CHAPTERS 8 AND 10
Probability and Statistics Unit

Ms. Romano is a health coach and nutritionist. Recently, she encouraged Matthew to eat a healthier breakfast by choosing a cereal with less sugar.

There are many different cereals, and the amount of sugar can vary widely. Matthew took a trip to the grocery store and recorded the amount of sugar in one serving of each cereal type.

Cereal	Amount of Sugar In a Single Serving (grams)
Cocoa Rounds	13
Flakes of Corn	4
Frosty Flakes	11
Grape Nuggets	7
Golden Nuggets	10
Honey Nut Squares	10
Raisin Branola	7
Healthy Living Flakes	7
Wheatleys	8
Healthy Living Crunch	6
Multi-Grain Squares	7
All Branola	5
Munch Crunch	12
Branola Flakes	5
Complete Flakes	4
Corn Crisps	3
Rice Crisps	4
Shredded Wheatleys	1
Puffs	22
Fruit Circles	11

1. Analyze the data collected. What conclusions can you draw about the amount of sugar in different cereals?

   The sugar amounts vary widely.

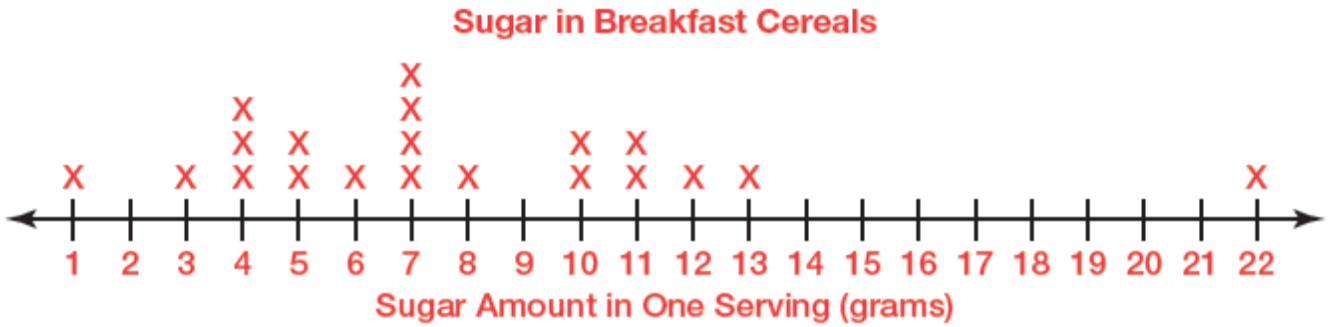
It can be difficult to analyze data listed in a table.
A better way to organize data is to create a graph.

A **dot plot** is a graph that displays discrete data distributed on a number line.

Discrete data has only a finite number of values. Discrete data can be counted or measured.

Dot plots are best used to organize and display the frequency of a small number of data points.

2. **Construct a dot plot** representing the amount of sugar in one serving of each breakfast cereal. **Label the number line** using intervals that include all the data values. **Place an "x"** above the number that represents each data value. Make sure you **name your dot plot**.



3. Analyze the dot plot. What conclusions can you draw about the amount of sugar in a single serving of breakfast cereal?

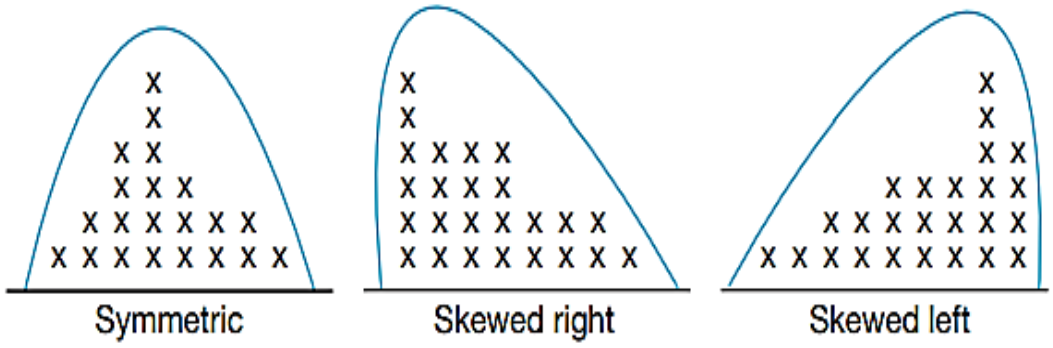
Most cereals have less than 10 grams of sugar per serving. Only one cereal has more than 13 grams of sugar per serving.

When you analyze the characteristics of a graph, ask yourself the following questions:

- What is the overall shape of the graph? Does it have any interesting patterns?
- Where is the approximate middle or center, of the graph?
- How spread out are the data values on the graph?

The overall shape of a graph is called the **data distribution**. It is the way data is spread out or clustered together.

The shape of the distribution can reveal a lot of information about the data. There are many different distributions, but the **most common** are **symmetric**, **skewed right**, and **skewed left** as shown.



4. Describe the properties of a data distribution that is:

   a) Symmetric — Most of the data values are in the middle and the remaining values are spread out in approximately the same pattern on either side of the middle.

   b) Skewed right — Most of the data values are left of center with fewer data values to the right. The data "tails off" to the right.

   c) Skewed left — Most of the data values are to the right of center with fewer data values to the left. The data "tails off" to the left.

5. Looking back at our dot plot, describe how the amount of sugar is distributed in one serving of breakfast cereal. Explain what this means in terms of the problem situation.

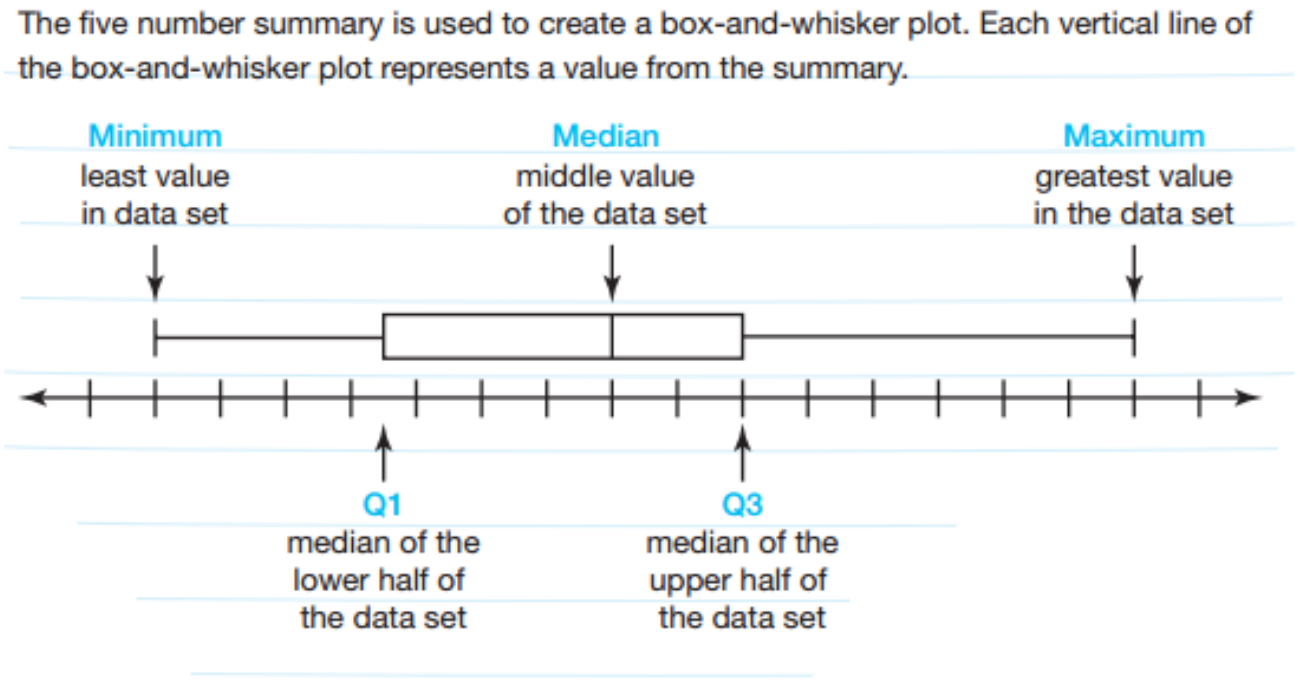
   The data is *skewed right.* This means most breakfast cereals have a low number of grams of sugar per serving while only a few cereals have a high number of grams of sugar per serving.

6. Do you think the conclusion you came to in Question 5 is true of **ALL** breakfast cereals? Why or why not?

   No. Matthew chose 20 breakfast cereals and there are more than 150 different cereals. Since he did not take into account all cereals, the conclusion is only true for the cereals he analyzed.

"Boxing It Up"

Another way to graphically represent a data distribution is a **box-and-whisker plot**. A box-and-whisker plot displays the data distribution using a **five number summary** that consists of the Minimum value, the 1st Quartile (Q1), the Median, the 3rd Quartile (Q3), and the Maximum value. **Quantitative data** is just another term for numerical data.



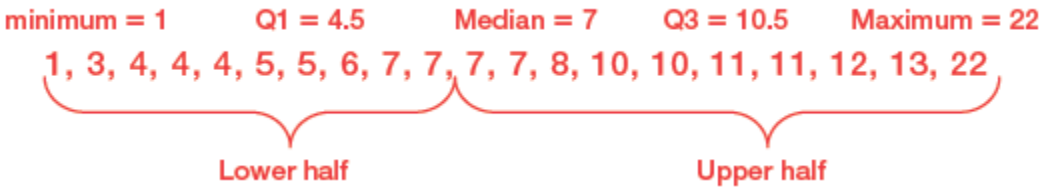
There are 4 sections of the graphical display:
Minimum to Q1, Q1 to Median, Median to Q3, and Q3 to Maximum.

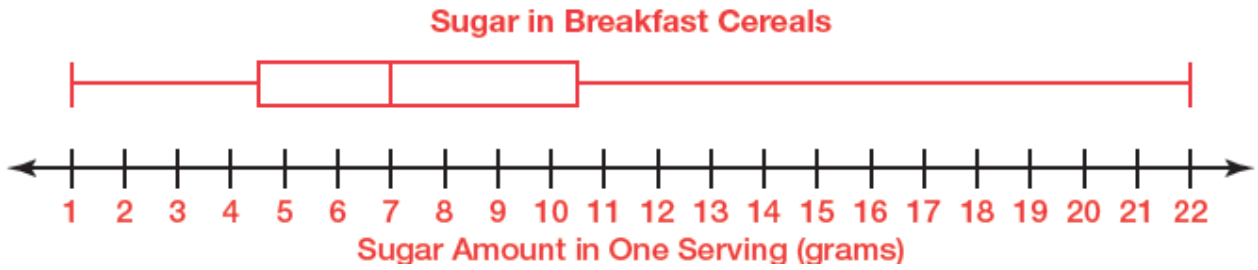
Each section of the box-and-whisker plot represents **25%** of the data set.

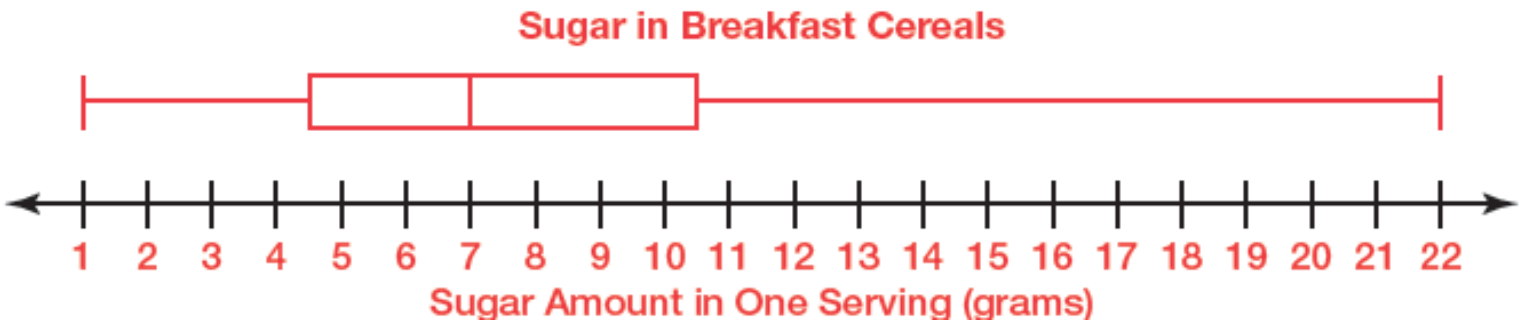
1. Determine the percentage of data values for each section of the box-and-whisker plot previously displayed.

   - Less than (<) Q1, Greater than (>) Q1 25%, 75%
   - Less than (<) Q3, Greater than (>) Q3 75%, 25%
   - Less than (<) the Median, Greater than (>) the Median 50%, 50%
   - Between Q1 and Q3 50%

2. Construct a box-and-whisker plot showing the amount of sugar in one serving of each breakfast cereal from Problem 1, How Much Sugar Is Too Much? Before you start constructing, list the data values in order.







3. Analyze the five number summary and box-and-whisker plot.

- 50% of the cereals had 7 or less grams of sugar & 50% had 7 or more grams of sugar
- 25% had less than 4.5 grams of sugar & 75% had more than 4.5 grams of sugar
- 75% had less than 10.5 grams of sugar & 25% had more than 10.5 grams of sugar
- 50% of the cereals had between 4.5 and 10.5 grams of sugar

4. Describe the data distribution shown in the box-and-whisker plot.

   The data distribution is skewed right.

"Weekend Gamers"

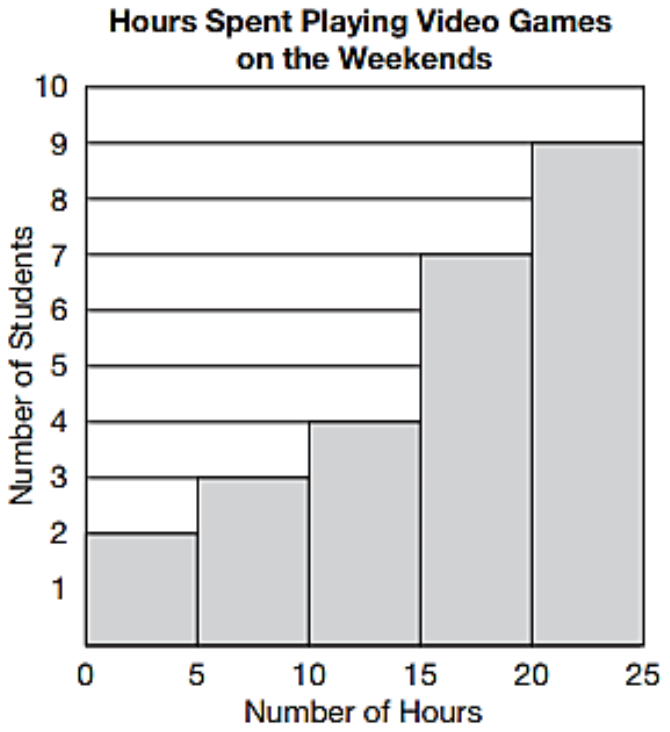
Another way to display quantitative or measurable data is to create a **histogram**. A histogram displays quantitative data using vertical bars.

The width of a bar in a histogram represents an *interval of data* rather than individual data values and is often called a **bin**. The value shown on the left side of the bin is the smallest data value in the interval.

The height of each bar indicates the frequency or the number of times the data values are included in any given bin.

Histograms are effective in displaying large amounts of continuous data. **Continuous data** is data that can take any numerical value within a range.

The histogram shown represents the data distribution for the number of hours students spend playing video games on the weekends. The data is gathered to the nearest half-hour.



1. What conclusions can you draw from the histogram about the number of hours students spend playing video games on the weekends?

   Only 2 students play less than 5 hours of video games & 9 students play at least 20 hours of video games on weekends.

2. Analyze the histogram.

   a. How many students play 5 to 9.5 hours of video games on weekends?
      3 students

   b. How many total students are included in the data?
      25 students

   c. How many students play exactly 22 hours of video games on the weekends?
      You can't tell because the data in the last bin could be any value from 20 to 24.5 hours.

   d. What percent of the students play 10 or more hours of video games on the weekends?
      20 of the 25 students play 10 or more hours of video games on the weekends. 20/ 25 = 0.8 or 80%

3. Describe the data distribution displayed by the histogram.

   The data is skewed left.