Truth, Lies, and Statistics: Data Analysis in Middle School

NCTM National Conference

Boston, Massachusetts

April 18, 2015

Mark D. Jones

"There are three types of lies -- lies, damn lies, and statistics."
— Benjamin Disraeli

Useful Website

For complete lessons that include student versions, teacher versions, resource pages, extra practice, and answers go to: http://cpm.org/teachers/stats_prob.htm

Citations

"Statistics & Probability." Statistics & Probability. Common Core State Standard Initiative, 2015. Web. 11 Apr. 2015. <http://www.corestandards.org/Math/Content/SP/>.

Common Core State Standards Writing Team. "Progressions for the Common Core." (n.d.): n. pag. 26 Dec. 2011. Web. 11 Apr. 2015. <http://www.commoncoreworks.org/cms/lib/DC00001581/Centricity/Domain/120/ccss_progression_sp_68_2011_12_26_bis.pdf>.

Common Core State Standards Data Analysis Progressions

In **Grade 6**, students build on the knowledge and experiences in data analysis developed in earlier grades. They develop a deeper understanding of variability and more precise descriptions of data distributions, using numerical measures of center and spread, and terms such as cluster, peak, gap, symmetry, skew, and outlier. They begin to use histograms and box plots to represent and analyze data distributions. As in earlier grades, students view statistical reasoning as a four-step investigative process:

- Formulate questions that can be answered with data
- Design and use a plan to collect relevant data
- Analyze the data with appropriate methods
- Interpret results and draw valid conclusions from the data that relate to the questions posed.

Such investigations involve making sense of practical problems by turning them into statistical investigations (MP1); moving from context to abstraction and back to context (MP2); repeating the process of statistical reasoning in a variety of contexts (MP8).

In **Grade 7**, students move from concentrating on analysis of data to production of data, understanding that good answers to statistical questions depend upon a good plan for collecting data relevant to the questions of interest. Because statistically sound data production is based on random sampling, a probabilistic concept, students must develop some knowledge of probability before launching into sampling. Their introduction to probability is based on seeing probabilities of chance events as long-run relative frequencies of their occurrence, and many opportunities to develop the connection between theoretical probability models and empirical probability approximations. This connection forms the basis of statistical inference. With random sampling as the key to collecting good data, students begin to differentiate between the variability in a sample and the variability inherent in a statistic computed from a sample when samples are repeatedly selected from the same population. This understanding of variability allows them to make rational decisions, say, about how different a proportion of "successes" in a sample is likely to be from the proportion of "successes" in the population or whether medians of samples from two populations provide convincing evidence that the medians of the two populations also differ.

Until **Grade 8**, almost all of students' statistical topics and investigations have dealt with univariate data, e.g., collections of counts or measurements of one characteristic. Eighth graders apply their experience with the coordinate plane and linear functions in the study of association between two variables related to a question of interest. As in the univariate case, analysis of bivariate measurement data graphed on a scatterplot proceeds by describing shape, center, and spread. But now "shape" refers to a cloud of points on a plane, "center" refers to a line drawn through the cloud that captures the essence of its shape, and "spread" refers to how far the data points stray from this central line. Students extend their understanding of "cluster" and "outlier" from univariate data to bivariate data. They summarize bivariate categorical data using two-way tables of counts and/or proportions, and examine these for patterns of association.

Common Core State Standards for Statistics

Grade 6

Develop understanding of statistical variability.

CCSS.Math.Content.6.SP.A.1

Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages*.

CCSS.Math.Content.6.SP.A.2

Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

CCSS.Math.Content.6.SP.A.3

Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

Summarize and describe distributions.

CCSS.Math.Content.6.SP.B.4

Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

CCSS.Math.Content.6.SP.B.5

Summarize numerical data sets in relation to their context, such as by:

CCSS.Math.Content.6.SP.B.5.a

Reporting the number of observations.

CCSS.Math.Content.6.SP.B.5.b

Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

CCSS.Math.Content.6.SP.B.5.c

Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

CCSS.Math.Content.6.SP.B.5.d

Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Grade 7

Use random sampling to draw inferences about a population.

CCSS.Math.Content.7.SP.A.1

Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.

CCSS.Math.Content.7.SP.A.2

Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. *For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be*.

Draw informal comparative inferences about two populations.

CCSS.Math.Content.7.SP.B.3

Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *For example, the mean height of players on the basketball team is 10 cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviation) on either team; on a dot plot, the separation between the two distributions of heights is noticeable*.

CCSS.Math.Content.7.SP.B.4

Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book*.

Grade 8

Investigate patterns of association in bivariate data.

CCSS.Math.Content.8.SP.A.1

Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

CCSS.Math.Content.8.SP.A.2

Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

CCSS.Math.Content.8.SP.A.3

Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height*.

CCSS.Math.Content.8.SP.A.4

Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. *For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?*

These first two activities are used to have students begin to see the advantages of displaying data in different ways. They also begin to analyze data and choose the best representation for their data.

1.1.1 What does this representation tell me?

Visualizing Information

Lesson Objective: Students will analyze strengths and weaknesses of histograms and scatter plots.

1-1. BIRTHDAY BONANZA

About 18,000,000 people on Earth share your birth date (unless you were born on February 29). Do you think one of them might be in your class?

Place a sticky note with your initials above the month of your birthday on the **histogram** (a graph of data) on the wall. If there are sticky notes already above your month, place yours directly above them so that all of the notes form a neat "tower" over that month.

a. Which month has the most birthdays in your class? Which has the fewest? How can you tell by looking at the histogram?
b. Can you tell by looking at the graph whether anyone shares the same birthday as you? Why or why not?
c. As a class, discuss how you could find the other students in class who were born in the same month that you were. Look for an organized way to accomplish this.
d. Ask the name and birth date for each student born in the same month as you. Did you find a "birthday twin"? If you are the only student in class born in your month, find the students born in the month just before or just after yours.

1-2. SLEEPY TIME

How much sleep do you get at night? On a sticky dot, write the time you usually go to bed and the time you usually get up. For example, the dot below shows that a student goes to bed at 10:00 p.m. and wakes up at 6:00 a.m.

On the **scatterplot** poster on the wall, find the time that you go to bed on the **horizontal axis** (the line that lies "flat"). Then trace straight up from that point high enough to be even with the time that you get up on the **vertical axis** (the line that stands straight up) and place your sticky dot on the graph.

When all the data is collected, work with your team to answer the questions below.

a. What is the most common bedtime for your class members? How can you tell?
b. Which dots represent the students who get the most sleep? The least sleep? How much sleep does each of these students get?
c. If you were to go to bed an hour earlier, how would your sticky dot move? What if you were to get up an hour earlier?
d. In general, how much sleep do students in your class get?

2.1.1 How can I represent data?

Dot Plots and Bar Graphs

Lesson Objective: Students will analyze the strengths and weaknesses of various graphical representations of data.

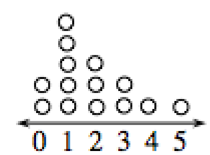
CCS Standard(s): **6.SP.4**

In this lesson, you will use multiple ways to visualize information and decide which ways of visualizing are better for various types of information. Consider these questions as you work through this lesson:

What does this organization of data tell us? What does it not tell us?
Is there a more useful way to visualize this information?

2-1. HOW MANY PETS?

Dot Plot



Many people have pets. Some people have many more than others. Some people have none.

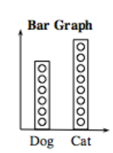
Write your initials on one of the sticky dots. Place a sticky dot above the number of pets you have on the class **dot plot**. If there are dots already over your number, place yours directly above them so that all of the dots form a neat "tower" over that number.

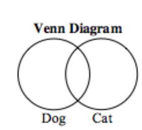
a. Which number of pets has the most occurrences in your class? Which has the fewest? How can you tell by looking at the dot plot?
b. How many pets do most of the students in your class have? How did you determine your answer?
c. Can you tell by looking at the dot plot whether anyone has the same type of pet(s) as you? Why or why not?
d. What other information can you determine from the class dot plot? Are there other questions you could ask about pets to create a different graph?

2-2. CATS AND DOGS

In the previous problem you looked at the number of pets that your classmates had. What if you want to know the types of pets people have?

For this activity, place your initials on two sticky dots. Then place one sticky dot on the class **bar graph** and one on the **Venn diagram** (shown below right). Then answer the questions below.





a. Were you able to place your dot easily on both graphs? Was there anyone who had a hard time placing their dot on either of the graphs? Explain.
b. Is there any information that is easier to see from looking at the bar graph? The Venn diagram?
c. What kinds of information are best represented in bar graphs? Venn diagrams?
d. Does the order of the bars matter on a bar graph?

2-5. LEARNING LOG

In your Learning Log, explain how bar graphs are similar to and different from dot plots. What additional information can a Venn diagram provide? Title this entry "Bar Graphs and Venn Diagrams" and label it with today's date.

2.1.2 How else can I represent data?

Histograms and Stem-and-Leaf Plots

Lesson Objective: Students will learn how to collect data and how to display the data in a stem-and-leaf plot and a histogram.

CCS Standard(s): **6.SP.4**

Have you ever noticed that when you are having fun, time seems to pass quickly, but when you are bored, what takes 10 minutes can feel like an hour? How good are the members of your class at estimating time? Today you will explore this question by learning new ways to collect, visualize, and analyze data.

2-11. ESTIMATING 60 SECONDS

Do you know how long 60 seconds is? Of course, it is 60 seconds long! But do you *really* know how long it takes for 60 seconds to pass? Would you be able to know when 60 seconds have passed without the help of a watch or a clock? Today you will conduct an experiment to see how accurately you and your classmates can do this.

Your task: You will close your eyes, put your head down, and estimate when 60 seconds have passed. Your teacher will tell you when to start your estimate. When you think 60 seconds have passed, raise your head and determine your time from the timer displayed by your teacher. Then record your time on the sticky note your teacher gave you.

Before you begin, discuss the following questions with your team:

What we are investigating
How many pieces of data are we collecting?
How will we measure it?
What unit of measurement will we use?

Be sure to remain quiet during the experiment. When everyone in the class has finished, be ready to share your time.

After the class data has been shared, discuss the following questions with your team. Be ready to explain your thinking to the class.

- Do you think the class would be more accurate at estimating 10 seconds or 60 seconds? What about 200 seconds? Why?
- What might affect the quality of the data?
- What do you expect the data to show?

2-12. USEFUL FORMS OF DATA

It is possible to organize items in a way that communicates information at a glance. In this problem, you will use the list of times from problem 2-11.

a. How could you rearrange the list to make it easier to find specific values? As a class, brainstorm ways to organize the data. Decide together how to rewrite the list.

b. One way to organize and display data is in a **stem-and-leaf plot**. The example at right represents the data 31, 31, 43, 47, 61, 66, 68, and 70. Think about how this plot is arranged and describe what you notice. For example, how would 42 be added to this plot? What about 102? Why do you think the space to the right of the 5 is blank.

Stem	Leaf
3	1 1
4	3 7
5	
6	1 6 8
7	0

c. Once the stem-and-leaf plot makes sense to you, work together to organize your class data from problem 2-11 in a similar way.

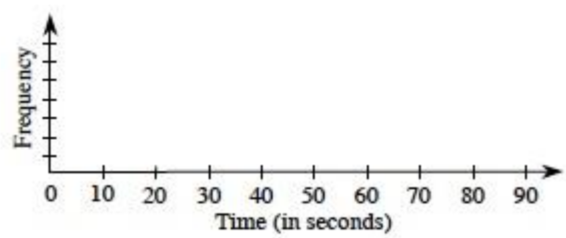
d. What do you notice about the class data? Discuss this with your team and then write down three observations you can make. Be ready to share your observations with the class and explain how you made them.

2-13. CREATING A HISTOGRAM

A histogram is another useful way to display data. You will explore one below.

a. In Lesson 2.1.1, you created a dot plot of pets. Why might a dot plot not be the best choice for graphing the 60-second data?

b. Another graph of data similar to a dot plot is called a histogram. Similar to a dot plot, this type of graph helps you see how many pieces of data are within each interval, such as between 0 and 10 seconds. Each interval is also called a **bin**. Following your teacher's directions, place a sticky note with your time from problem 2-11 on the class histogram. Copy the histogram into your notes, using the height of bars to represent the number of sticky notes.



c. Examine the graphed data. What statements can you make that describe how your class performed in the experiment? Were most students able to make a good estimate of 60 seconds? How can you tell?

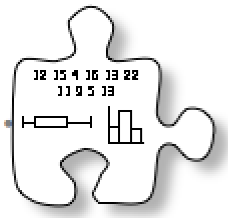
d. What if the histogram is formed in intervals of 20 seconds so it has five bins instead (0 – 19, 20 – 39, 40 – 59, 60 – 79, and 80 – 99)? What would be the same or different? Would it affect how you describe the performance of your class in the experiment? What if it was formed using intervals of 5 seconds? How would this change the histogram?

2-14. LEARNING LOG

In your Learning Log, compare the histogram in problem 2-13 with the stem-and-leaf plot from problem 2-12. What connections can you make between the two data displays? How are these data displays the same or different? How do they compare to a dot plot? Title this entry "Histograms and Stem-and-Leaf Plots" and label it with today's date.

8.1.1 How can I describe the data?

Measures of Central Tendency



Lesson Objective: Students will be introduced to measures of central tendency and will develop methods to find the mean, median, and range of a set of data.

CCS Standard(s): **6.SP.2**, **6.SP.3**, **6.SP.5c**

Data can be described and displayed in a variety of ways. You have already worked with many displays of data, such as tables and graphs. Today, you will focus on how to analyze and describe data numerically. Specifically, you will learn how to make statements to describe a set of data numerically. As you do today's activities, ask yourself these questions:

How can I describe the data?
What number best describes what is "typical?"
How are data measures useful?

8-1. TAKING A CENSUS

Have you ever heard of a census? A census is a collection of data that describes the people living in a country. The United States government performs a census every ten years. The government uses the data to learn such things as how the population is changing, where people live, what types of families exist, and what languages are spoken. For example, in the year 2000, there were 281,421,906 people surveyed for the census, and about 8,000,000 of them lived in New York, NY.

Today you will take a census of your class to answer the question, "*What is the size of a typical family for the students in your math class?*"

Your task: Obtain one sticky dot for each person in your team. On your sticky dot, write down the number of people in your family. Then place your sticky dot above the appropriate number on the class dot plot. Work with your class to answer the following questions:

- What is the difference between the largest piece and the smallest piece of data in your class? This difference is called the **range**. It is one way to measure the "spread" or variability of the data.
- What number falls right in the middle of all the class data when the data is sorted in order? This number is called the **median**.

8-2. Once each day, Erika tracks the depth of the water in her local creek. Her first nine measurements, in inches, are below.

16 15 13 12 17 14 11 9 11

a. What is the median of her data?
b. Erika's next three measurements, in inches, are 9, 10, and 9. What is the new median?

8-3. WHAT IS AVERAGE?

Now obtain one cube (or other manipulative) from your teacher to represent each person in your family.

a. Work with your classmates to organize yourselves into a human dot plot.
b. If the cubes were redistributed so that everyone in the class had the same number of cubes, how many cubes would each person have? This is called the **mean** (or the **arithmetic average**) of the data.

8-4. An **outlier** is a piece of data that is much larger or much smaller than the rest of the data. Imagine that a student with a family of 20 people joined your class. How do you think the range and the measures of central tendency (mean and median) of your class's data would change with this additional piece of data? Which measure would change the most?

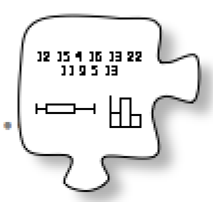
8-5. COMPUTING THE MEAN

In problem 8-3, you found the mean number of cubes in the class by sharing cubes evenly among all students. Now you will explore how this method translates into a mathematical strategy.

a. How could you use numbers and symbols to represent what happens when everyone puts all of their cubes together?
b. How could you use numbers and symbols to represent what happens when the big pile is distributed evenly among all of the people?
c. As you have discovered, one way to calculate the mean for a set of data is to add all of the data together (like combining all of the cubes) and then divide by the number of pieces of data (like distributing the cubes evenly among all of the people). Calculate the mean for the typical family size of students in your math class. How does your answer compare to the one that you got in problem 8-3? Be sure to record your work carefully.

8.1.2 What is a typical value?

Choosing Mean or Median

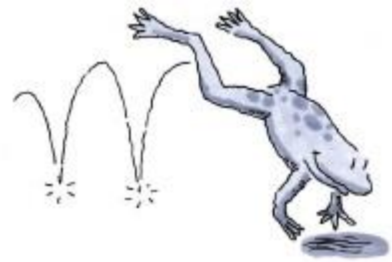


Lesson Objective: Students will compare two data sets using mean, median, and range. Students will choose between median and mean to describe the "typical" middle value in a distribution of data.

CCS Standard(s): **6.SP.2**, **6.SP.3**, **6.SP.5c**, **6.SP.5d**

You are exposed to a huge amount of information every day in school, in the news, in advertising, and in other places. It helps to have tools to be able to understand what the data means. Today you will turn your attention to what you can learn from data.

8-12. JUMPING FROG JUBILEE



In Lesson 1.1.4, you examined results of the Calaveras County frog-jumping contest. You saw that the 2009 competition winner jumped 20.5 inches farther than the 2008 winner. Do you think that the winner in 2009 was a special frog? Or were the frogs in 2009 better jumpers, in general, than the frogs in 2008?

With your team, brainstorm ways that you could compare the two groups of frogs. Recall some of the measures of central tendency from the previous lesson.

2008		2009	
Frog Name	**Jump Length**	**Frog Name**	**Jump Length**
Skeeter Eater	231.5 in.	For the Sign	252 in.
Warped	230 in.	Alex Frog	236.5 in.
Greg Crome Dome	229 in.	Shakit	231.5 in.
R.G.	227 in.	Six Mile Shooter	226.75 in.
The Well Ain't Dry	221.5 in.	Spare the Air Everyday	223.25 in.
Winner	220.5 in.	Hooper	223.25 in.
7 lb 8 oz. Baby	217 in.	Jenifer's Jumper	222.25 in.
Delbert Sr.	216.5 in.	Dr. Frog	185.25 in.

a. Your first job is to make a graphical representation of the data. Many statisticians say that the first and most important step in analyzing any data is to make a graphical representation. How can a representation help you analyze the data? The bin size of each histogram should be 10 inches. Why is a histogram a good choice?
b. What was more obvious when you looked at the histograms compared to looking at the list of data? Is there information that is easier to see on the histograms?
c. The technology tool shows the mean and the median of the data below the histogram. However, if the mean and the median were not labeled, would there be a way to determine the median or mean from a histogram alone? Explain.

8-13. Use the list of data in problem 8-12 or the histograms that you or your teacher made to continue to analyze and compare the frog jumps for 2008 and 2009. Answer the following questions.

a. What was the range of the jumps in each year? What does this tell you about the frog jumps?
b. What was the typical jump length of the frogs each year? How did you find this value?
c. Were the jumps all about the same, or were some jumps outliers? Name any outliers and explain why you think they are outliers.
d. Compare as completely as you can the 2008 jumps to the 2009 jumps. Compare the center, shape , spread (range), and outliers. Then draw a conclusion: were one year's frogs a better group of jumpers than the other? How do you know?

8-14. CHOOSING MEAN OR MEDIAN

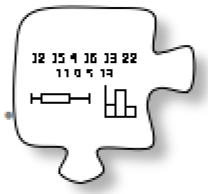
It is important to look at the distribution of the data when deciding whether to use the mean or the median.

a. In 2008, which represents a "typical" jump better, the mean or the median?
b. What if the 9^{th}-place jump in 2008 was very small, such as 160 inches? Redraw your histogram from problem 8-12 and include this new jump.

c. Without using your calculator, make a prediction. How does adding this outlier affect the mean and the median of the 2008 data?
d. Use your calculator to test your prediction from part (c).
e. Look at your histogram from part (b). Does the mean or does the median better represent a "typical" jump in 2008?
f. When does the median represent a typical jump better than the mean does?

8-15. Even though there are outliers in the data, the mean and the median in 2009 are almost the same. Why?

8.1.4 How can I display variability?



Box Plots and Interquartile Range

Lesson Objective: Students will construct and interpret box plots and compare sets of data.

CCS Standard(s): **6.SP.3**, **6.SP.4**, **6.SP.5a**, **6.SP.5b**, **6.SP.5c**, **6.SP.5d**

"Get your facts first, and then you can distort them as much as you please." This is a quote from Mark Twain, a famous American writer and humorist (1835–1910). He also said, *"Facts are stubborn, but statistics are more pliable."* What do you think he meant? Much of what you learn and interpret about different sets of data is based on how it is presented.

In this lesson, you will use several mathematical tools to look at data in different ways. As you work, use these questions to help focus your discussions with your team:

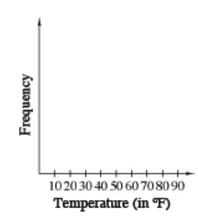
What can we conclude based on this representation?
What cannot be concluded based on this representation?
How are the representations related?

8-44. CLIMATE CHANGE?

Is the planet getting hotter? Experts look at the temperature of the air and the oceans, the kinds of molecules in the atmosphere, and many other kinds of data to try to determine how the earth is changing. However, sometimes the same data can lead to different conclusions because of how the data is represented.

Your teacher will provide you with temperature data from November 1, 1975, and from November 1, 2000. To make sense of this data, you will first need to organize it in a useful way.

a. Your teacher will assign you a city and give you two sticky notes. Label the appropriately colored sticky note with the name of the city and its temperature in 1975. Label the other sticky note with its city name and temperature in 2000.
b. Follow the directions of your teacher to place your sticky notes on the class histogram. Use the axis at the bottom of the graph to place your sticky note.

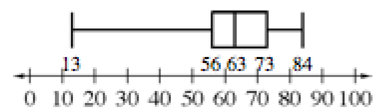


c. How many cities were measured for this study?
d. Describe the spread and shape of each of the histograms that you have created. Which measure of central tendency would you use to describe a typical temperature for each year? Justify your choice.

8-45. Look at the graphs of temperatures the class created. According to these histograms, what can you say about the temperatures on November 1 in 1975 and 2000? Do the graphs show that one date was warmer than the other? Or are the temperatures basically the same? Be prepared to share your reasoning.

8-46. The histograms your class made in problem 8-44 display data along the horizontal axis. Another way to display the data is to form a **box plot**, which divides the data into four equal parts, or **quartiles**. To create a box plot, follow the steps below with the class or in your team.

a. With a sticky dot provided by your teacher, plot the 1975 temperature for your city on a number line in front of the class.
b. What is the median temperature for 1975? Place a vertical line segment about one-half inch long marking this position above the number line on your resource page.
c. How far does the data extend from the median? That is, what are the minimum and maximum temperatures in 1975? Place vertical line segments marking these positions above the number line.
d. The median splits the data into two sets: those that come before it and those that come after it when the data is ordered from least to greatest, like it is on the number line. Find the median of the lower set (called the **first quartile**). Mark the first quartile with a vertical line segment above the number line.
e. Look at the temperatures that come after the median on your number line. The median of this portion of data is called the **third quartile**. Mark the third quartile with a vertical line segment above the number line.
f. Draw a box that contains all of the data points between the first and third quartiles. Your graph should be similar to a box with outer segments like the one at right.



g. What does the box plot tell you about the temperatures of the cities in 1975 that the dot plot did not?

8-47. With your team, create a new box plot of the temperature data for the same cities on November 1, 2000 on your resource page. Be sure to identify each of the values below.

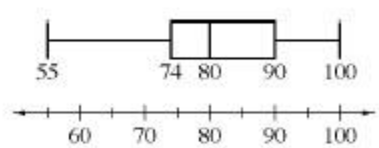
- The minimum and maximum data values (endpoints of the segments)
- The median temperature
- The first and third quartiles

8-48. USING A BOX PLOT TO MEASURE SPREAD

a. By looking at your box plots from problems 8-46 and 8-47, determine which year of temperature data has the greatest range.
b. In addition to range and mean absolute deviation, there is another way to calculate spread. The **interquartile range (IQR)** is found by calculating the difference between the third quartile and the first quartile. It is the range of the middle 50% of the data. Calculate the interquartile range for the temperature data from 1975.
c. Calculate the interquartile range for the temperature data from 2000.
d. Now which year of temperature data seems to have the most spread measured with IQR.
e. What advantages are there in using the IQR to measure spread instead of range or mean absolute deviation?

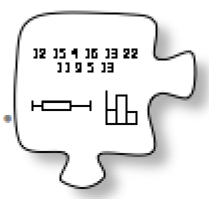
8-49. Look at the histograms and box plots of temperatures the class created. Compare the center, shape, spread, and outliers for both sets of data. According to these histograms, what can you say about the temperatures on November 1 in 1975 and 2000? Do the graphs show that one date was warmer than the other? Or are the temperatures basically the same? Be prepared to share your reasoning.

8-50. The box plot below shows the speeds of cars measured by the Royal Canadian Mounted Police (RCMP) on a certain section of roadway over the course of a week. Speeds are given in kilometers per hour, or kph.



a. What was the median speed measured? What were the highest and lowest speeds?
b. Did most drivers go a particular speed? How do you know?
c. Can you tell how many drivers' speeds were measured that week?
d. If the RCMP measured the speeds of 332 cars that week, then how many drivers were going faster than 80 kph? How many drove faster than 90 kph? Explain how you know.
e. What is the IQR for this set of data?
f. Can you tell if the speeds between 80 kph and 90 kph were closer to 80 kph or closer to 90 kph? Explain.

8.1.5 Which representation should I use?



Comparing and Choosing Representations

Lesson Objective: Students will construct three different representations of a single set of data. They will then determine what kinds of information the representation communicates and decide which representation is most useful for answering a specific question.

CCS Standard(s): **6.SP.4**

When you want to represent data visually, you have several different representations to choose from. In this chapter, you have used three different representations: histograms, stem-and-leaf plots, and box plots. Today you will compare these representations to decide when one might be a better way to communicate information from a set of data than the other two representations would be.

8-61. DATA SET DECISIONS

Ms. Anderson's math class took a chapter test. She wants to find a way to display the data from the class test scores so that the students can easily understand the data.

- Michaela thinks that the best representation is a stem-and-leaf plot because it organizes information quickly.
- Gabe thinks that Ms. Anderson should use a histogram because it gives a picture that is easier to look at.

- Geri does not agree with either Michaela or Gabe. She thinks a box plot shows important information that the other two representations do not show.

a. Discuss with your team which representation Ms. Anderson should use to display the data. On your paper, justify your choice.
b. The test scores for the class are: 78, 62, 91, 51, 55, 93, 76, 82, 65, 85, 79, 83, 55, and 72. On the Lesson 8.1.5A Resource Page make a stem-and-leaf plot, a histogram, and a box plot for this data. Use bin widths of 10. Be sure to include a title and labels on all the representations.

8-62. The principal of Ms. Anderson and her class's school, Mr. Siebers, wants to get information about Ms. Anderson's latest test scores (listed in problem 8-61). He is interested in the class median, but he also wants to know what percentage of students scored lower than a C+ (less than 77 points) on the test.

a. What is the median of the class test scores? Which graph lets you see the median most quickly?
b. What percentage of the students scored lower than a C+? Which graph lets you see this most easily?

8-63. Mrs. Smith, another math teacher, also wants to look at Ms. Anderson's test scores. She wants to see information about how many students earned more than 60 points on the test.

a. Which representation(s) would show her this information? Justify your answer.
b. Is your answer to this problem the same as or different from your answer to part (a) of problem 8-62? Why or why not?

8-64. Ms. Anderson still does not know what kind of representation she should use. She knows that each representation shows different kinds of information. Using the representations that you made in problem 8-61:

- List and explain what measures of central tendency or other information will be easily read from each kind of graph listed below.
- List the information that will not be easy to obtain. Justify your answers.
    a. Histogram
    b. Stem-and-leaf plot
    c. Box plot

8-65. LEARNING LOG

Write a Learning Log entry that describes what you have learned about descriptive statistics in this section. For example, what kinds of data displays can you use? What numerical information can you give to describe center and spread? How can you describe the shape of data? Title this entry "Representations, Center, Shape, and Spread of Data" and include today's date.

8.2.1 How can I answer the question using statistics?

Statistical Questions

Lesson Objective: Students will identify statistical questions. They will understand that statistical questions can anticipate variability in the answers.

CCS Standard(s): **6.SP.1**, **6.SP.5b**

All of the collecting, organizing, analyzing, interpreting, and presenting data you have done so far in this course is called *statistics*. Statistics is an important branch of mathematics. It allows you to make conjectures about data sets and populations. It allows you to report overall trends instead of just guessing what is happening.

Can you think of ways you have used statistics in your daily life? The basic statistics you have learned so far is useful in many parts of everyday life. If you want to buy pizza for your school pizza party, for example, you can use statistics to make a good estimate of how many pizzas to purchase so that everyone gets enough to eat and there is not much left over. How could you go about doing this? One way is to take a survey and ask a statistical question.

Statistical questions are questions that have answers with variability. For example, "*How tall are you?*" is not a statistical question because there is only one answer, and you would not expect various answers. However, "*How tall are the students in your class?*" is a statistical question, because you would get a variety of answers from the students in your class.

8-71. Read each question below and decide if it is a statistical question. If possible, reword the question so that it is a statistical question.

a. How old are you?
b. How many books did each student in your school read last year?
c. What is your favorite color?
d. How many days are in a week?
e. What is the batting average of a major league baseball player?
f. How many points did the basketball team score during their last game?
g. How much cereal is in a box of Tase-T-Squares?
h. How long does it take for a student in your class to get to school?
i. What is the capital of Maine?

8-72. SURVEYING THE CLASS

Work with your team to make up some questions to ask the class. Then design the graphs that will best represent the answers.

Your task: Write down three questions that you could ask students in the class. The questions should help you learn more about them. Think about a way to display the responses for each question. Then contribute your ideas to your team and, as a team, decide on your three favorite questions to ask. For each question, decide whether the answers should be shown on a histogram, a box plot, a bar graph, a Venn diagram, or another (better) way to show the data, if there is one.

Try to ask questions that will give you the information you want, questions that allow for variability in the answers. For example, asking *"Do you play sports?"* will get "*yes*" and "*no*" answers, with no information about what types of sports people play (no variability in the answers). However, the question, "*What sport(s) do you play?*" will enable you to learn if you have soccer players, swimmers, or other athletes in your class. This type of question will gather a variety of answers.

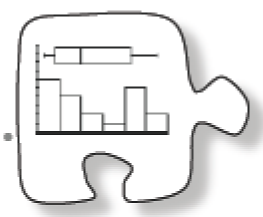
Some sample questions are provided below to help you get your conversations started.

- *"How many hours was the longest car or bus trip you have been on?"*
- *"How many cousins do you have?"*
- *"How did you get to school this morning?"*

8-73. Analyze the data you collected from your class. Was there variability in the answers? Work with your team to create a poster for one or more of your questions using the statistics that you have learned so far. Your poster will need to include the following information:

- The question(s) asked;
- A table or list of the data collected;
- The number of responses;
- Details about how the data was collected and any units of measurement necessary;
- A graphical representation of the responses;
- Analysis of the data, including center, shape, and spread; and,
- Your conclusions and how they are supported by appropriate statistics.

8.1.2 How can I compare the results?



Comparing Distributions

Lesson Objective: Students will compare two populations based on making inferences from samples. Students will quantify thedifference between the medians as a multiple of the IQR.

CCS Standard(s): **7.SP.3**

Today you will continue to develop your methods for comparing two distributions of data.

8-19. Josh is just starting a round of golf. This first hole is 130 yards long. He needs to decide which club to use for his first shot. He has kept careful records about how close his first shots came to the hole, all from this same distance. His records include data from his use of two different golf clubs, a wedge and an 8-iron, over the past year.

Wedge, distance to hole (yards): 0 3 1 2 7 2 15 25 5 22
8-iron, distance to hole (yards): 19 12 12 8 3 11 5 7 10 13 8 10 11 20

a. Create a histogram and box plot for each of the clubs. Place the box plot above the histogram on the same number line for each club to make a combination plot. Use a bin width of 5 yards.

b. To find the "typical" distance that Josh hits the ball from the hole with a wedge, is the mean or the median a better choice? Find the typical distance Josh hits the ball from the hole with a wedge and compare it to the typical distance he hits the ball with an 8-iron.
c. Advise Josh which club to use. Explain your thinking.

8-20. It is just as important to consider the spread of the data as it is to consider the center when comparing data sets.

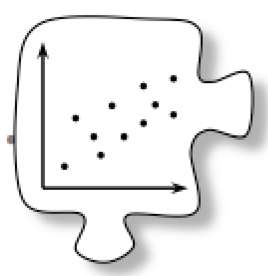
a. Calculate the Interquartile Range (IQR) for each golf club in the previous problem. With which club is Josh more consistent?
b. If Josh decided to use the 8-iron, he could "typically" expect to hit the ball so that it lands between 8 and 12 yards from the hole. This is indicated by the box on the box plot display and corresponds with the IQR. If Josh decided to use the wedge, what is a "typical" interval of distances from the hole he could expect the ball to land?
c. Compare the typical interval of distances for the 8-iron with the interval you found for the wedge. Do you wish to modify your advice to Josh? Explain.

8-21. Mr. Webb has only one more starting position available on his basketball team, but two students have tried out for it. He wants to choose the student who is likely to score the most points. The two students from whom he can choose are described below.

a. In her most recent games, Jana scored: 7 46 9 6 11 7 9 11 19 7 9 11 9 55 11 7 points, while Alejandra scored 13 15 9 18 13 17 17 15 points. Which girl has the higher average (mean) number of points?
b. Which student do you think Mr. Webb should select and why? Use parallel box plots (two box plots on the same number line) to support your explanation.
c. Why was the mean not a good measure of the girl's typical performance?
d. Calculate the IQR to measure the variability of each girl's performance.
e. Who had the higher median and by how many points? How large is the difference between the medians measured by how many IQRs would fit into it?

7.1.2 Is there a relationship?

Organizing Data in a Scatterplot



Lesson Objective: Students will create scatterplots and identify whether there is a relationship between two sets of data. Students will draw a line of best fit and use it to make predictions.

CCS Standard(s): **8.SP.1**

In Lesson 7.1.1, you looked at single data sets, such as world population. Often, you need to compare two measurements to answer a question or to see a connection between two types of data. For example, comparing the odometer reading of a car to the price of a car can help determine if these factors are related. In this lesson, you will study scatterplots, a new tool for visually presenting data, as a way to relate two sets of measurements. You will be asked to analyze the data to see if you can make predictions or come to any conclusion about the relationships that you find.

As you work with your team today, use these focus questions to help direct your discussion:

How can I organize data?
Can I use this data to make a prediction?
What does a point represent?
Is there a connection between the two variables?

7-13. HOW MUCH IS THAT CAR?

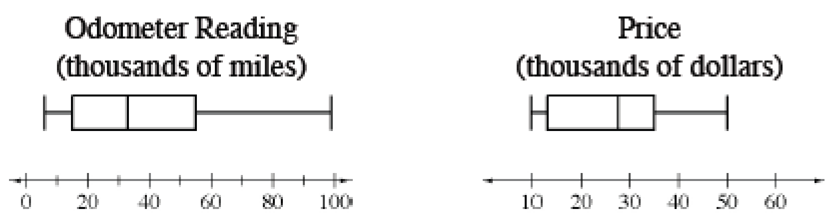
Nate and Rick were discussing cars again. Nate claimed that cars with lower odometer readings were more expensive than cars with higher odometer readings. His evidence was that his car with 23,000 miles was worth more than Rick's car with 31,000 miles. To investigate Nate's claim, the boys collected data from several car advertisements and found the information in the table at right.

Nate's Data from Car Ads

Odometer Reading (thousands of miles)	Price (thousands of $)
35	$38
55	$16
6	$50
28	$30
50	$26
31	$35
15	$28
99	$10
99	$13

Does the information in the table support Nate's claim? That is, do you believe Nate's claim that cars with a lower odometer reading cost more money?

7-14. Melissa looked at the data from problem 7-13 and said, *"I need to be able to see the data as a picture. I cannot tell if there is a relationship from the lists of numbers."* She decided to use a box plot. Her box plots for odometer reading and price are shown below. Do these pictures help you decide if Nate is correct? Why or why not?



7-15. Melissa wondered if a coordinate graph could help determine if there was a relationship in Nate's data from problem 7-13.

a. Follow the directions below to create a scatterplot of the data for Melissa or use 7-15 Student eTool (Desmos).
   - Set up a graph showing Odometer Reading on the *x*-axis and Price on the *y*-axis.
   - Label equal intervals on each axis so that all of the data will fit on the graph.
   - Plot the data points from problem 7-13.

b. Describe the scatterplot you just created. What do you notice about how the points are placed on the graph? Do you see any patterns?

c. Place an additional point on your graph for Nate's car that has an odometer reading of 23,000 miles. Explain your strategy for deciding where to put the point.

d. When a relationship exists, one way to help show a trend in the data is to place a line or curve that, in general, represents where the data falls. This line, sometimes called a **line of best fit**, does not need to touch any of the actual data points. Instead, it shows where the data generally falls. The line is a mathematical model of the data. Models of data help you describe the data more easily and help you make predictions for other cars with different mileages.

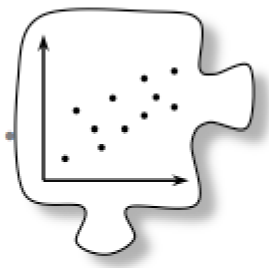
   With your team, decide where a line of best fit could be placed that would best model the data points. Are there any limits to where your line makes sense?

e. Using the line of best fit, can you predict the price of a car with an odometer reading of 80,000 miles? If so, explain how the line of best fit helps. If not, explain why it is not helpful.

f. Based on the scatterplot, would you agree with Nate's claim that cars with a higher odometer reading cost less? Use the scatterplot to justify your answer.

7.1.3 What is the relationship?

Identifying and Describing Association



Lesson Objective: Students will continue to develop their understanding of different associations and will consider the direction of an association. Students will create and use scatterplots to make predictions, if possible, and identify when it is not possible to make predictions.

CCS Standard(s): **8.SP.1**, **8.SP.2**

When is it reasonable to make a prediction? For example, when you know the height of a tree, can you predict the size of its leaves? Or if you know the outdoor temperature for the day, can you predict the number of glasses of water you will drink during the day?

In Lesson 7.1.2, you found that some data sets were related and others were not. In this lesson, you will look at different situations and decide if they show relationships that allow you to make predictions. As you work with your team today, use these focus questions to help direct your discussion:

When one value goes up, what happens to the other one?
Is there a relationship between one thing changing and the other changing?
Can I make a prediction?

7-23. The students in Mr. Carle's science class have been experimenting with different factors that they think may affect how tall a plant will grow. Each team planted seeds in several pots using different experimental conditions.

With your team, read the questions that the students investigated. Write a team prediction (hypothesis) for the results for each experiment. Assume that all other variables are controlled (meaning that they will not affect the experimental outcome). Write your hypothesis so that it indicates a directional relationship between the independent and dependent variables.

a. Does the amount of fertilizer affect the plant height?
b. Does how deep the seed is planted in each pot affect the plant height?
c. Does the number of seeds in each pot affect plant height?
d. Does the size of the pot affect plant height?
e. Does the number of hours of sunlight per day affect the plant height?

7-24. After three weeks, the teams measured the heights of their plants and recorded the data. The Team 1 data table and the question Team 1 investigated are included below. On graph paper, make a scatterplot for the data gathered. Be sure that you:

- Clearly label your axes.
- Mark the scale at equal intervals.
- Title your graph appropriately (such as with the experimental question).

Team 1: Does the amount of fertilizer affect the plant height?

Amount of fertilizer over 3 week period (ml)	5	10	15	20	25	30
Height of plant (cm)	14	17	21	31	33	40

7-25. Your teacher will assign your team one of the remaining sets of data. Prepare a scatterplot poster for your assigned set of data. Be sure your graph has a title and that the axes are correctly labeled. Also make sure that the points on your graph will be easily seen from across the room.

Team 2: Does the depth of seed in each pot affect the plant height?

Depth of seed in pot (cm)	3	6	9	12	15	18
Height of plant (cm)	24	21	18	12	6	0

Team 3: Does the number of seeds in each pot affect the plant height?

Number of seeds planted in pot	1	2	3	4	5	6
Height of plant (cm)	21	24	27	21	25	20

Team 4: Does the size of the pot affect the plant height?

Diameter of pot (cm)	3	6	9	12	15	18
Height of plant (cm)	21	20	16	20	22	17

Team 5: Does the number of hours of sunlight per day affect the plant height?

Amount of light per day (hours)	1	3	5	7	9	11
Height of plant (cm)	3	10	11	24	30	34

7-26. Examine your scatterplot for Team 1 from problem 7-24. Also look at the scatterplot posters created by your classmates for Teams 2 through 5. Then answer the following questions.

a. For each of the graphs of data, does there appear to be a relationship? Describe the relationship by completing the appropriate sentence below.

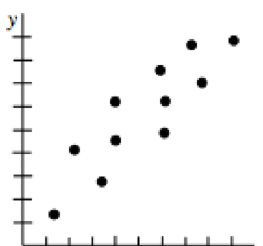
As ____ gets larger, then ____ gets ____.

OR

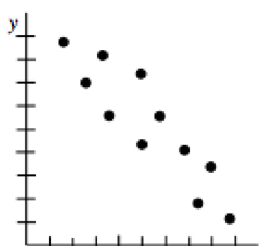
There appears to be no relationship between ____ and ____.

b. DIRECTION OF ASSOCIATION

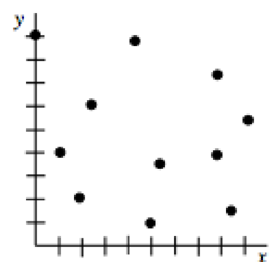
In a scatterplot, if there appears to be no relationship between the variables, then the points in the scatterplot have **no association**. But if one variable generally increases as the other variable increases, there is said to be a **positive association**. If one variable generally decreases as the other variable increases, there is said to be a **negative association**. See some examples below.



Positive Association



Negative Association

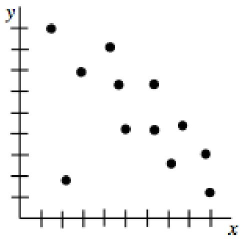


No Association

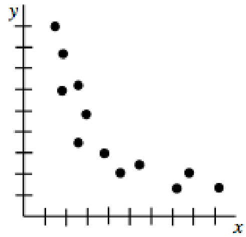
Review each of the graphs of the plant-experiment data and decide on the direction of the association. That is, decide if there is a positive association, a negative association, or no association.

c. FORM OF ASSOCIATION

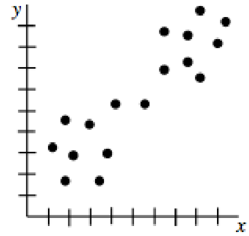
When there is a positive or negative association, the shape of the pattern is called the **form** of the association. Associations can have a **linear form** or a **non-linear form**, and the form can be made up of **clusters** of data. See some examples below.



Negative **linear** association (for example, gas mileage decreases as the weight of cargo on a truck increases)



Negative **non-linear** association (for example, temperature of a cup of coffee decreases over time)



Positive **linear** association with **clusters** (for example, height increases as shoe size increases; one cluster is mostly girls and the other cluster is mostly boys)

Review each of the graphs of the plant-experiment data that has an association, and decide on the form of the association. That is, decide if it is linear or non-linear, and whether it has clusters or no clusters.

d. OUTLIERS

An **outlier** is a piece of data that does not seem to fit into the pattern. Do there appear to be any outliers in any of the example scatterplots above?

e. Now go back and look at your team predictions (hypotheses) for each question in problem 7-23. Were your predictions accurate? Explain your reasoning.

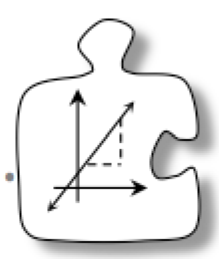
7-27. When there is an association, predictions can be made. One way to help make predictions is to draw a line (or curve) of best fit for the data.

a. Find your graph of Team 1's data from problem 7-24. Work with your team to draw a *straight* line that models (represents) the trend of the data on this graph. The line does not need to intersect each of the points, and the line does not need to pass through the origin.
b. Now use your line of best fit to predict the height of the plant when 12 milliliters of fertilizer are given to the plant over a 3-week period.
c. What is the *y*-intercept? Interpret the *y*-intercept in this situation.

7-28. LEARNING LOG

In this Learning Log entry, describe how a graph can help you tell if there is an association between two sets of data. That is, what does it look like if there is an association? What does it look like if there is no association? What are the forms an association might have? What does an outlier look like? Title this entry "Associations" and label it with today's date.

7.3.1 How can I use an equation?



Using Equations to Make Predictions

Lesson Objective: Students will use their understanding of $y = mx + b$ to write the equation for a line of best fit to represent scattered data that is roughly linear by calculating the slope from two points. Students will use this equation and a graph to make and justify predictions. Students will interpret the slope and y-intercept of a best fit line in context.

CCS Standard(s): **8.SP.3**

Previously, you have learned to find and extend patterns in data and to make predictions using rules, equations, and graphs. Today you will apply these mathematical tools to a real situation in which your data does not make a perfect pattern.

7-87. AURÉLIE'S BIKE RIDE

To prepare for a 30-mile ride for charity, Aurélie has been biking every weekend. To predict how long the charity event will take her to complete, Aurélie has been keeping track of her time and distances. She tries to ride at a constant pace, but of course that is not easy to do.

a. Would a box plot, scatterplot, or histogram be most useful in helping Aurélie make a prediction?
b. Considering that Aurélie is trying to predict how long it will take to complete the ride, which is the independent variable and which is the dependent variable? Make a scatterplot.
c. Use a line of best fit to predict how long it will take Aurélie to complete the charity ride. Remember that the line does not need to intersect each of the points and that the line does not need to pass through the origin to model the data.

Aurélie's Data

Weekend	Time (minutes)	Distance (miles)
Feb 6	68	7
Feb 13	88	17
Feb 20	35	3.5
Feb 27	150	25
Mar 6	104	14
Mar 13	80	11

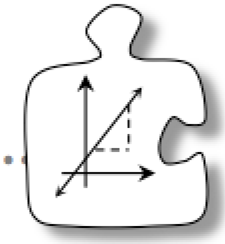
7-88. USING AN EQUATION TO MAKE PREDICTIONS

Sometimes it is more convenient and more accurate to use an equation to make a prediction rather than making a prediction by reading a graph.

a. Choose two points that lie on your line of best fit. These points can be given data points or lattice points on the coordinate grid. Use these two points to calculate the slope of your line of best fit. What does the slope mean in terms of Aurélie's bike riding?
b. What is the y-intercept of your line of best fit? What does the y-intercept mean in terms of Aurélie's training?
c. Write the equation of the line of best fit in $y = mx + b$ form. Identify your variables.
d. Use your equation to predict how long the charity ride would take Aurélie to complete. How did your prediction compare to the prediction you made from the graph in part (c) of problem 7-87?

7.3.2 How can I describe the association?

Describing Association Fully



Lesson Objective: Students will continue to write the equation for a line of best fit to represent scattered data that is roughly linear. Students will be able to fully describe an association between two numerical variables using form, direction, strength, and outliers.

CCS Standard(s): **8.SP.2**, **8.SP.3**

You have already seen that form, direction, and outliers can be used to describe an association. Today you will complete the description of an association by considering strength.

7-97. Do you remember the "Newton's Revenge" roller-coaster problem from Chapter 1? You will now return to the problem, using your knowledge of $y = mx + b$ to solve it. The problem (problem 1-24) is summarized below.

Newton's Revenge, the new roller coaster, has a tunnel that thrills riders with its very low ceiling. The closest the ceiling of the tunnel ever comes to the seat of the roller-coaster car is 200 cm. Although no accidents have been reported yet, it is said that very tall riders have stopped riding the roller coaster.

a. To help determine whether the tunnel is safe for any rider, no matter how tall, you collected data in problem 1-24. The height and reach were both measured in centimeters. Use the data from the class at the right. Which is the dependent variable? As you plot the data, consider whether the plot is useful for making predictions. If not, can you change the plot to make it more useful?

b. Work with your team to draw a straight line that models the data on this graph. Use two points from the line that you drew to calculate the slope. Then find the equation of this line. Identify the variables.

Height (cm)	Reach (cm)
166.4	127
169	133
172.8	133
179	139
170	139
183	137
162.5	121
165	126
157.5	128
165	123
169	132
156	119

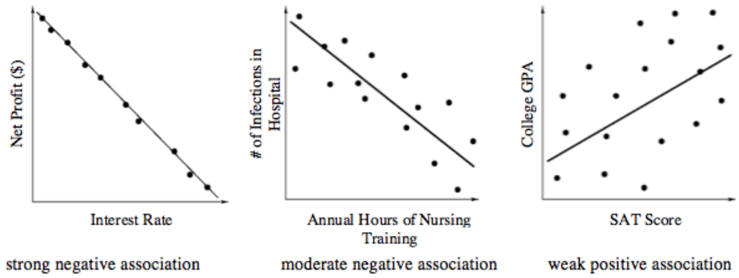
7-98. Once you have an equation that can best represent the data in problem 7-97, you will be able to use the equation to verify that the roller coaster is safe. The amusement park wants Newton's Revenge to be safe for tall riders. For example, remember that one of the tallest NBA players in history was Yao Ming, who is 7 feet 6 inches (about 228.6 cm) tall. Is the roller coaster safe for him? Explain.

7-99. The slope and y-intercept can sometimes give you more information about the situation you are studying, but sometimes their interpretation makes no sense.

a. What is the slope of the line of best fit in Newton's Revenge? Interpret the slope in this problem situation. Does your interpretation make sense?
b. Interpret the y-intercept in this situation. Does your interpretation make sense?
c. Make a conjecture about why your interpretation for the y-intercept does not make sense in a real-world situation, but can still be part of the equation to model this situation.

7-100. STRENGTH OF ASSOCIATION

Although considering the direction of an association (positive or negative) is important in describing it, it is just as important to consider the **strength of the association**. Strength is a description of how much scatter there is in the data away from the line of best fit. Some examples are below.



a. Fully describe the association in “Newton’s Revenge” (problem 7-97), considering the form, direction, strength, and outliers. See the Math Notes box in Lesson 7.1.3 to review some of these terms.
b. Look back at Aurélie’s scatterplot from the previous lesson (problem 7-87). Fully describe the association between the distance she bikes and the time it takes her.

7-101. LEARNING LOG

Make an entry in your Learning Log. Title it “Describing Association” and label it with today’s date. What are the four things you need to consider when describing an association? Give examples of each item. What information does an equation give you? When might an equation be more useful than a graph? What additional information does the slope tell you about the association?

Jumping Frog Jubilee Contest Data

2008

Frog Name	Jump Length (inches)
Skeeter Eater	231.5
Warped	230
Greg Crome Dome	229
R.G.	227
The Well Ain't Dry	221.5
Winner	220.5
7 lb 8 oz. Baby	217
Delbert Sr	216.5

2009

Frog Name	Jump Length (inches)
For the Sign	252
Alex Frog	236.5
Shakit	231.5
Six-Mile Shooter	226.75
Spare the Air Every Day	223.25
Hooper	223.25
Jenifer's Jumper	222.25
Dr. Frog	185.25

High-Temperature Data

	City	1975 (°F)	2000 (°F)
1	Anchorage, AK	13	33
2	Spokane, WA	52	44
3	Billings, MT	62	44
4	Juneau, AK	29	45
5	Bangor, ME	53	48
6	Bellingham, WA	52	53
7	Albuquerque, NM	67	53
8	Denver, CO	60	54
9	Portland, OR	57	54
10	Seattle, WA	54	54
11	Boston, MA	60	56
12	New York, NY	56	58
13	Duluth, MN	55	60
14	Bismarck, ND	66	61
15	Baltimore, MD	61	62
16	Washington, D.C.	62	62
17	Philadelphia, PA	59	62
18	El Paso, TX	83	65
19	Lansing, MI	55	66
20	Phoenix, AZ	77	67
21	San Francisco, CA	67	67
22	Sacramento, CA	71	68
23	Los Angeles, CA	71	69
24	Raleigh, NC	63	70
25	Des Moines, IA	72	73
26	Kansas City, MO	73	74
27	Chicago, IL	60	75
28	Oklahoma City, OK	76	76
29	Louisville, KY	70	76
30	Topeka, KS	74	77
31	Atlanta, GA	66	79
32	Orlando, FL	79	82
33	Baton Rouge, LA	81	84
34	Honolulu, HI	84	85
35	New Orleans, LA	80	86

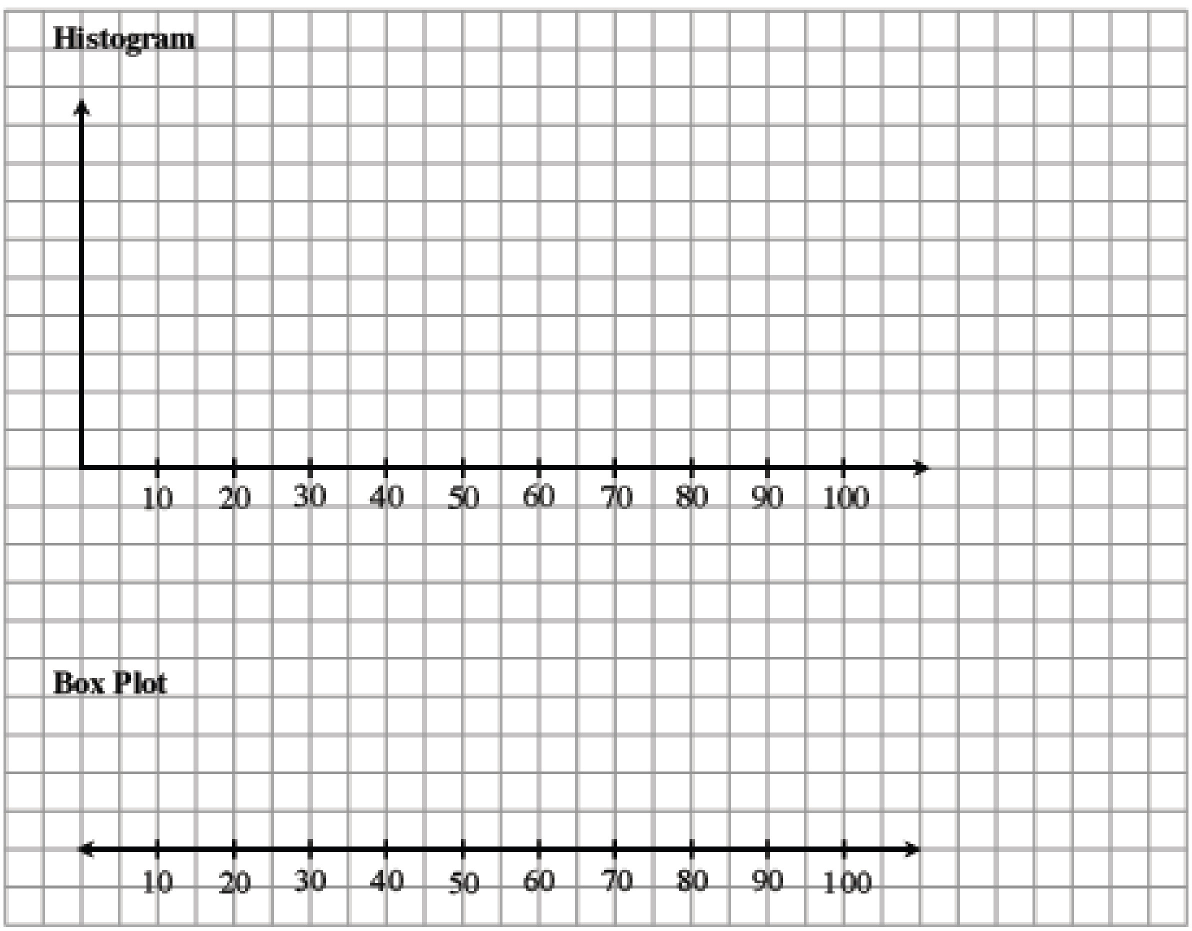
1975

2000

Data-Set Decisions

Stem-and-Leaf

Stem	Leaves



Comparing Mileage and Price