

Start Your Day the Right Way

Graphically Representing Data

8.1

LEARNING GOALS

In this lesson, you will:

- Represent and interpret data displayed on dot plots.
- Represent and interpret data displayed on histograms.
- Represent and interpret data displayed on box-and-whisker plots.

KEY TERMS

- dot plot
- discrete data
- data distribution
- symmetric distribution
- skewed right distribution
- skewed left distribution
- box-and-whisker plot
- five number summary
- histogram
- bin
- frequency
- continuous data

Many nutritional experts call breakfast the most important meal of the day, and many people start their day with a bowl of cereal. However, cereal was not always an option. In the late 1800s, most people's diets consisted mainly of meat products, including breakfasts of pork and beef. However, John Harvey Kellogg and his brother William Keith Kellogg, both of whom worked at a health spa, began creating vegetarian-based breakfast options for their guests using grains. It was actually by mistake that they created some of the first flakes of wheat cereal. This mistake was an immediate success! Just a few years later, the Kellogg Company was selling more than one million cases of cereal a year.

Some of today's cereals still contain the healthy whole grains that the Kelloggs used in their original recipe. However, there are many other cereals that contain other ingredients that are not quite as healthy. What are some healthy cereal options in the stores today? What are some cereals that might not be considered as healthy? What is the difference between these two types of cereal?

8

PROBLEM 1 How Much Sugar Is Too Much?

Ms. Romano is a health coach and nutritionist. Recently, she encouraged Matthew to eat a healthier breakfast and recommended a cereal with less sugar. There are many different cereals and it seems like the amount of sugar in each type varies widely. Matthew took a trip to the grocery store and recorded the sugar amount that each cereal has in one serving.

Cereal Name	Sugar Amount in One 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

© 2012 Carnegie Learning

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



It may be difficult to properly analyze data in a table. One way to better organize the data is to create a graph. A **dot plot** is a graph that shows how *discrete data* are distributed using a number line. **Discrete data** are data that has only a finite number of values or data that can be "counted." Dot plots are best used to organize and display the number of occurrences of a small number of data points.

Remember, if a value occurs more than once, place an "x" above the number each time the value occurs.



2. Construct a dot plot to represent the sugar amount in one serving of each breakfast cereal. Label the number line using intervals that will 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 sugar amounts in one serving of breakfast cereal from the dot plot?

8

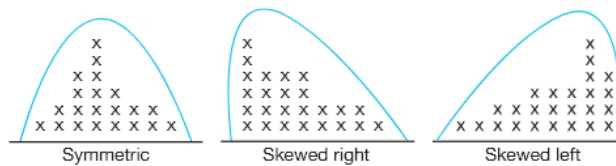
4. Jordan states that those numbers on the number line that do not contain any data values should be eliminated. Toni disagrees and says that all the numbers on the number line must be included even if there are no data values for that particular number. Who is correct? Explain your reasoning.



When you analyze a graphical display, you can look at several characteristics of the graph to draw conclusions. For example, you can ask yourself:

- 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*. The **data distribution** is the way in which the 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.



© 2012 Carnegie Learning

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

a. symmetric.

b. skewed right.

c. skewed left.

8

In a **symmetric distribution** of data, the left and right halves of the graph are nearly mirror images of each other. There is often a “peak” in the middle of the graph.

In a **skewed right distribution** of data, the peak of the data is to the left side of the graph. There are only a few data points to the right side of the graph.

In a **skewed left distribution** of data, the peak of the data is to the right side of the graph. There are only a few data points to the left side of the graph.

6. Describe the distribution of the sugar amount in one serving of breakfast cereal. Explain what this means in terms of the problem situation.



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

8

PROBLEM 2 Boxing It Up



Another graphical representation that displays the distribution of quantitative data is a *box-and-whisker plot*. A **box-and-whisker plot** displays the data distribution based on a *five number summary*. The **five number summary** consists of the minimum value, the first quartile (Q1), the median, the third quartile (Q3), and the maximum value.

Quantitative data is just another term for numerical data.



The five number summary is used to create a box-and-whisker plot. Each vertical line of the box-and-whisker plot represents a value from the summary.

Minimum least value in data set	Median middle value of the data set	Maximum greatest value in the data set
--	--	---

Q1
median of the
lower half of
the data set

Q3
median of the
upper half of
the data set

The lines connecting the minimum and Q1, and Q3 and the maximum are known as the whiskers.

There are four 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 percent of the data set.



© 2012 Carnegie Learning



1. Determine each percent of data values for the given sections of the box-and-whisker plot shown in the worked example. Explain your reasoning for each.

- a. Less than Q1
Greater than Q1

- b. Less than Q3
Greater than Q3

- c. Less than the median
Greater than the median



- d. Between Q1 and Q3



8



- Construct a box-and-whisker plot of the sugar amount 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.



- Analyze the five number summary and box-and-whisker plot. What conclusions can you draw about the sugar amount in one serving of breakfast cereal from these representations?

Interpret the data in terms of percents.



- Describe the data distribution shown in the box-and-whisker plot. Interpret the meaning of the distribution in terms of this problem situation.

© 2012 Carnegie Learning

5. Damon states that more breakfast cereals have over 10 grams of sugar per serving than have under 5 grams of sugar per serving because the whisker connecting the maximum and Q3 is longer than the whisker connecting the minimum and Q1. Is Damon correct? Explain why or why not.

8



PROBLEM 3 Weekend Gamers

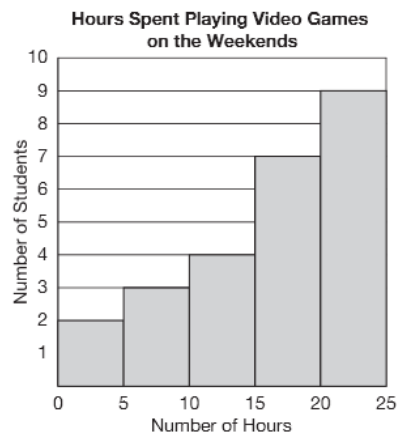


Another way to display quantitative data is to create a *histogram*. A **histogram** is a graphical way to display quantitative data using vertical bars. The width of a bar in a histogram represents an interval of data and is often referred to as a **bin**. A bin is represented by intervals of data instead of showing individual data values. The value shown on the left side of the bin is the least data value in the interval.

The height of each bar indicates the **frequency**, which is the number of data values included in any given bin.

Histograms are effective in displaying large amounts of *continuous data*. **Continuous data** is data which 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.



8




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

2. Jonae and Tyler must identify the greatest value represented in the bin beginning with 15. Their responses are shown.

 **Jonae**

The bin that begins with the interval 15 includes all data values from 15 to 20.

 **Tyler**

The bin that begins with the interval 15 includes all data values from 15 thru, but not including 20.

a. Explain why Tyler's answer is correct and why Jonae's answer is incorrect.

b. Represent the bin that contains 15 as an inequality.

3. Analyze the histogram.

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

b. How many total students are included in the data? Explain your reasoning.

- c. Marcel states that between 0 and 5 students spend 2 hours playing video games on weekends. Is Marcel's statement correct? Explain why or why not.

8

- d. How many students play 22 hours of video games on the weekends? Explain your reasoning.

- e. What percent of the students play 10 or more hours of video games on the weekends? Explain your reasoning.



4. Describe the data distribution displayed by the histogram. Interpret its meaning in terms of this problem situation.

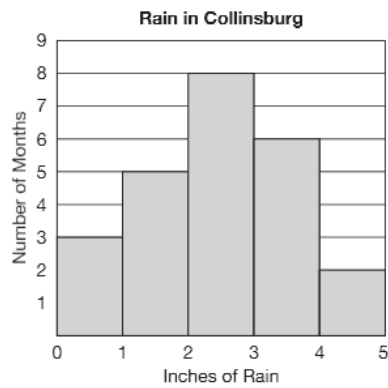
8

Talk the Talk



Analyze each data representation to answer the questions. Justify your reasoning using the characteristics of each representation.

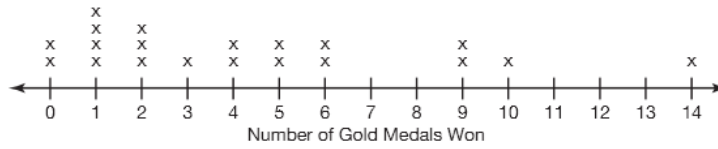
1.



- a. Describe the information represented in the histogram.
- b. How many months are represented on the histogram? Describe how you determined your answer.
- c. Identify the intervals represented by each bin.
- d. How many months had 4 or more inches of rain?
- e. Describe the data distribution and interpret its meaning in terms of this problem situation.

2. **Participants Who Won Gold Medals at the Special Olympics**

8



- Describe the information represented in the dot plot.
- How many participants are represented in the dot plot?
- How many participants won 10 or more medals?
- Describe the data distribution and interpret its meaning in terms of this problem situation.

3. **Volunteers Hours at the Local Animal Shelter**



- Describe the information represented in the box-and-whisker plot.
- How many people are represented on the box-and-whisker plot?
- What percent of the people volunteered 14 or more hours?

© 2012 Carnegie Learning

8

- d. What percent of people volunteered less than 11 hours?
- e. How many hours did the middle 50 percent of the people volunteer?
- f. Describe the data distribution and interpret its meaning in terms of this problem situation.
4. Analyze each visual display shown. Describe what information each display provides. Be sure to include advantages and limitations and any specific characteristics for each visual display.
- table
 - dot plot
 - five number summary
 - box-and-whisker plot
 - histogram



Be prepared to share your solutions and methods.