Section 5.3: Organizing and Graphing Quantitative Data

Objective: Organize quantitative data into frequency tables

An easy way to compile quantitative data would be to make a frequency or relative frequency table as we did with qualitative data.

Commute time to school

The data in the table below are the commute time to school (in minutes) for a group of students attending a particular mathematics course. We would like to create a frequency table.

10	15	17	20	25	20	3	30	17	15	5	15
60	8	25	15	25	22	38	10	14	30	30	18

Since there are no natural categories for this data we must create what are called classes. Classes divide the number line into smaller pieces.

First let's find the minimum commute time (3 minutes). Next, find the maximum commute time (60 minutes). We can start our first class at 3 minutes or back up a few minutes. Let's start the first class at 1 minute and use a class width of 10 minutes. The first class is "1 – 10". The 1 is considered the lower limit of the class and the 10 is considered the upper limit of the class.

$$1 - 10$$

Lower class limit   Upper class limit
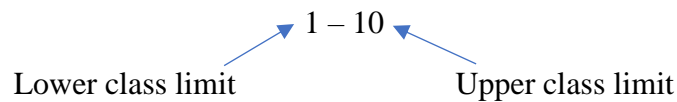
Next, we will determine the lower limits of the other classes. Add the class width of 10 to our first lower limit. This will give us 11. If we continue to add 10 the remaining lower limits will be 21, 31, 41, and 51. The upper limits are determined by filling in the numbers that approach the next class' lower limit but do not equal it. For example in the second row we chose 20 because it is the closest whole number less than 21.

Commute Time
1 – 10
11 – 20
21 – 30
31 – 40
41 – 50
51 – 60

The difference is 10

The class width of 10 is evident by looking at the differences in consecutive lower class limits. This is also the case with consecutive upper class limits. They are all 10. Once we made all the classes and made sure that the minimum and maximum can be placed in the table, we see that there are 6 classes. Now we must determine how many commute times fall into each of the classes. There are several strategies to tallying up the counts. You can mark them off as you go, or possibly put unique symbols or marks next to the ones that are in the same classes. This will help you avoid classifying a value twice or forgetting a data value altogether. Once all the symbols are there you can count them.

10 *	15✋	17✋	20✋	25❖	20✋	3 *	30❖	17✋	15✋	5 *	12✋
60■	8 *	25❖	15✋	25❖	22❖	38⌘	10 *	14✋	30❖	30❖	18✋

	Commute Time	**Frequency**
*	1 – 10	5
✋	11 – 20	10
❖	21 – 30	7
⌘	31 – 40	1
⊠	41 – 50	0
■	51 – 60	1

Vocabulary:

Classes   =   range of data values used to group the data.
Lower class limit =   the smallest value that goes in a class.
Upper class limit =   the largest value that goes in a class.
Class width   =   the difference between two consecutive lower class limits.


Objective: Create a histogram for quantitative data

We can take our frequency table and create a graph of the information. This graph is called a histogram. A histogram is like a bar graph. The classes are along the horizontal axis and the frequencies are demonstrated with the vertical axis. The bars need to touch in a histogram because we want to imply that the classes are adjacent and represent a continuum on the number line.

Steps to sketch a histogram:

1) Draw the horizontal axis with the lower class limits equally spaced along it.
2) Draw the vertical axis with the frequencies equally spaced along it.
3) Create the bars (rectangles)

The histogram for the "Commute time to school" is shown below.
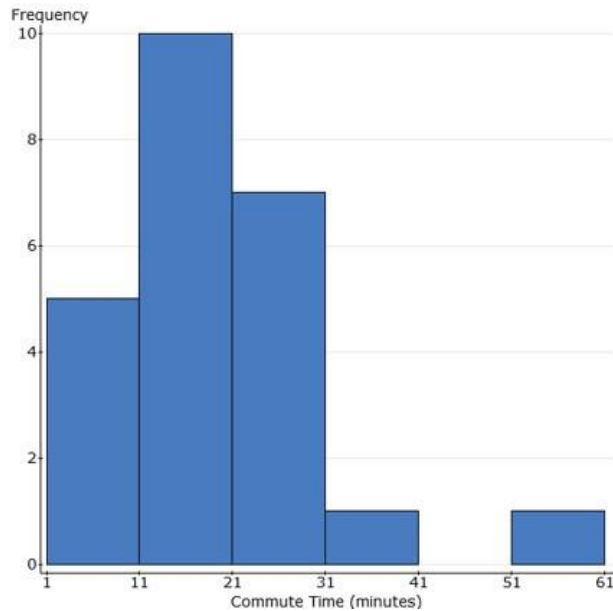
You can see the height of the bars are exactly as our frequency table (5, 10, 7, 1, 0, and 1). The lower class limits are along the horizontal axis. The first bar gives us the commute times that were in the class 1 – 10. A histogram allows us to visualize the data and see how the data is spread out or possibly similar to one another. Notice there is not a bar for 41 – 50. This is because the frequency was zero for that class. Frequencies of zero are the only reasons there are gaps in histograms.

Objective: Create Relative Frequencies

We can also add a column to the frequency table to represent relative frequency. This is similar to what was done for frequency tables for categorical data.

$$\text{Relative frequency for a class} = \frac{\text{frequency in that class}}{\text{sum of all frequencies}}$$

| Commute Time	Frequency	Relative Frequency
1 – 10	5	$\frac{5}{24} = 0.21$
11 – 20	10	$\frac{10}{24} = 0.42$
21 – 30	7	$\frac{7}{24} = 0.29$
31 – 40	1	$\frac{1}{24} = 0.04$
41 – 50	0	0
51 – 60	1	$\frac{1}{24} = 0.04$
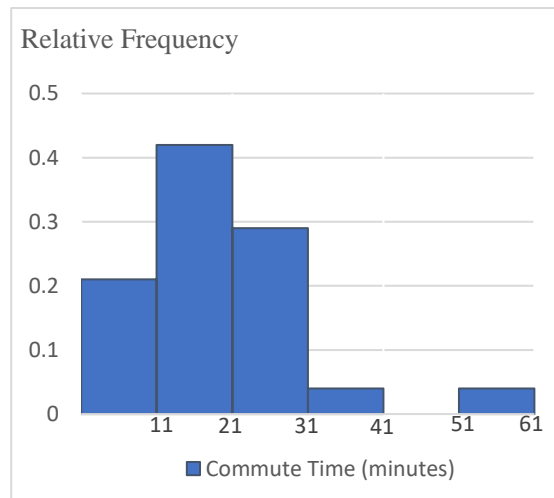
The relative frequencies, like the histogram, allow us to see that the majority of the class had a commute time between 11 and 30 minutes.

Objective: Discuss the shape of histograms

Histograms are valuable tools to display data. There are features that we tend to describe with words. For example, it is helpful to mention how many peaks (humps) are present in the graph. Does the histogram have a single, central peak or several separated peaks? A histogram with one main peak is dubbed **uni**modal; histograms with two peaks are **bi**modal; histograms with three or more peaks are called **multi**modal.

This is an example of a **unimodal** histogram. This histogram displays 90 different rainfalls at a national park. They measured the pH level of each rainfall.
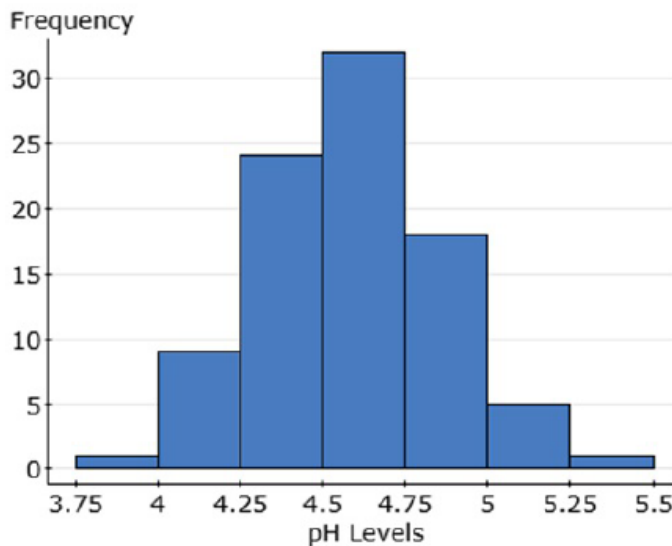
This is an example of a bimodal histogram. These are the heights of 75 singers in a chorus.
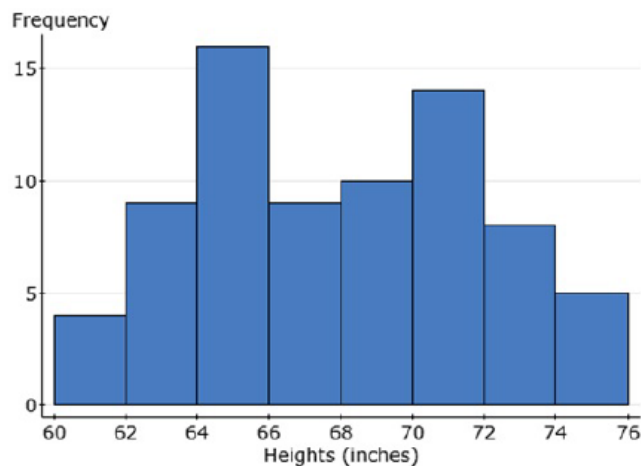
Another way we describe histograms is to discuss their symmetry (or lack there of). If you can fold the histogram along a vertical line through the middle and have the edges match pretty closely, the histogram is considered symmetric.
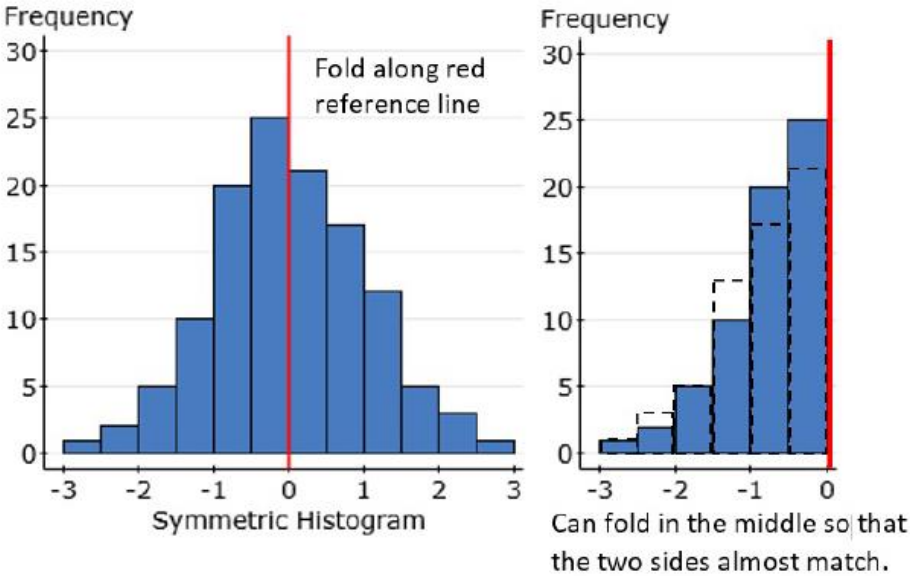
The (usually) thinner ends of a distribution are called the tails. If one tail stretches out farther than the other, the histogram is said to be skewed to the side of the longer tail.

Below the histogram on the left is said to be skewed left, while the histogram on the right is said to be skewed right.
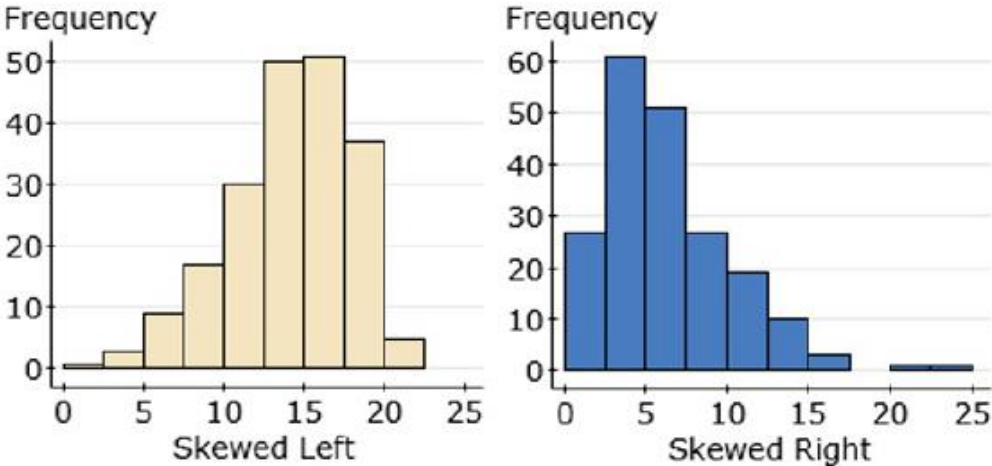
Objective: Create dotplots for quantitative data

Definition. *A* **dotplot** *is a visual representation of quantitative data and provides a graphical display of the data distribution*.

Steps to construct a dot plot:

1) Begin by drawing a number line that reflects the range of values.
2) Plot each data point by placing a dot over the appropriate value. For any repeated value stack the points.

The dotplot for the "Commute time to school" is shown below.

Commute Time (minutes)
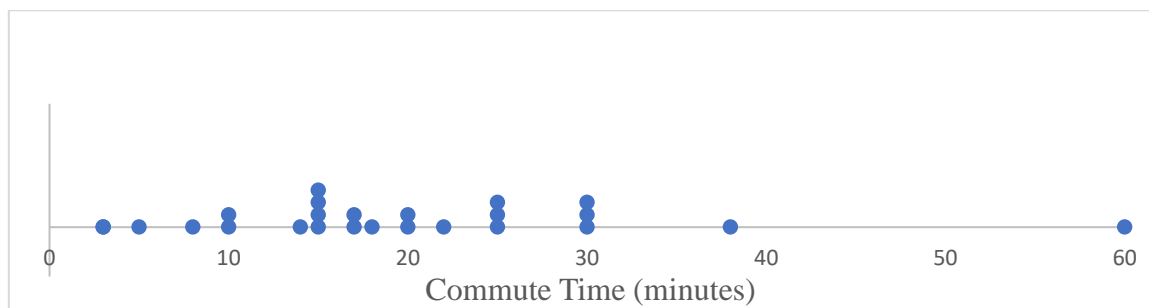
Using the dotplot we can answer many questions about our data.

What observation occurred the most?	We have four stacked dots above 15, so we conclude that 15 minutes is the observation that occurred the most.
How many students were surveyed?	Each point represents a reported commute time. By counting all the points on the graph we get the total number of students surveyed. We have 24 dots so the answer is 24 students.
What proportion of students commute to school in 20 minutes or less?	Count the points on the graph that are stacked at the value 20 or below it. 15 students commute in 20 minutes or less. Divide 15 by 24 (total students) to get the proportion. The answer is 15/24=0.625

Objective: Create stem-and-leaf plots for quantitative data

Definition. *A* **stem-and-leaf plot** *is another way to arrange quantitative data. The plot separates each value into two parts: the stem (such as everything of the leftmost digits) and the leaf (such as the rightmost digit).*

Steps to construct a stem-and-leaf plot:

1) Draw a vertical bar

2) Identify the stems and list them down on the left side of the vertical bar
3) List to the right of the vertical bar the leaves corresponding to their stems

Consider the Commute to school data

10	15	17	20	25	20	3	30	17	15	5	15
60	8	25	15	25	22	38	10	14	30	30	18

For this data, the stems are the tens, namely, 0, 1, 2, 3, 4, 5, 6 and the leaves are all the ones. For example, the data value 10 is separated into its stem of 1 and the leaf of 0. The data value 15 is separated into its stem of 1 and the leaf of 5. All the values are separated in the same way and arranged as shown below.

```
0 | 358
1 | 0045555778
2 | 002555
3 | 0008
4 |
5 |                    2|5  means  25
6 | 0
```

Note: The leaves are arranged in increasing order.

5.3 Practice

1) The frequency distribution below summarizes employee years of service for a particular corporation.

Years of service	Frequency
1 – 5	4
6 – 10	15
11 – 15	20
16 – 20	9
21 – 25	5
26 – 30	2

   a) How many employees participated in the survey?
   b) What is the class width?
   c) Identify the lower class limits.
   d) Identify the upper class limits.

2) A business magazine was conducting a study into the amount of travel required for mid-level managers across the US. They surveyed a group of managers and asked them the number of days they spent traveling each year. The frequency distribution below summarizes the results.

Days Traveling	Frequency
0 – 6	15
7 – 13	21
14 – 20	27
21 – 27	9
28 – 34	2
35 – 41	1

   a) How many managers participated in the survey?
   b) What is the class width?
   c) Identify the lower class limits.
   d) Identify the upper class limits.

3) A teacher at your college selected a group of students from all of his classes and asked them the number of credits each took that semester.

| 9	6	10	3	9	7	10	9	12	9	6	9
12	9	7	10	9	6	9	7	7	14	9	8

a) Create a frequency and a relative frequency table to display the data. Let your first lower limit be 3 and the class width be 3 credits.
b) Create a frequency histogram.
c) Create a relative frequency histogram.
d) How does the frequency histogram and relative frequency histogram compare?

4) Twenty-four students were asked the number of hours they sleep each night. The results of the survey are listed below.

7	9	5	8	8	11	12	7	10	7	8	9
8	6	4	3	9	10	7	6	12	5	6	11

a) Create a frequency and a relative frequency table to display the data. Let your first lower limit be 3 and the class width be 2 hours.
b) Create a frequency histogram.
c) Create a relative frequency histogram.
d) How does the frequency histogram and relative frequency histogram compare?

5) In a survey, 26 voters were asked their ages. The results are shown below.

| 43	56	28	63	67	66	52	48	37	56	40	60	62
66	45	21	35	49	32	56	61	53	69	31	48	59

a) Create a frequency table for the ages. Let your first lower limit be 19 and the class width be 10.
b) Create a frequency histogram.

6) The following histograms describe the number of minutes per day that a group of 50 elementary and 50 middle school students spent on a computer.

Daily Time Spent on Computer (Elementary)

Daily Time Spent on Computer (Middle)
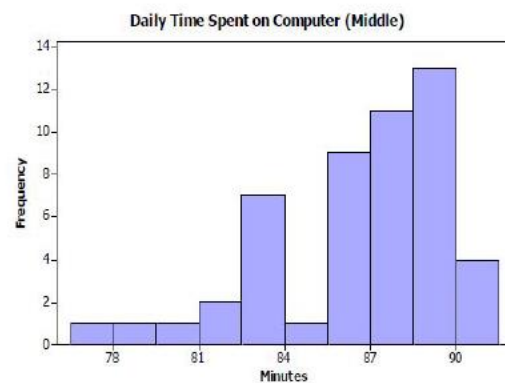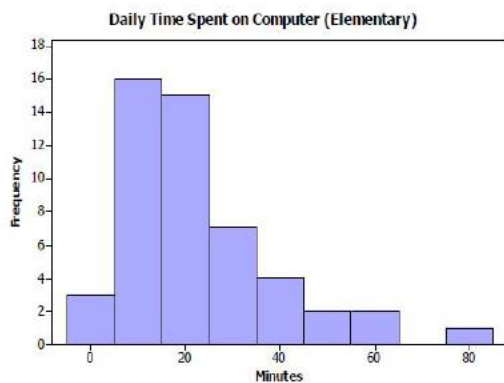
a) Which histogram appears to be skewed to the right?

b) Which histogram appears to be skewed to the left?

7) The following histograms represent the pulse rate (beats per minute) for a group of 40 males and 40 females.
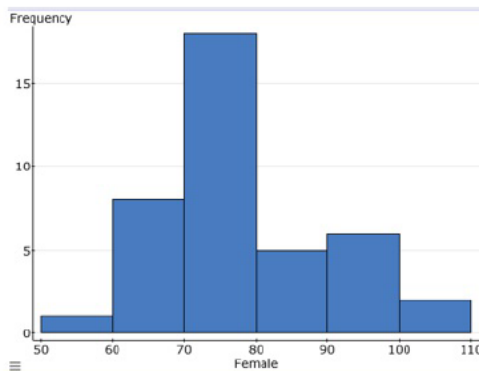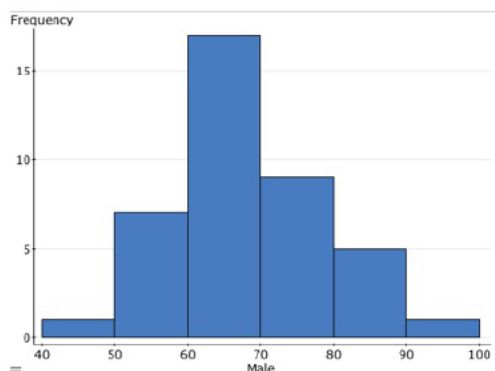
a) Describe the shape of the two histograms.
b) Compare the histogram of pulse rate of males with females. Is there a notable difference between pulse rates of females and males?

8) The stem and leaf plot below shows the number of laps run by each participant in a marathon. Refer to the stem and leaf plot to answer the following questions.

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

a) Construct a dotplot.
b) How many participants ran less than 48 laps?
c) How many participants ran at least 50 laps?
d) What observation occured the most?

9) The stem and leaf plot below shows the ages of a group of patients who had strokes caused by stress. Refer to the stem and leaf plot to answer the following questions.

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

a) Construct a dotplot.
b) How many patients were at most 36 years old?
c) How many patients were 72 and older?
d) What observation occurred the most?

5.3 Answers

1.
   a) 55 employees
   b) Class width: 5
   c) Lower class limits: 1, 6, 11, 16, 21, 26
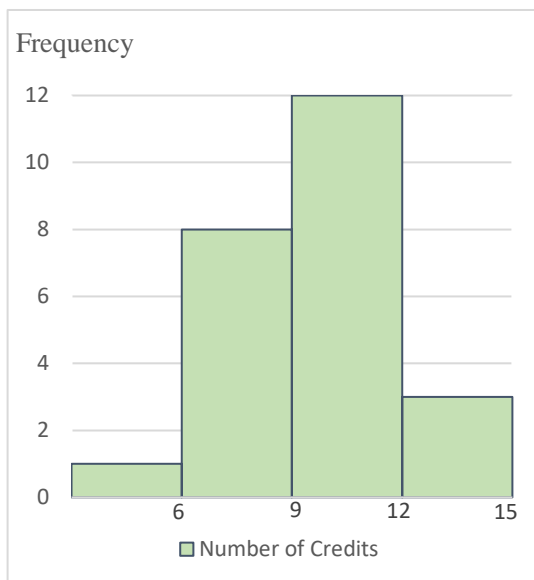   d) Upper class limits: 5, 10, 15, 20, 25, 30

2.
   a) 75 managers
   b) Class width: 7
   c) Lower class limits: 0, 7, 14, 21, 28, 35
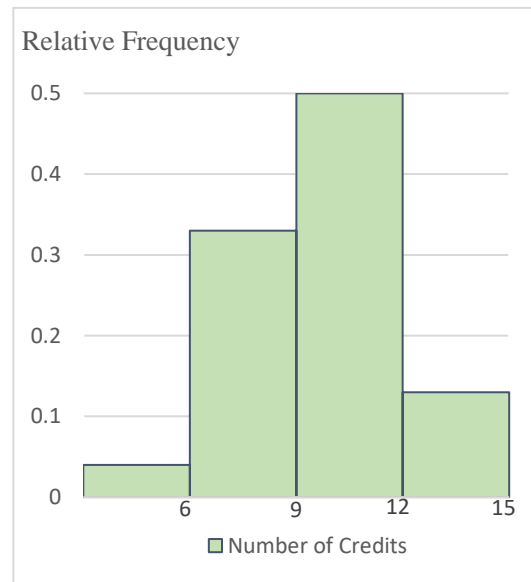   d) Upper class limits: 6, 13, 20, 27, 34, 41

3.
   a)

Number of credits	Frequency	Relative Frequency
3 – 5	1	0.04
6 – 8	8	0.33
9 – 11	12	0.50
12 – 14	3	0.13

b)

Frequency — Number of Credits

c)



Relative Frequency — Number of Credits

d) The frequency histogram and relative frequency histogram have the same shape because frequencies and relative frequencies are proportional. The only difference is on the units of the $y-$axis.
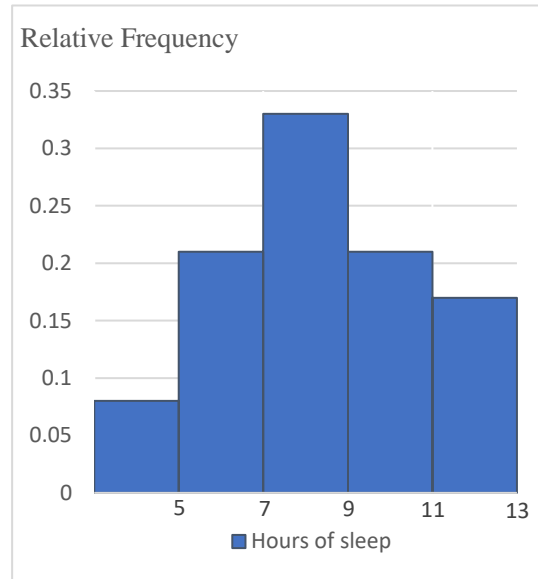
4.

a)

Hours of Sleep	Frequency	Relative Frequency
3 – 4	2	0.08
5 – 6	5	0.21
7 – 8	8	0.33
9 – 10	5	0.21
11 – 12	4	0.17

b)

Frequency

Hours of sleep

c)

Relative Frequency

Hours of sleep
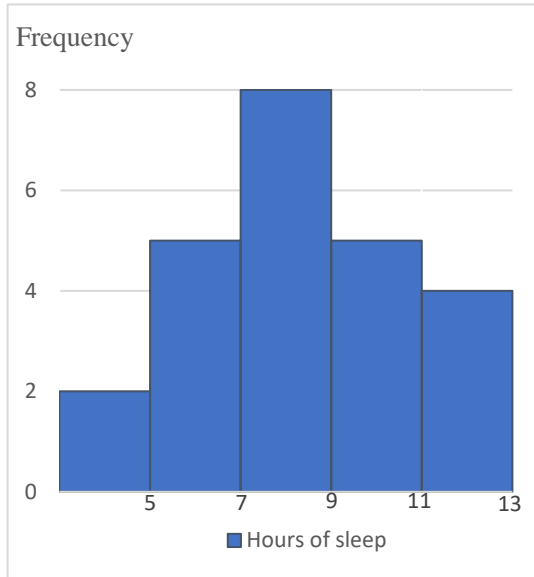
5. a)

Age	Frequency
19 – 28	2
29 – 38	4
39 – 48	5
49 – 58	6
59 – 68	8
69 – 78	1

b)

Frequency

Ages
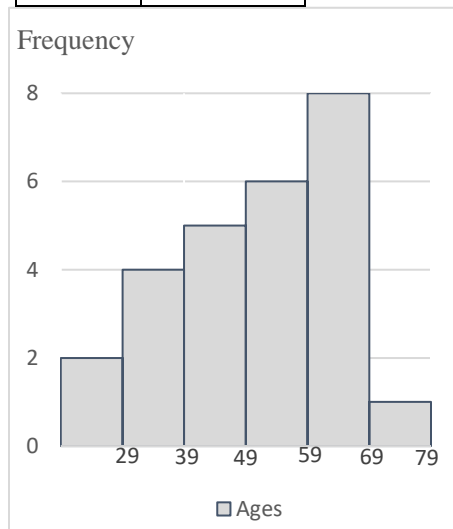
6.
   a) Elementary school students' histogram appears to be skewed to the right.
   b) Middle school students' histogram appears to be skewed to the left.
7.
   a) The histogram of pulse rate of males appears to be unimodal and approximately symmetric. The histogram of pulse rate of females appears to be bimodal and skewed to the right.
   b) The pulse rates of males appear to be generally lower than the pulse rates of females.

8.

Number of laps
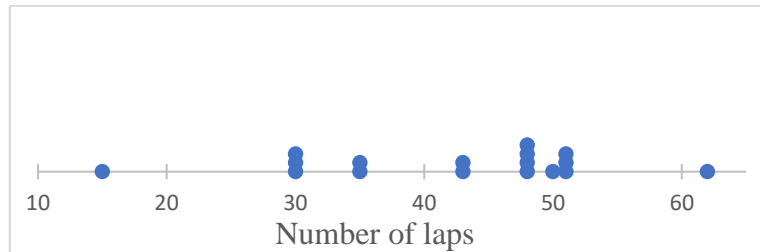
   a)
   b) 8 participants ran less than 48 laps
   c) 5 participants ran at least 50 laps
   d) 48 occured the most
9.

Age
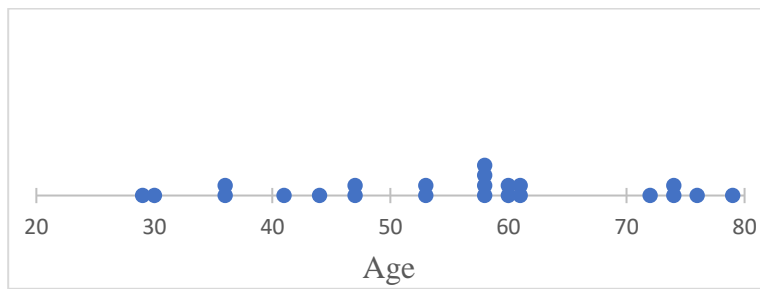
   a)
   b) 4 patients were at most 36 years old
   c) 5 patients were 72 and older
   d) 58 occured the most