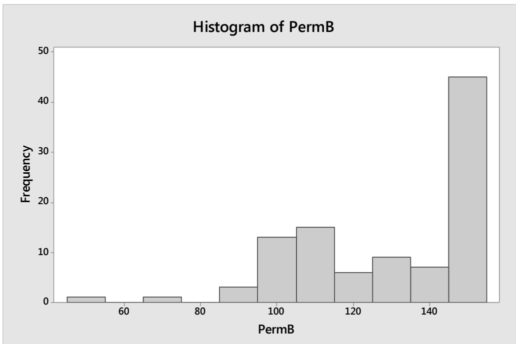
From Exercise 2.33, the mean is 73.62. Because the data are skewed to the right, we will use Chebyshev's rule. At least 8/9 or 88.9% of the observations will fall within 3 standard deviations of the mean. This interval is

$\bar{y} \pm 3s \Rightarrow 73.62 \pm 3(14.48) \Rightarrow 73.62 \pm 43.44 \Rightarrow (30.18,\ 117.06)$. Thus, at least 88.8% of the measurements for Group A will fall between 30.18 and 117.06.

d. The range for Group B is 99.60. $R = 150.00 - 50.40 = 99.60$

The standard deviation for Group B is 21.97. $s = \sqrt{482.75} = 21.97$

Using MINITAB, a histogram of Group B data is:

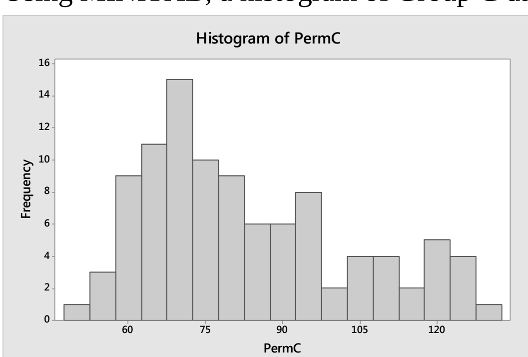


From Exercise 2.33, the mean is 128.54. Because the data are skewed to the left, we will use Chebyshev's rule. At least 8/9 or 88.9% of the observations will fall within 3 standard deviations of the mean. This interval is
$\bar{y} \pm 3s \Rightarrow 128.54 \pm 3(21.97) \Rightarrow 128.54 \pm 65.91 \Rightarrow (62.63,\ 194.45)$. Thus, at least 88.8% of the measurements for Group B will fall between 62.63 and 194.45.

e. The range for Group C is 76.80. $R = 129.00 - 52.20 = 76.80$

The standard deviation for Group C is 20.05. $s = \sqrt{401.94} = 20.05$

Using MINITAB, a histogram of Group C data is:



From Exercise 2.33, the mean is 83.07. Because the data are skewed to the left, we will use Chebyshev's rule. At least 8/9 or 88.9% of the observations will fall within 3 standard deviations of the mean. This interval is
$\bar{y} \pm 3s \Rightarrow 83.07 \pm 3(20.05) \Rightarrow 83.07 \pm 60.15 \Rightarrow (22.92,\ 143.22)$. Thus, at least 88.8% of the measurements for Group C will fall between 22.92 and 143.22.

f. From all of the analyses, Group B appears to result in higher permeability measurements.

The interval of the $\bar{y} \pm 3s$ for Group B is shifted to the right of that for Group C. Also, the histogram for Group B is skewed to the left, while the histogram for Group C is skewed to the right. Most of the observations for Group B are to the right of the observations from Group C.

2.42 a. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: PLANTS

```
Variable   N  StDev  Variance   Range
PLANTS    20  2.743     7.524  10.000
```

The range is 10.0, the variance is 7.524, and the standard deviation is 2.743.

b. Using MINITAB and eliminating the largest observation, the descriptive statistics are:

Descriptive Statistics: PLANTS_minus

```
Variable       N  StDev  Variance  Range
PLANTS_minus  19  2.244     5.035  8.000
```

The range is 8.0, the variance is 5.035, and the standard deviation is 2.244. By eliminating the largest observation, the range decreases from 10.0 to 8.0, the variance decreases from 7.524 to 5.035, and the standard deviation decreases from 2.743 to 2.244.

c. Using MINITAB and eliminating both the largest and smallest observations, the descriptive statistics are:

Descriptive Statistics: PLANTS_minus2

```
Variable        N  StDev  Variance  Range
PLANTS_minus2  18  2.218     4.918  8.000
```

The range is 8.0, the variance is 4.918, and the standard deviation is 2.218. By eliminating the largest and smallest observations, the range decreases from 10.0 to 8.0, the variance decreases from 7.524 to 4.918, and the standard deviation decreases from 2.743 to 2.218.

2.43 a. If we assume that the distributions of scores are mound-shaped, then we know that approximately 95% of all observations are within 2 standard deviations of the mean. For flexed arms, this interval is $\bar{y} \pm 2s \Rightarrow 59 \pm 2(4) \Rightarrow 59 \pm 8 \Rightarrow (51, 67)$. For extended arms, this interval is $\bar{y} \pm 2s \Rightarrow 43 \pm 2(2) \Rightarrow 43 \pm 4 \Rightarrow (39, 47)$. Since these intervals do not overlap, the scores for those with extended arms tend to be smaller than those with flexed arms. Thus, this supports the researchers' theory.

b. Changing the standard deviations: The interval for flexed arms is $\bar{y} \pm 2s \Rightarrow 59 \pm 2(10) \Rightarrow 59 \pm 20 \Rightarrow (39, 79)$. The interval for extended arms is $\bar{y} \pm 2s \Rightarrow 43 \pm 2(15) \Rightarrow 43 \pm 30 \Rightarrow (13, 73)$. Since these intervals significantly overlap, there is no evidence to support the researchers' theory.

2.44 The mean and standard deviation for defective ("true") modules are $\bar{y} = 61.51$ and $s = 67.49$.

The mean and standard deviation for nondefective ("false") modules are $\bar{y} = 26.17$ and $s = 37.64$. In both cases, the standard deviations are greater than the means, so the distributions are skewed to the right. Using Chebyshev's rule, we know that at least ¾ of the observations will be within 2 standard deviations of the mean. For the defective modules, this interval is $\bar{y} \pm 2s \Rightarrow 61.51 \pm 2(67.49) \Rightarrow 61.51 \pm 134.98 \Rightarrow (-73.47, 196.49)$ or $(0, 196.49)$. For the nondefective modules, this interval is $\bar{y} \pm 2s \Rightarrow 26.17 \pm 2(37.64) \Rightarrow 26.17 \pm 75.28 \Rightarrow (-49.11, 101.45)$ or $(0, 101.45)$. In general, the nondefective modules tend to have fewer lines of code than the defective modules.

2.45 a. For the private wells, $\bar{y} = 1.00$ and $s = 0.950$. Assuming that the distribution is approximately mound-shaped, approximately 95% of the observations will be within 2 standard deviations of the mean. This interval is $\bar{y} \pm 2s \Rightarrow 1.00 \pm 2(0.95) \Rightarrow 1.00 \pm 1.90 \Rightarrow (-0.90, 2.90)$.

b. For the public wells, $\bar{y} = 4.56$ and $s = 10.39$. Assuming that the distribution is approximately mound-shaped, approximately 95% of the observations will be within 2 standard deviations of the mean. This interval is $\bar{y} \pm 2s \Rightarrow 4.56 \pm 2(10.39) \Rightarrow 4.56 \pm 20.78 \Rightarrow (-16.22, 25.34)$.

2.46 The interval would be $\bar{y} \pm 2s \Rightarrow 19.5 \pm 2(4.7) \Rightarrow 19.5 \pm 9.4 \Rightarrow (10.1, 28.9)$. Since we know that approximately 95% of all the observation will be between 10.1 and 28.9, a SNR value of 30 would be unusual because it does not fall in the above interval.

2.47 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: Strength

Variable	N	Mean	StDev	Minimum	Median	Maximum
Strength	10	234.74	9.91	215.70	234.55	248.80

The mean and standard deviation are $\bar{y} = 234.74$ and $s = 9.91$. Assuming that the data are approximately mound-shaped and symmetric, the interval of the mean plus or minus two standard deviations is $\bar{y} \pm 2s \Rightarrow 234.74 \pm 2(9.91) \Rightarrow 234.74 \pm 19.82 \Rightarrow (214.92, 254.56)$. Approximately 95% of all the observations will be between 214.92 and 254.56.

2.48 a. Assuming that the distribution of velocities at 15 feet is approximately mound-shaped and symmetric, then approximately 95% of all observations will be within 2 standard deviations of the mean. This interval will be $\bar{y} \pm 2s \Rightarrow 936 \pm 2(10) \Rightarrow 936 \pm 20 \Rightarrow (913, 956)$. Approximately 95% of all observations will be between 913 and 956 feet per second.

b. Because an observation of 1000 is not in the above interval, it would not be a likely value for a bullet manufactured by Winchester. The bullet is not likely to be manufactured by Winchester.

2.49 a. From the histogram, the approximate 30th percentile would be 10%.

b. From the histogram, the approximate 95th percentile would be 90%.

2.50 a. From Exercise 2.16, the proportion of *fup/fumic* ratios that fall above 1 is $14/416 = 0.034$. Thus, 1 is the $100 - 3.4 = 96.6^{th}$ percentile.

b. From Exercise 2.16, the proportion of *fup/fumic* ratios that fall below 0.4 is $289/416 = 0.695$. Thus 0.4 is the 69.5th percentile.

2.51 a. Using the Empirical Rule, the 84th percentile would correspond to 1 standard deviation above the mean. Thus, the 84th percentile would be $\$126,417 + \$15,000 = \$141,417$.

b. Using the Empirical Rule, the 2.5th percentile would correspond to 2 standard deviations below the mean. Thus, the 2.5^{h} percentile would be $\$126,417 - 2(\$15,000) = \$126,417 - \$30,000 = \$96,417$.

c. $z = \dfrac{y-\mu}{\sigma} = \dfrac{\$100,000 - \$126,417}{\$15,000} = -1.76$

2.52 The 75th percentile score of 10 micrograms means that 75% of the Everglade sites had total phosphorus (TP) levels less than or equal to 10 micrograms. This level was probably selected because most (75%) of the sites had values less than this. Therefore, only 25% of the sites would have values greater than 10 micrograms.

2.53 a. In the text, it was given that the mean number of sags is 353 and the standard deviation of the number of sags is 30. The *z*-score for 400 sags is $z = \dfrac{y-\mu}{\sigma} = \dfrac{400-353}{30} = 1.57$. A value of 400 sags is 1.57 standard deviations above the mean number of sags.

b. In the text, it was given that the mean number of swells is 184 and the standard deviation of the number of swells is 25. The *z*-score for 100 swells is $z = \dfrac{y-\mu}{\sigma} = \dfrac{100-184}{25} = -3.36$. A value of 100 swells is 3.36 standard deviations below the mean number of swells. This would be a very unusual value to observe.

2.54 a. To find the 10th percentile, we use the formula $i = p(n+1)/100) = 10(115+1)/100 = 11.6$. Thus, the 10th percentile is approximately the 12th observation, once the data have been arranged in order. The 10th percentile is 21. This means that 10% of aa egg lengths are less than or equal to 21 millimeters and 90% are greater than or equal to 21 millimeters.

b. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: Egg Length

```
Variable      N   Mean  StDev
Egg Length  115  61.06  45.46
```

The *z*-score for an egg length of 205 millimeters is $z = \dfrac{y-\bar{y}}{s} = \dfrac{205-61.06}{45.46} = 3.17$. An egg length of 205 millimeters is 3.17 standard deviations above the mean egg length.

2.55 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: Score

```
Variable    N    Mean  StDev
Score     186  94.441  5.335
```

a. The z-score for the *Nautilus Explorer*'s score of 74 is $z=\dfrac{y-\bar{y}}{s}=\dfrac{74-94.441}{5.335}=-3.83$.
The *Nautilus Explorer*'s score of 74 is 3.83 standard deviations below the mean sanitation score.

b. The z-score for the *Rotterdam*'s score of 86 is $z=\dfrac{y-\bar{y}}{s}=\dfrac{86-94.441}{5.335}=-1.58$. The *Rotterdam*'s score of 86 is 1.58 standard deviations below the mean sanitation score.

2.56 No. The 90th percentile of the study sample had a lead concentration of 0.00372. Thus, more than 90% of the study sample had lead concentrations less than the EPA *Action Level* of 0.015. The water is safe.

2.57 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: ZETA without, ZETA with GYPSUM

```
Variable           N     Mean  StDev
ZETA without      50  -52.070  2.721
ZETA with GYPSUM  50  -10.958  1.559
```

a. The z-score for a zeta potential measurement for solutions prepared without calcium/gypsum of -9.0 is $z=\dfrac{y-\bar{y}}{s}=\dfrac{-9.0-(-52.07)}{2.721}=15.83$

b. The z-score for a zeta potential measurement for solutions prepared with calcium/gypsum of -9.0 is $z=\dfrac{y-\bar{y}}{s}=\dfrac{-9.0-(-10.958)}{1.559}=1.26$.

c. The solution prepared with calcium/gypsum is more likely to have a zeta potential measurement of -9.0 because the z-score of 1.26 is reasonable. The z-score of 15.83 is highly unlikely.

2.58 The z-score for a salary of \$180,000 is $z=\dfrac{y-\mu}{\sigma}=\dfrac{\$180,000-\$126,417}{\$15,000}=3.57$. If the distribution is mound-shaped and symmetric, the Empirical Rule says that essentially all of the observations will be within 3 standard deviations of the mean. A salary of \$180,000 is more than 3 standard deviations from the mean. The claim is not believable.

2.59 a. The median is $m=170$. Half of the clinkers had barium content less than or equal to 170 mg/kg and half of the clinkers had barium content greater than or equal to 170 mg/kg.

b. $Q_L=115$. 25% of the clinkers had barium content less than or equal to 115 mg/kg and

75% of the clinkers had barium content greater than or equal to 115 mg/kg.

c. $Q_U = 260$. 75% of the clinkers had barium content less than or equal to 260 mg/kg and 25% of the clinkers had barium content greater than or equal to 260 mg/kg.

d. $IQR = Q_U - Q_L = 260 - 115 = 145$.

e. Lower inner fence $= Q_L - 1.5(IQR) = 115 - 1.5(145) = -102.5$.
Upper inner fence $= Q_U + 1.5(IQR) = 260 + 1.5(145) = 477.5$.

f. Because no clinkers had barium content levels beyond the inner fences, there is no evidence of outliers.

2.60 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: REB-LENGTH

```
Variable     N  Minimum    Q1  Median     Q3  Maximum
REB-LENGTH  13     4.90  5.64   10.94  12.64    17.83
```

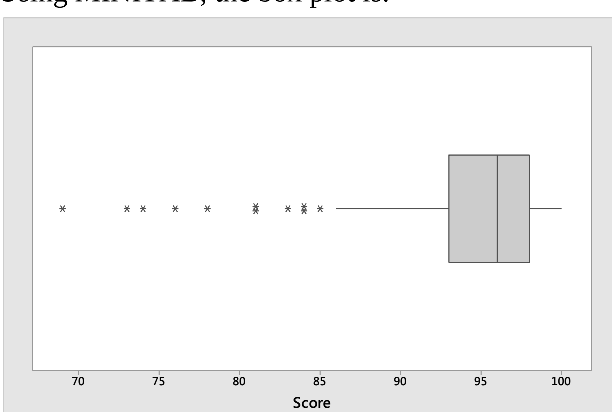
$IQR = Q_U - Q_L = 12.64 - 5.76 = 7.0$
The lower inner fence $= Q_L - 1.5(IQR) = 5.64 - 1.5(7.0) = -4.86$.
The upper inner fence $= Q_U + 1.5(IQR) = 12.64 + 1.5(7.0) = 23.14$.

Since no observation falls beyond the inner fences, there is no evidence of outliers.

2.61 a. The z-score for 400 sags is $z = \frac{y-\mu}{\sigma} = \frac{400-353}{30} = 1.57$. We would not consider 400 sags to be unusual because the z-score is less than 2.

b. The z-score for 100 swells is $z = \frac{y-\mu}{\sigma} = \frac{100-184}{25} = -3.36$. We would consider 100 swells to be unusual because it is more than 3 standard deviations fro the mean.

2.62 a. Using MINITAB, the box plot is:



Using MINITAB, the descriptive statistics are:

Descriptive Statistics: Score

```
Variable    N    Mean  StDev  Minimum      Q1  Median      Q3  Maximum
Score     186  94.441  5.335   69.000  93.000  96.000  98.000  100.000
```

$IQR = Q_U - Q_L = 98 - 93 = 5$

The lower inner fence $= Q_L - 1.5(IQR) = 93 - 1.5(5.0) = 85.5$.

The upper inner fence $= Q_U + 1.5(IQR) = 98 + 1.5(5.0) = 105.5$.

Because there are many observations (11) below the lower fence (stars), there is evidence of several outliers.

b. The *z*-score for 85 is $z = \dfrac{y - \bar{y}}{s} = \dfrac{85 - 94.441}{5.335} = -1.77$. A score of 85 would not be considered an outlier.

The *z*-score for 84 is $z = \dfrac{y - \bar{y}}{s} = \dfrac{84 - 94.441}{5.335} = -1.96$. A score of 84 would not be considered an outlier.

The *z*-score for 83 is $z = \dfrac{y - \bar{y}}{s} = \dfrac{83 - 94.441}{5.335} = -2.14$. A score of 83 would be considered a suspect outlier.

The *z*-score for 81 is $z = \dfrac{y - \bar{y}}{s} = \dfrac{81 - 94.441}{5.335} = -2.52$. A score of 81 would be considered a suspect outlier.

The *z*-score for 78 is $z = \dfrac{y - \bar{y}}{s} = \dfrac{78 - 94.441}{5.335} = -3.08$. A score of 78 would be considered an outlier. Any score less than 78 would also be considered an outlier.

c. Some of the observations that are considered outliers using the box plot are not considered outliers using the *z*-score method. Two different methods are employed that are not exactly the same.

2.63 The *z*-score for 1.80% is $z = \dfrac{y - \mu}{\sigma} = \dfrac{1.80 - 2.00}{0.08} = -2.50$. A reading of 1.80% zinc phosphide is 2.5 standard deviations below the mean value. This would be a suspect outlier. There is some evidence to indicate that there is too little zinc phosphide in the day's production.

2.64 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: TRANERR

Results for INTRINSICS = No

```
Variable  PROJECTIONS  N      Mean  StDev
TRANERR   No           1  0.000003      *
          Yes          5     25.16   6.80
```

Results for INTRINSICS = Yes

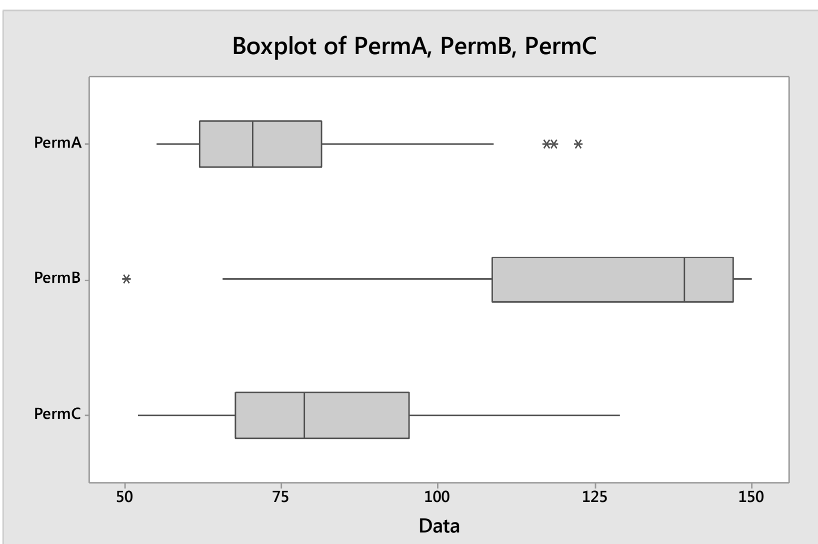
```
Variable  PROJECTIONS  N   Mean  StDev
TRANERR   No           5  1.620  0.792
```

a. For trials with perturbed intrinsics but no perturbed projections, $\bar{y} = 1.62$ and $s = 0.792$.

b. For trials with perturbed projections but no perturbed intrinsics, $\bar{y} = 25.16$ and $s = 6.80$.

c. For trials with perturbed intrinsics but no perturbed projections, an error of 4.5 would have a z-score of $z = \frac{y - \bar{y}}{s} = \frac{4.5 - 1.62}{0.792} = 3.64$. For trials with perturbed projections but no perturbed intrinsics, an error of 4.5 would have a z-score of $z = \frac{y - \bar{y}}{s} = \frac{4.5 - 25.16}{6.80} = -3.04$. Neither camera perturbation would be very likely because both z-scores are greater than 3 in magnitude. Because the z-score for trials with perturbed projections but no perturbed intrinsics is smaller in magnitude, the most likely camera perturbation would be perturbed projections but no perturbed intrinsics.

2.65 Using MINITAB, boxplots for the three groups are:



a. There are 3 observations beyond the inner fences for Group A that are suspect outliers. They have values 117.3, 118.5, and 122.4.

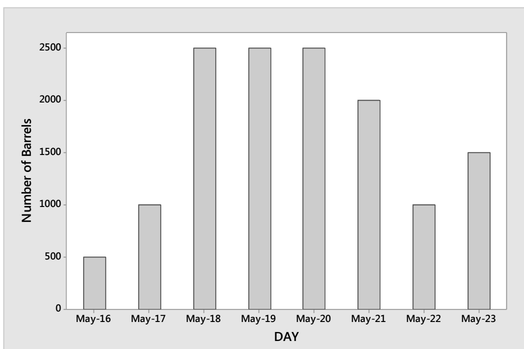
b. There is 1 observation beyond the inner fences for Group B that is a suspect outlier. The value is 50.4.

c. There are no observations beyond the inner fences for Group C. There are no suspect outliers for Group C.

2.66 In this graph, both the width and the height of the bars increase as the percentage of teeth increase. The reader may equate the area of the bars with the percentage of teeth in each category. Another problem with this graph is that the vertical axis is stretched. Rather than starting at 0, the vertical axis starts at 5. By starting at 5, it appears that the difference between the percentages for Heavy and Slight are much greater than they really are.

2.67 a. By using the cumulative number of barrels collected per day, it looks like BP was collecting more barrels of oil on each successive day, when they were collecting only about an average of 1500 barrels each day.

b. Using MINITAB, the bar chart of the data is:



From the graph, we see that the amount of oil collected per day increased from May 16 to May 18, then remained constant for three days, then decreased for two days before increasing again on May 23.

2.68 No. Even though the mean, 95.52, is higher than the interarrival time of 80, the standard deviation is almost as large as the mean. Since we know we cannot have negative interarrival times, we know the distribution of the interarrival times is skewed to the right. In addition, the median interarrival time is much smaller than the mean, again indicating that the distribution of the interarrival times is skewed to the right. When the data are highly skewed, the median is a much better measure of the "typical" observation. Because the median, 70.88, is much smaller than the interarrival time of 80, there is some evidence that this was not an inside job.

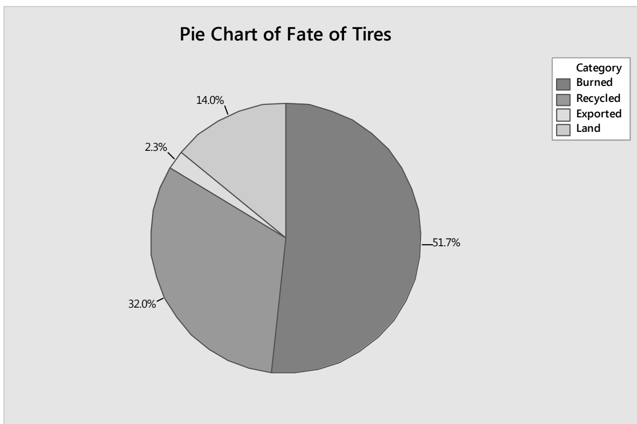
2.69 a. The variable measured for each scrapped tire is the fate of the tire.

b. There are 4 classes or categories: burned for fuel, recycled into new products, exported, or land disposed.

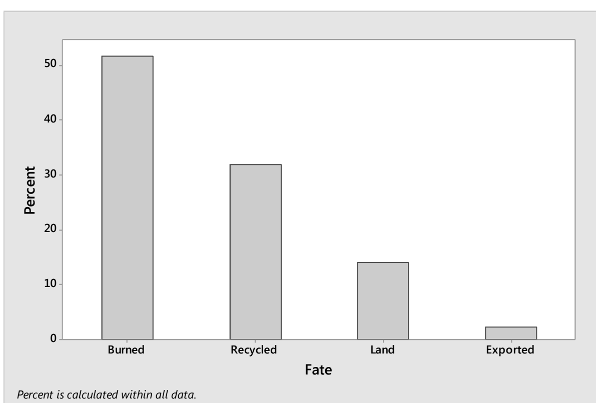
c. The class relative frequencies are computed by dividing the class frequencies by the total number of tires. The class relative frequencies are:

Fate of Tires	**Frequency (millions)**	**Relative Frequency**
Burned for Fuel	155	155/300 = 0.517
Recycled into new products	96	96/300 = 0.320
Exported	7	7/300 = 0.023
Land disposed	42	42/300 = 0.140
Totals	300	1.000

d. Using MINITAB, the pie chart is:



e. Using MINITAB, the Pareto chart is:

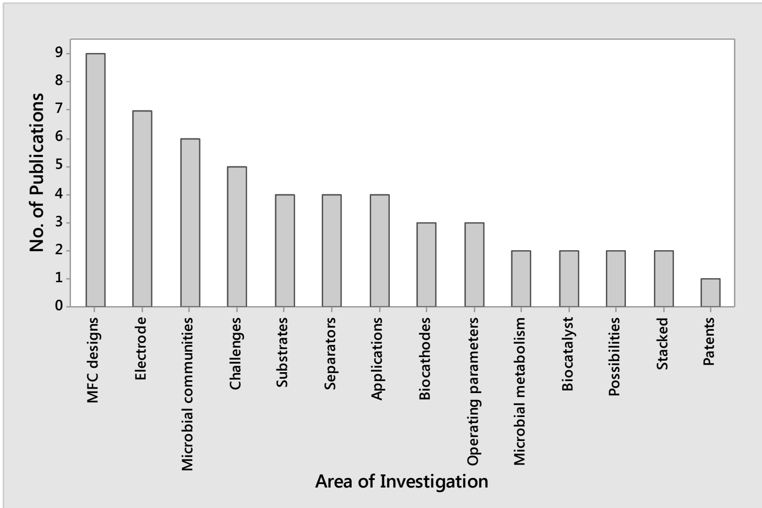


Over half of all scrapped tires are burned for fuel. About a third of scrapped tires are recycled. Only a very small percentage of scrapped tires are exported.

2.70 a. The qualitative variable measured for each research article is area of investigation.

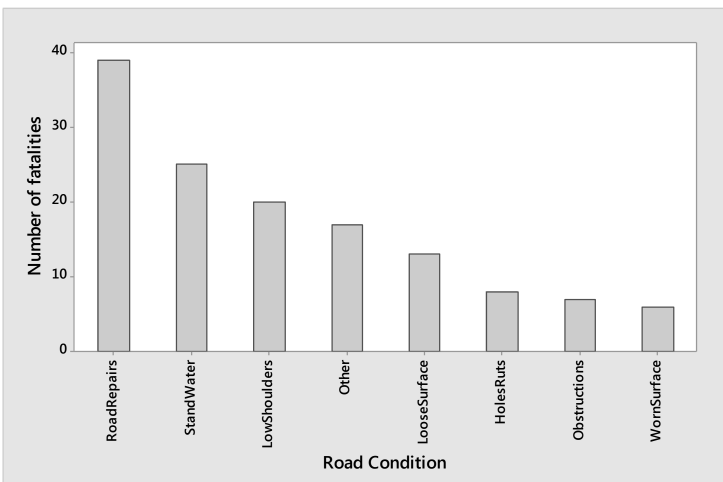
b. The graph is a bar chart.

c. Using MINITAB, the Pareto diagram is:



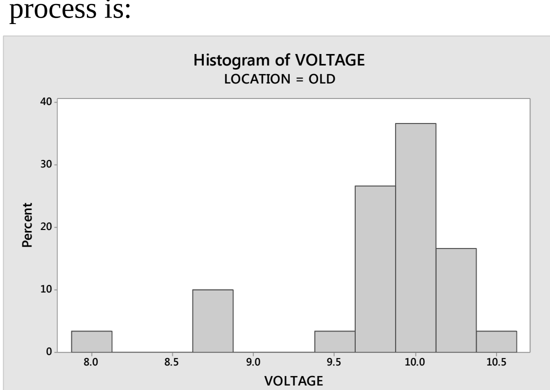
The investigation with the largest proportion of MFC research articles is MFC designs.

2.71 Using MINITAB, the Pareto diagram is:



Most of the fatalities are due to road repairs, standing water, and low shoulders. Very few fatalities are due to worn road surface, obstructions without warning, and holes and ruts.

2.72 a. Using MINITAB, the relative frequency histogram for the voltage readings of the old process is:



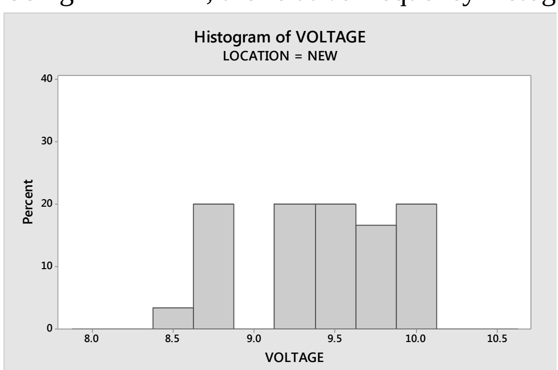
b. Using MINITAB, the stem-and-leaf display is:

Stem-and-Leaf Display: VOLTAGE

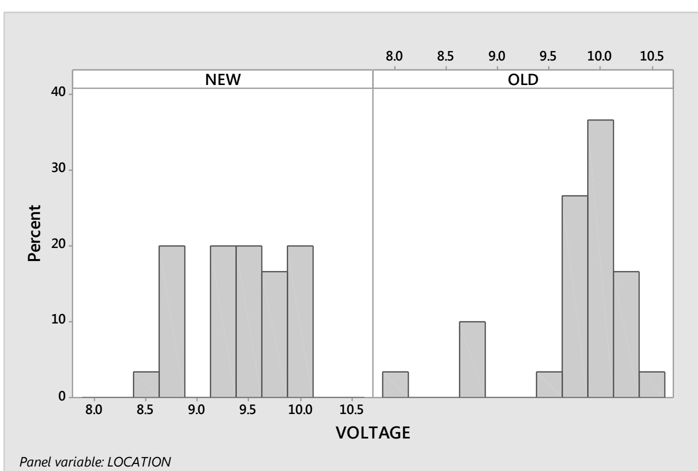
```
Stem-and-leaf of VOLTAGE  Loc = OLD    N  = 30
Leaf Unit = 0.10


 1    8   0
 1    8
 1    8
 3    8   77
 4    8   8
 4    9
 4    9
 5    9   5
 7    9   77
(10)  9   8888889999
 13   10  000000111
 4    10  222
 1    10  5
```

Both graphs give about the same information.

c. Using MINITAB, the relative frequency histogram is:



d. Using MINITAB, the relative frequency histograms for both locations are:



From the graphs, it appears that the new process is not as good as or better than the old process.

e. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: VOLTAGE

```
                                                                       N for
Variable   LOCATION   N     Mean   StDev  Median                   Mode   Mode
VOLTAGE    NEW       30  9.4223  0.4789  9.4550                    8.82      2
           OLD       30  9.8037  0.5409  9.9750  8.72, 9.8, 9.84, 9.87      2

The data contain at least five mode values. Only the smallest four are shown
```

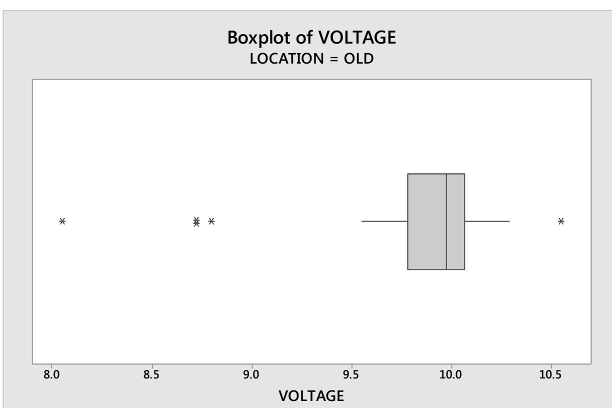
The average voltage for the old process is 9.8037 while the average voltage for the new process is 9.4223. The median for the old process is 9.9750 while the median for the new process is 9.4550. There are several modes for the old process while the mode for the new process is 8.82. It appears that the distribution of the old process is somewhat skewed to the left while the distribution of the new process appears to be somewhat symmetric. Since the distribution of the old process is skewed, the median would be a better measure of central tendency.

f. The z-score for the old process: $z = \dfrac{y - \bar{y}}{s_y} = \dfrac{10.5 - 9.8037}{0.5409} = 1.29$

g. The z-score for the new process: $z = \dfrac{y - \bar{y}}{s_y} = \dfrac{10.5 - 9.4223}{0.4789} = 2.25$

h. A voltage reading of 10.5 is more likely to occur at the old location because the z-score is closer to 0.

i. Using MINITAB, the box plot is:



All the observations marked with stars indicate possible outliers.

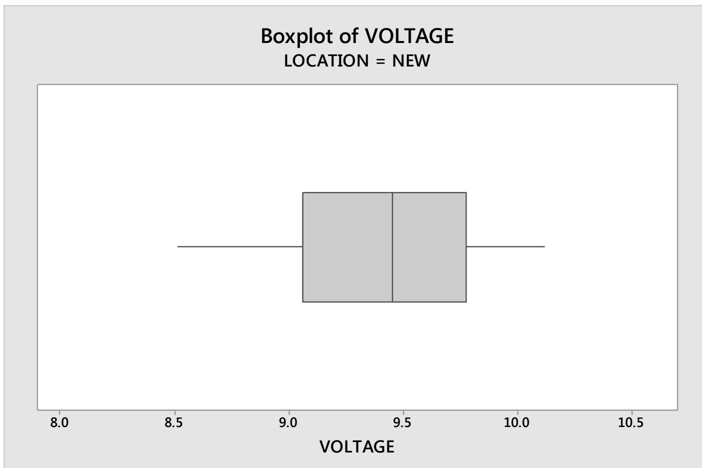
j. The smallest observation is 8.05. Its z-score is: $z = \dfrac{y - \bar{y}}{s_y} = \dfrac{8.05 - 9.8037}{0.5409} = -3.24$.

The next smallest observation is 8.72. Its z-score is: $z = \dfrac{y - \bar{y}}{s_y} = \dfrac{8.72 - 9.8037}{0.5409} = -2.00$.

The largest observation is 10.55. Its z-score is: $z = \dfrac{y - \bar{y}}{s_y} = \dfrac{10.55 - 9.8037}{0.5409} = 1.38$.

Using the z-score method there would be only one outlier, 8.05, because its z-score is greater than 3 in magnitude. There would also be one suspect outlier, 10.55, because its z-score is between 2.00 and 3.00 in magnitude.

k. Using MINITAB, the box plot is:



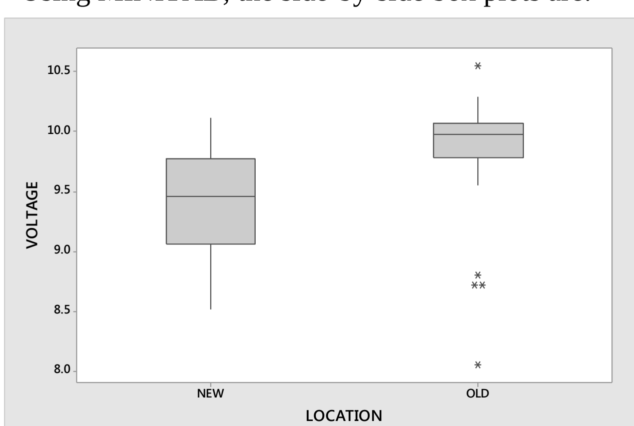
There do not appear to be any outliers as there are no stars on the plot.

l. The smallest observation is 8.51. Its z-score is: $z = \dfrac{y - \bar{y}}{s_y} = \dfrac{8.51 - 9.4223}{0.4789} = -1.90$.

The largest observation is 10.12. Its z-score is: $z = \dfrac{y - \bar{y}}{s_y} = \dfrac{10.12 - 9.4223}{0.4789} = 1.46$.

Using the z-score method there would be no outliers as no observations have z-scores greater than 2 in magnitude.

m. Using MINITAB, the side-by-side box plots are:



The mean and median of the old process are greater than the mean and median of the new process. The new process does not appear to have any outliers, while the old process appears to have many suspect outliers.

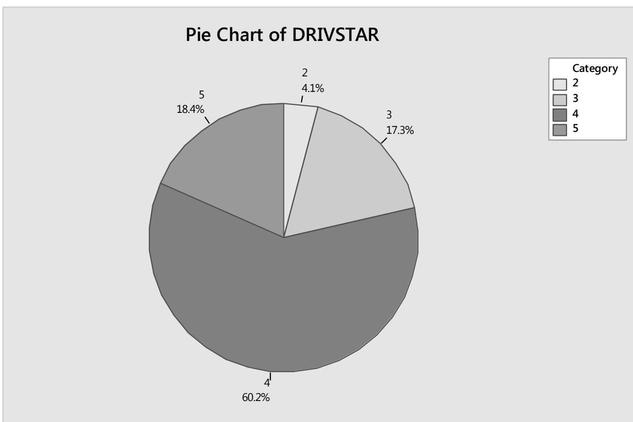
2.73 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: ROUGH

```
Variable   N   Mean  StDev
ROUGH     20  1.881  0.524
```

We know that approximately 95% of all observations will be within 2 standard deviations of the mean. This interval is $\bar{y} \pm 2s \Rightarrow 1.881 \pm 2(0.524) \Rightarrow 1.881 \pm 1.048 \Rightarrow (0.833, 2.929)$

2.74 Using MINITAB, the pie chart is:



Approximately 60% of all crash tests resulted in a score of 4 stars. About an equal number of cars had crash results of 3 or 5 stars. Very few cars had crash results of 2 stars.

2.75 A driver-head-injury rating of 408 has a *z*-score of $z = \frac{y - \bar{y}}{s_y} = \frac{408 - 603.7}{185.4} = -1.06$. A head-injury rating of 408 is a little over one standard deviation below the mean. This is not an unusual rating.

2.76 a. Using MINITAB, the descriptive statistics are:

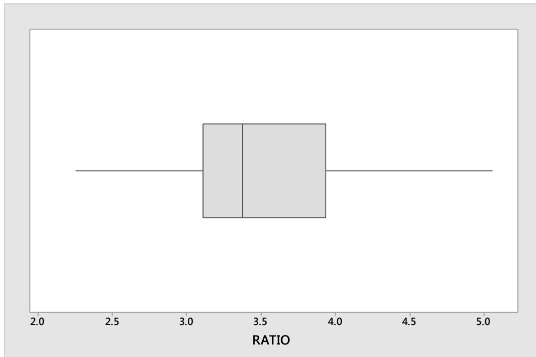
Descriptive Statistics: RATIO

```
                                                                      N for
Variable N    Mean  StDev     Q1 Median      Q3  RANGE   IQR         Mode   Mode
RATIO    26  3.507  0.635  3.113  3.375   3.938  2.810  0.825  2.73, 4.09      2
```

The average A/Be ratio is 3.507. The median A/Be ratio is 3.375. Half of all A/Be ratios are less than 3.375 and half are greater than 3.375. The mode A/Be ratio is 4.09. This observation appears 2 times in the data set.

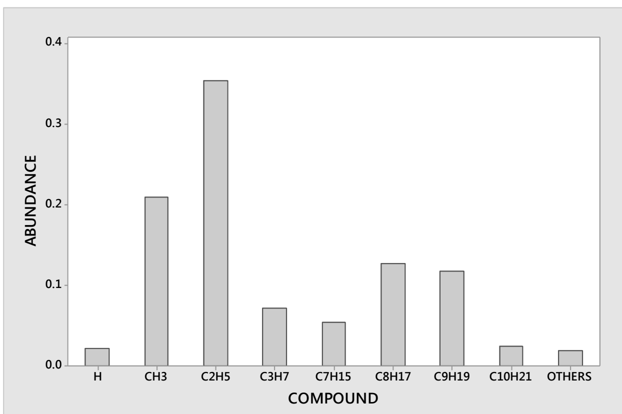
b. The range is 2.810. The difference between the largest and smallest observations is 2.810. The standard deviation is 0.635. We would expect most of the observations to be within 2 standard deviations, $2(0.635)=1.27$,of the mean. The interquartile range is 0.825. Fifty percent of the observations will fall within the interquartile range.

c. Using MINITAB, the box plot is:



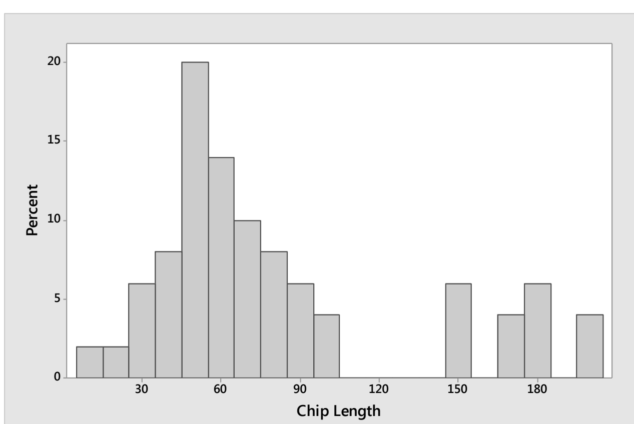
Since there are no stars, there do not appear to be any outliers.

2.77 Using MINITAB, the bar graph is:



The red dye component with the highest abundance is C_2H_5 with 35.4%. The next highest red dye component is CH_3 with 21.0%. The three components with the least abundance are $C_{10}H_{21}$ (2.5%), H (2.1%), and Others (1.9%).

2.78 a. Using MINITAB, the relative frequency histogram is:

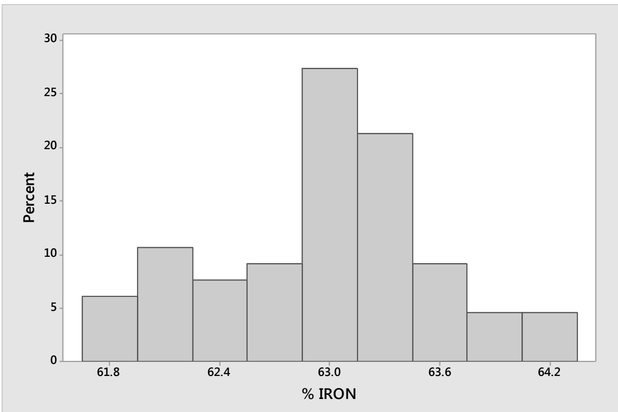


b. No. Only about 4% of all drill chips are 190mm or longer.

2.79 a. The population is all possible bulk specimens of Chilean lumpy iron ore in a 35,325-long-ton shipload of ore.

b. Answers may vary. One possible objective is to estimate the percentage of iron ore in the shipment.

c. Using MINITAB, the relative frequency histogram is:



d. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: PCTIRON

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
PCTIRON	66	62.963	0.609	61.680	62.573	63.010	63.362	64.340

$\bar{y} = 62.963$ and $s = 0.609$

e. $\bar{y} \pm 2s \Rightarrow 62.963 \pm 2(0.609) \Rightarrow 62.963 \pm 1.218 \Rightarrow (61.745, 64.181)$

64 of the 66 observations or 96.97% of the observations fall in this interval. This does not agree with the Empirical Rule. The Empirical Rule states that approximately 95% of the observations will fall within 2 standard deviations of the mean.

f. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: PCTIRON

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
PCTIRON	66	62.963	0.609	61.680	62.573	63.010	63.362	64.340

The 25th percentile is 62.573, the 50th percentile is 63.010, and the 75th percentile is 63.362. To find the 90th percentile, we calculate $i = p(n+1)/100 = 90(66+1)/100 = 60.3$. The 90th percentile is $y_{(i)} = y_{(60)} = 63.71$.

25% of the observations are less than or equal to 62.573. 50% of the observations are less than or equal to 63.010. 75% of the observations are less than or equal to 63.362. 90% of the observations are less than or equal to 63.71.

2.80 a. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: AntSpecies

Variable	N	Mean	Median	Mode	N for Mode
AntSpecies	11	12.82	5.00	4, 5	3

The average number of species of ants per site is 12.82. The median number of species per site is 5.00. Half of the sites have less than or equal to 5 species of ants. There are 2 modes: 4 and 5. The most frequent number of species of ants found per site is 4 and 5. Each of these occurred 3 times.

b. Since this distribution has a couple of very large values and is skewed to the right, the median would be the best measure of central tendency.

c. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: PlantCov

Variable	Region	N	Mean	Median	Mode	N for Mode
PlantCov	Dry Steppe	5	40.40	40.00	40	2
	Gobi Desert	6	28.00	26.00	30	2

Dry Steppe sites: The mean total plant cover percentage is 40.40. The median total plant cover percentage is 40.00. The mode total plant cover percentage is 40.

d. **Gobi Desert**: The mean total plant cover percentage is 28.00. The median total plant cover percentage is 26.00. The mode total plant cover percentage is 30.

e. Yes. The 3 measures of central tendency for the Dry Steppe sites are all larger than the corresponding measures for the Gobi Desert sites.

2.81 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: SCRAMS

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
SCRAMS	56	4.036	3.027	0.000	2.000	3.000	5.750	13.000

To find the 95th percentile, we calculate $i = p(n+1)/100 = 95(56+1)/100 = 54.15$. The 95th percentile is the 54th observation, $y_{(i)} = y_{(54)} = 9$. Thus, 95% of all observations are less than or equal to 9. A value of 11 would not be very likely.

A score of 11 would be $z = \frac{y - \bar{y}}{s} = \frac{11 - 4.036}{3.027} = 2.30$ standard deviations above the mean. A score greater than 2 standard deviations from the mean is not very likely.

2.82 Using MINITAB, the descriptive statistics are:

Descriptive Statistics: HOURS

Variable	N	Mean	StDev	Variance	Minimum	Median	Maximum	Range
HOURS	50	117.82	15.01	225.33	88.00	117.50	150.00	62.00

a. From the output, $\bar{y} = 117.82$, $m = 117.50$, and the 5 modes are 97, 112, 124, 128, and 131.

b. From the output, $Range = 62$, $s^2 = 225.33$, and $s = 15.01$.

c. $\bar{y} \pm s \Rightarrow 117.82 \pm 15.01 \Rightarrow (102.81, 132.83)$ There are 31 observations in this interval or $31/50 = 0.62$.

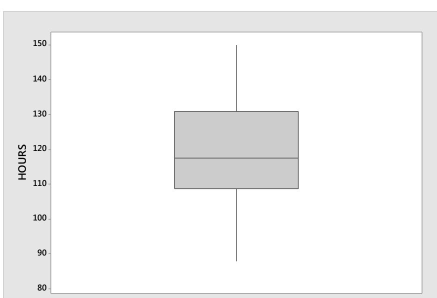
$\bar{y} \pm 2s \Rightarrow 117.82 \pm 2(15.01) \Rightarrow 117.82 \pm 30.02 \Rightarrow (87.80, 147.84)$ There are 49 observations in this interval or $49/50 = 0.98$.

$\bar{y} \pm 3s \Rightarrow 117.82 \pm 3(15.01) \Rightarrow 117.82 \pm 45.03 \Rightarrow (72.79, 162.85)$ All 50 observations are in this interval or $50/50 = 1.00$.

These data do not follow the Empirical Rule real well. The Empirical Rule says that about 68% of the observations will be within 1 standard deviation of the mean. This data set only has 62%. The Empirical Rule says that about 95% of the observations will be within 2 standard deviations of the mean. This data set has more at 98%. The Empirical Rule says that almost all of the observations will be within 3 standard deviations of the mean. This data set has 100%.

There do not appear to be any outliers.

d. Using MINTAB, the boxplot of the data is:



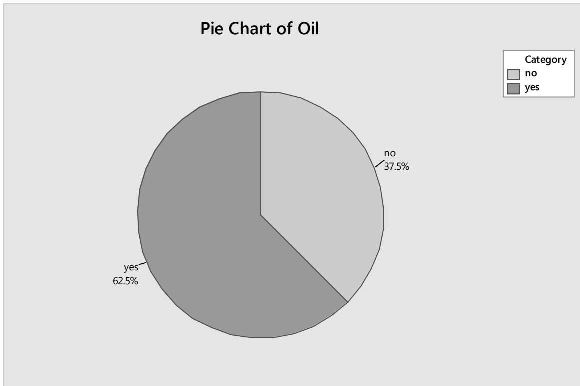
There do not appear to be any outliers.

e. To find the 70th percentile, we calculate $i = p(n+1)/100 = 70(50+1)/100 = 35.7$. The 70th percentile is the 36th observation, $y_{(i)} = y_{(36)} = 128$. Thus, 70% of all observations are less than or equal to 128.

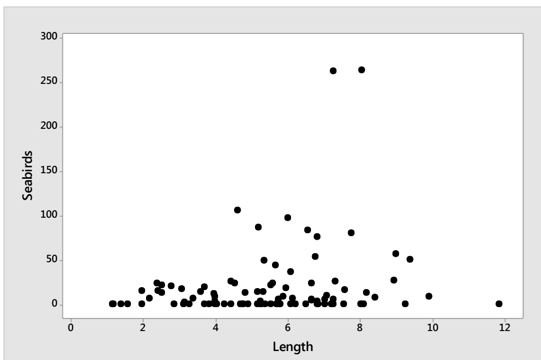
2.83 a. The number of seabirds present and the lengths of the transects are quantitative. Whether the transect was in an oiled area or not is qualitative.

b. The experimental unit is a transect.

c. Using MINITAB, the pie chart is:



d. Using MINITAB, a scatterplot is:



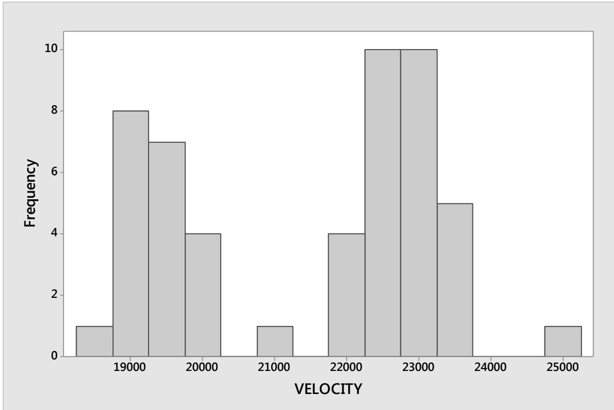
e. From the output, the means for the two groups are similar as are the medians and standard deviations. It appears that the distributions of seabird densities are similar for transects in oiled and unoiled areas.

f. The data appear to be skewed, so we will use Chebyshev's Rule. At least 75% of the observations will fall within 2 standard deviations of the mean. For unoiled transects, this interval is $\bar{y} \pm 2s \Rightarrow 3.27 \pm 2(6.70) \Rightarrow 3.27 \pm 13.40 \Rightarrow (-10.13, 16.67)$. Since a density cannot be negative, the interval should be $(0, 16.67)$.

g. The data appear to be skewed, so we will use Chebyshev's Rule. At least 75% of the observations will fall within 2 standard deviations of the mean. For oiled transects, this interval is $\bar{y} \pm 2s \Rightarrow 3.495 \pm 2(5.968) \Rightarrow 3.495 \pm 11.936 \Rightarrow (-8.441, 15.431)$. Since a density cannot be negative, the interval should be $(0, 15.431)$.

h. It appears that unoiled transects is more likely to have a seabird density of 16 because 16 falls in the interval in part f, but not in part g.

2.84 a. Using MINITAB, a histogram of the velocities of galaxy cluster A1775 is:



b. Yes, there is evidence to support a double cluster theory. There are two different mounds of data in the graph.

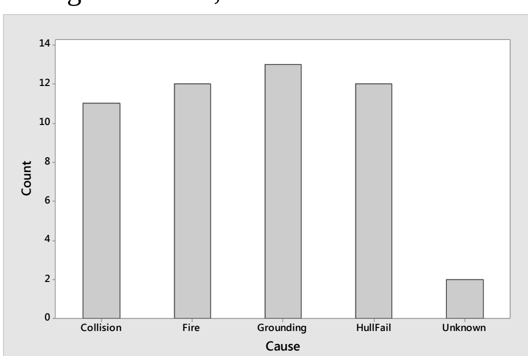
c. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: Sorted VELOCITY

Variable	Cluster	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
VELOCITY	A1775A	21	19462	532	18499	19084	19408	19774	20785
	A1775B	30	22838	561	21911	22491	22780	23146	24909

d. A galaxy velocity of 20,000 km/s is more likely to belong to cluster A1775A because all observations in this cluster have velocities less than or equal to 20,785 km/s.

2.85 a. Using MINITAB, a bar chart is:



Because no bar is way taller than the others, there does not appear to be one cause that is more likely than the others.

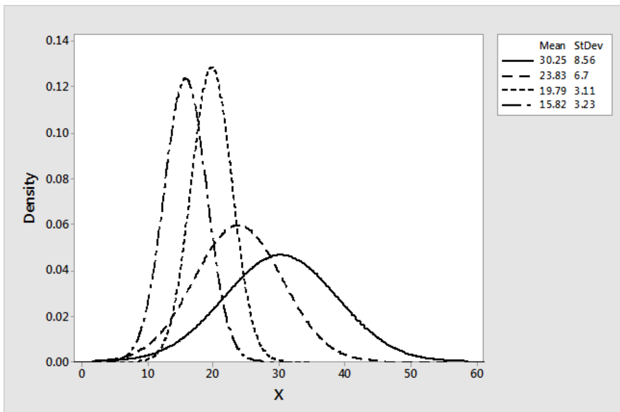
b. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: Spillage

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Spillage	50	59.82	53.36	21.00	31.00	39.50	63.50	257.00

The average spillage amount is 59.82 thousand metric tons and the median spillage amount is 39.50 thousand metric tons. The standard deviation is 53.36 thousand metric tons.
The graph of the spillage amounts is skewed to the right. Thus, we will use Chebyshev's Rule to describe the data. We know that at least 8/9 or 88.9% of the observations will fall within 3 standard deviations of the mean. This interval is
$\bar{y} \pm 3s \Rightarrow 59.82 \pm 3(53.36) \Rightarrow 59.82 \pm 160.08 \Rightarrow (-100.26,\ 219.90)$. Because we cannot have a negative spillage amount, the interval would be $(0,\ 219.90)$. Thus, we are pretty sure that the amount of the next spillage will be less than 219.9 thousand metric tons.

2.86 a. Using the Empirical Rule means that the data are approximately normally distributed for each condition for males and females separately. Therefore, there will be 4 approximately normal distributions. Using MINITAB, the graphs would look something like the following:



b. Male/1 lift per minute: $\bar{y} \pm 2s \Rightarrow 30.25 \pm 2(8.56) \Rightarrow 30.25 \pm 17.12 \Rightarrow (13.13,\ 47.37)$

Male/4 lift per minute: $\bar{y} \pm 2s \Rightarrow 23.83 \pm 2(6.70) \Rightarrow 23.83 \pm 13.40 \Rightarrow (10.43,\ 37.23)$

Female/1 lift per minute: $\bar{y} \pm 2s \Rightarrow 19.79 \pm 2(3.11) \Rightarrow 19.79 \pm 6.22 \Rightarrow (13.57,\ 26.01)$

Female/4 lift per minute: $\bar{y} \pm 2s \Rightarrow 15.82 \pm 2(3.23) \Rightarrow 15.82 \pm 6.46 \Rightarrow (9.36,\ 22.28)$

For each of the intervals, approximately 95% of all the observations for the particular category will fall within the interval.

c. We would expect that an average male could safely lift a box weighing 25 kilograms from the floor to knuckle height at a rate of 4 lifts per minute because 25 is in the interval constructed in part b for males/4 lifts per minute. However, we would not expect that an average female could safely lift a box weighing 25 kilograms from the floor to knuckle height at a rate of 4 lifts per minute because 25 is not in the interval constructed in part b for females/4 lifts per minute.

2.87 a. The figure portrays quantitative data because diameters are measured using numbers.

b. A frequency histogram is used to display the data.

c. There are about 80 observations between 1.0025 and 1.0035 and about 63 observations between 1.0035 and 1.0045. Thus, between 1.0025 and 1.0045, we have about 143 observations. This proportion is $143/500 = 0.286$.

d. Yes. The shape of the distribution is almost mound-shaped, except for the interval from 0.9995 and 1.0005 and the interval from 0.9985 and 0.9995. The number of observations in the interval 0.9995 and 1.0005 is bigger than what would be expected and the number of observations in the interval 0.9985 and 0.9995 is smaller than what would be expected.