Minitab

ANSWER KEY: DESCRIBING DATA *GRAPHICALLY*
This answer key provides solutions to the corresponding student activity sheet.

Describing Data *Graphically*

The data for these exercises are in the Minitab file ***DescribingDataGraphically_Activity.mtw***.

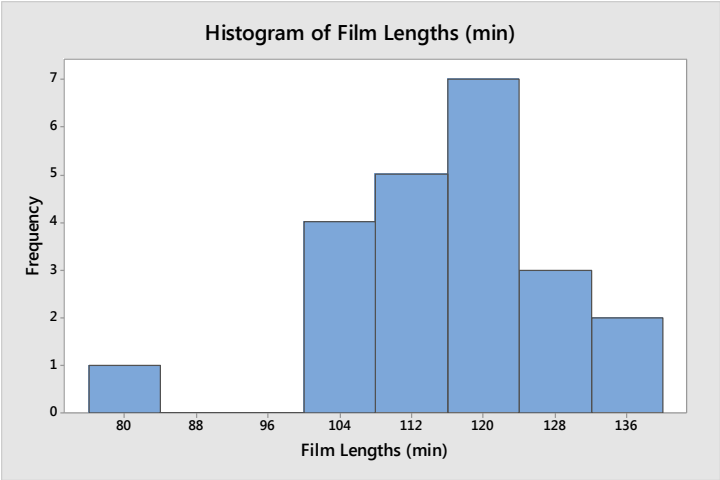
Exercise 1

(a) Construct a histogram of this data in Minitab.

Solution: The histogram will have bins of width 8 with the first bin starting at 76 and the last bin ending at 140. The midpoints of the bars will be displayed in this first histogram. If x is the film length, then the bin with midpoint 80 ranges from 76 to 84 and contains x's such that $76 \leq x < 84$.

You should notice the following items about this graph:

1. Correct histogram will have bins with midpoints
2. Correct labeling of axes, including time units

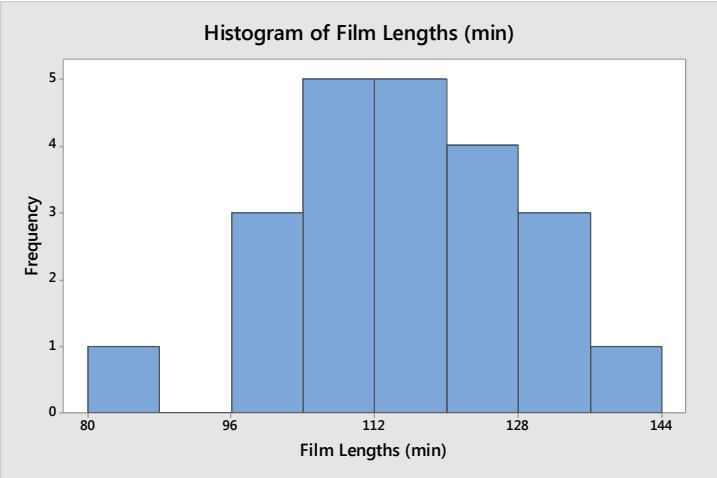


(b) Change the histogram bins to **cutpoints** (boundary values), instead of **midpoints**.

Solution: The histogram will have bins of width 8 with the first bin starting at 80 and the last bin ending at 144. The cutpoints of the bars will be displayed in this histogram. If x is the film length, then the bin with cutpoint 80 ranges from 80 to 88 and contains x's such that $80 \leq x < 88$.

You should notice the following items about this graph:

1. Correct histogram will have bins with cutpoints
2. Correct labeling of axes, including time units



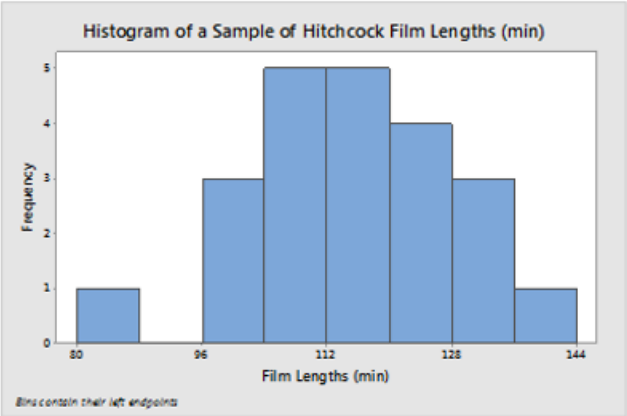
(c) Add the following enhancements to the histogram from part **(b)**.

1. Add a **footnote** to your histogram.
2. Include a more descriptive **title** with the histogram.

Solution: The histogram will have bins of width 8 with the first bin starting at 80 and the last bin ending at 144. The cutpoints of the bars will be displayed in this histogram. If x is the film length, then the bin with cutpoint 80 ranges from 80 to 88 and contains x's such that $80 \leq x < 88$. The correct histogram is displayed below on the left (the histogram on the right pertains to part **(d)** below).

You should notice the following items about this graph:

1. Correct histogram will have bins with cutpoints
2. Correct labeling of axes, including time units
3. More descriptive title than "Histogram of Film Lengths (min)"
4. Footnote explaining the binning in the histogram



(d) Which film appears to be an outlier with respect to film length?

Solution: Although we haven't formally defined outliers, it appears that the film *Rope* has an unusually small film length with respect to the other film lengths.

(e) Construct a stem-and-leaf plot of the "Film Lengths (min)" data in Minitab. Let Minitab choose the increment value.

Solution: Minitab uses an increment of 5 minutes.

Stem-and-Leaf Display: Film Lengths (min)

```
Stem-and-leaf of Film Lengths (min)  N  = 22
Leaf Unit = 1.0

 1    8   1
 1    8
 1    9
 1    9
 4    10  133
 8    10  5888
 10   11  13
(4)   11  6679
 8    12  000
 5    12  68
 3    13  02
 1    13  6
```

(f) Use your plot in part **(e)**. Ignore the first column for now [1, 1, 1, 1, 4, 8, 10, (4), 8, 5, 3, 1], and interpret row 6. What are the lengths of these Hitchcock films?

Solution: 105, 108, 108, and 108 minutes, respectively

(g) What is the longest film length from this sample of Hitchcock films?

Solution: 136 minutes

(h) Are there any Hitchcock films in this sample that have lengths between 85 and 100 minutes?

Solution: **No**

(i) Now let's use the first column, or the "count" column, of the stem-and-leaf plot for Hitchcock film lengths. How many of the sample Hitchcock films have lengths less than 110 minutes?

Solution: **8 films**

(j) What is the mode or modes of the sample of Hitchcock film lengths?

Solution: The data is bimodal: **108 minutes and 120 minutes**, each with a frequency of 3

(k) This is a personal preference question. Which graph do you prefer for gathering information about the length of Hitchcock films—the histogram or stem-and-leaf plot? Briefly state why.

Solution: Some possible plusses and minuses of each:

Histogram plusses:

+ A histogram is easy to construct once you determine the bin widths.
+ A histogram is easy to read and understand if the binning method is explained clearly.
+ In general, many non-statistical people have seen a histogram and get the basic idea of what it's representing.
+ The histogram displays the basic shape, center, and spread of the data, in addition to outliers.
+ The histogram provides a quick visual "summary" of the data.

Histogram minuses:

– The histogram does not show the values of the data contained within each of the bins.
– It can be time-consuming to determine the binning structure for a histogram.
– Different binning structures give different views of the same data set, which can create confusion when reading the graph.

Stem-and-leaf plot plusses:

+ In a stem-and-leaf plot, the actual data values are part of the graphic. You can compute the mean, median, mode, standard deviation, and range of a data set by using the data in the plot.
+ A stem-and-leaf plot is easy to construct once you determine the bin increments.
+ A stem-and-leaf plot is easy to read and understand once it is explained to the person reading the graph.
+ The stem-and-leaf plot displays the basic shape, center, and spread of the data, in addition to outliers.

Stem-and-leaf plot minuses:

- In general, many non-statistical people have never seen a stem-and-leaf plot. Detailed instructions need to be provided before the person reading the graphic can understand and construct one.
- It can be hard to determine the best increment for the plot's rows.
- Different incrementing can give different views of the same data set, which can create confusion when reading the graph.

Exercise 2

Below are stem-and-leaf plots of n = 40 Statistics Exam 1 scores. One plot uses an increment of 10 and the other uses an increment of 5, where the **increment** indicates the difference in value between stems.

What is revealed about the data by the second stem-and-leaf plot (with an increment of 5) that is not visibly apparent in the first stem-and-leaf plot (with an increment of 10)?

Stem increments of 10:

Stem-and-Leaf Display: Exam I Scores

```
Stem-and-leaf of Exam I Scores  N  = 40
Leaf Unit = 1.0
 9    6  034667899
 17   7  00122244
(19)  8  0011111223445557899
 4    9  0358
```

Stem increments of 5:

Stem-and-Leaf Display: Exam I Scores

```
Stem-and-leaf of Exam I Scores  N  = 40
Leaf Unit = 1.0
 3    6  034
 9    6  667899
 17   7  00122244
 17   7
(12)  8  001111122344
 11   8  5557899
 4    9  03
 2    9  58
```

What does the second stem-and-leaf plot (with an increment of 5) reveal about the data that is not visibly apparent in the first stem-and-leaf plot (with an increment of 10)?

Solution: There are NO test scores between and including 75 to 79.

Exercise 3

(a) How many teams had salaries of at least 70 million dollars?

Solution: 11 teams

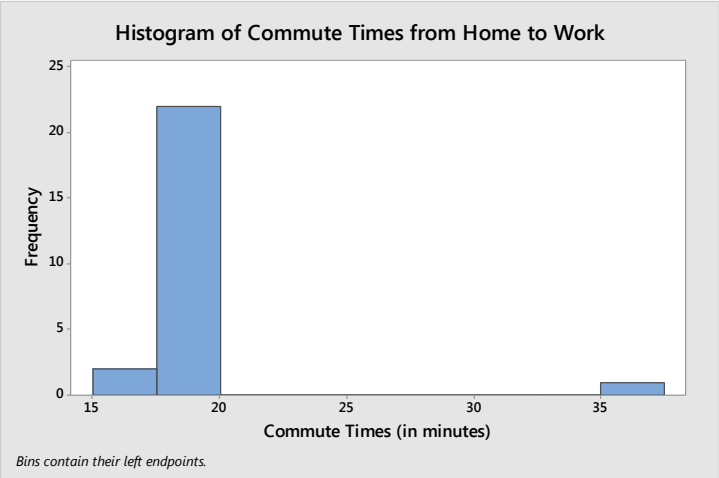
(b) What is the median team salary for the 30 NBA teams?

Solution: Since there are 30 teams, then the median is the average of the 15^{th} and 16^{th} data points. Thus, the median team salary is **67.5 million dollars**.

Exercise 4

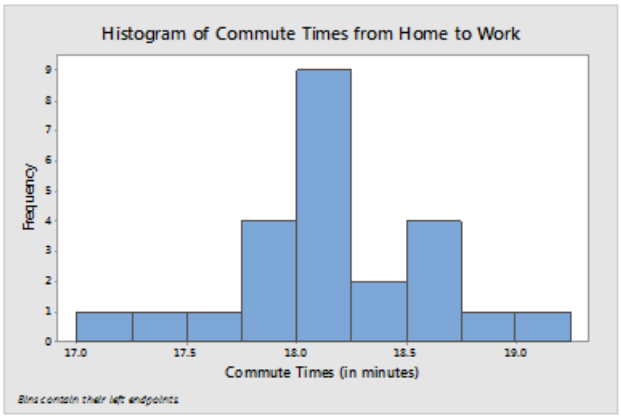
(a) Construct a histogram in Minitab of your professor's commute times.

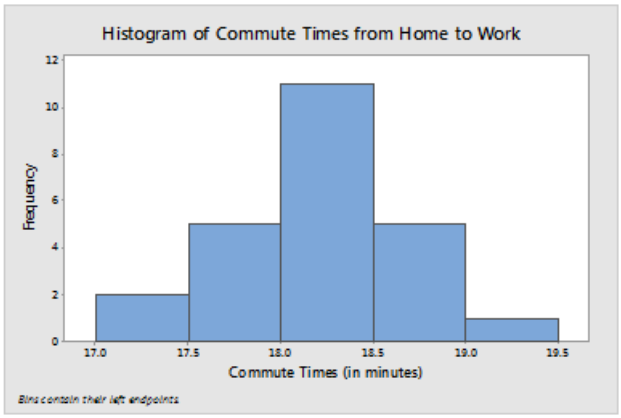
Solution:



(b) The 21st commute time, 37.4 minutes, reflects a day when your professor left home without his laptop computer and had to turn around to retrieve it. Remove this outlier from the data set and reconstruct the histogram.

Solution: Depending on the bin width chosen, most students will have one of the graphs below.





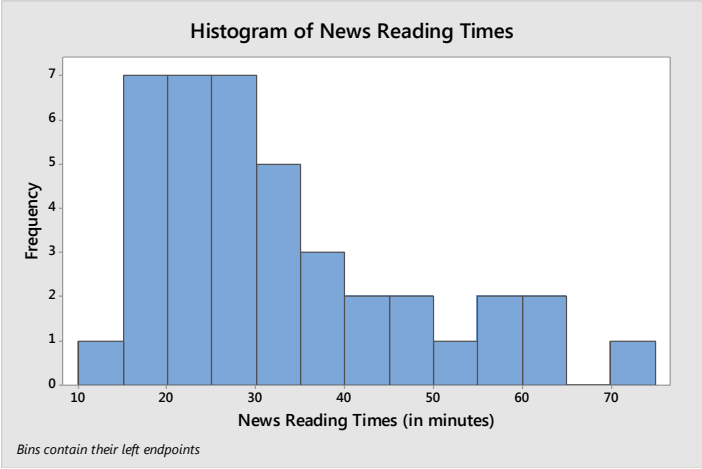
(c) How many days did your professor's commute time fall between 18 minutes (inclusive) and 18.5 minutes (exclusive); i.e. 18 ≤ commute time < 18.5?

Solution: 11 days

Exercise 5

(a) Construct a histogram of this data using cutpoints.

Solution: The graph should look fairly identical to one of the histograms below. Make sure the title is clear, time units are provided, and there is a note about the bin structure.



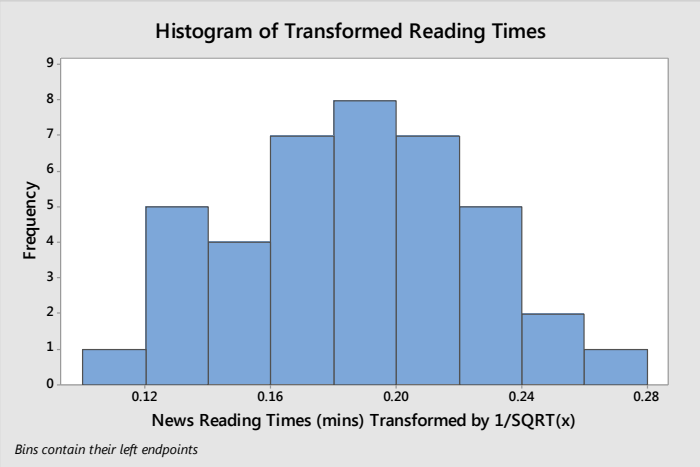
(b) What type of skewness, if any, does this data display?

Solution: Positively skewed or right skewed.

(c) Calculate $1/\sqrt{x}$ for each data value and put the new transformed data values in a new column in Minitab.

(d) Construct a histogram of the transformed data using cutpoints.

Solution: The graph should look fairly identical to the histogram below if a bin width of 0.02 is used. Make sure the title indicates that the histogram was constructed with transformed data. Time units should still be provided, as well as a note about bin structure.



(e) What is the effect of the transformation on the data?

Solution: The transformation has affected the positive skewness of the data. The transformed data appears to be symmetric and even bell-shaped.