THE UNIVERSITY OF KANSAS
Achievement & Assessment Institute

ASKING QUESTIONS AND DESCRIBING DATA

6.SP.1-3

CONTENTS

The types of documents contained in the unit are listed below. Throughout the unit, the documents are arranged by lesson.

Enhanced Learning Maps developed these materials under a grant from the Department of Education, PR/Award # S368A150013. Contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

ASKING QUESTIONS AND DESCRIBING DATA

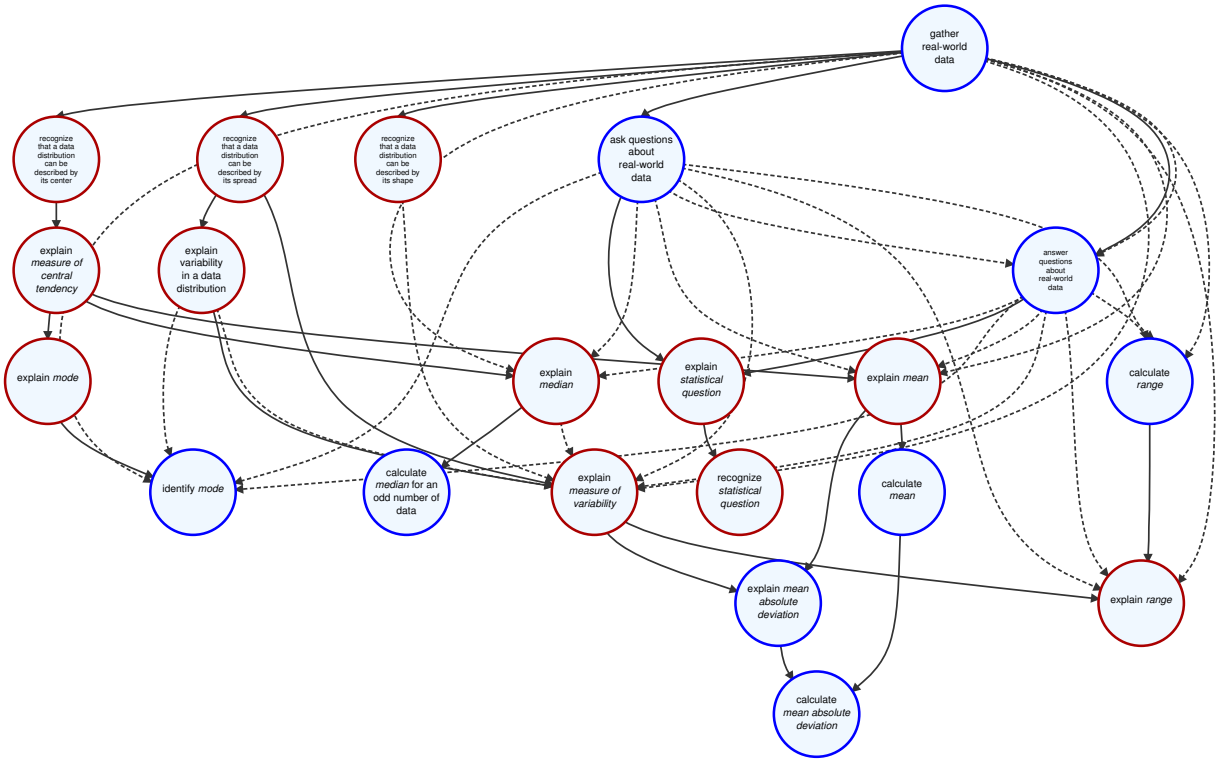
LEARNING MAP INFORMATION

STANDARDS

6.SP.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.*

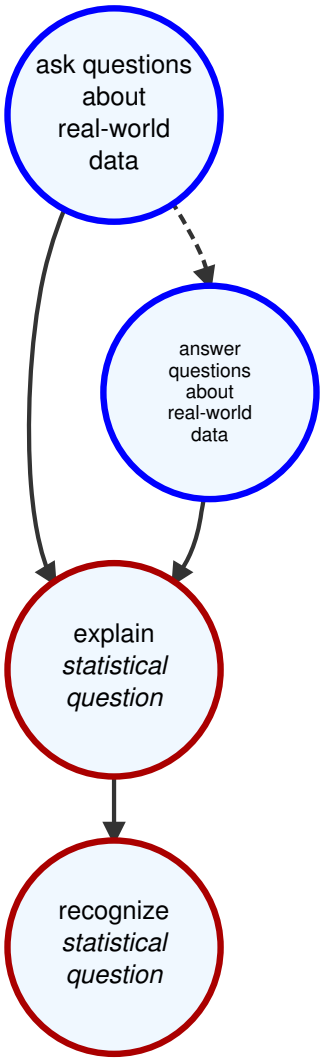
6.SP.2 Understand that a set of data has a distribution that can be described by its center (mean, median, or mode), spread (range), and overall shape and can be used to answer a statistical question.

6.SP.3 Recognize that a measure of center (mean, median, or mode) for a numerical data set summarizes all of its values with a single number, while a measure of variation (range) describes how its values vary with a single number.



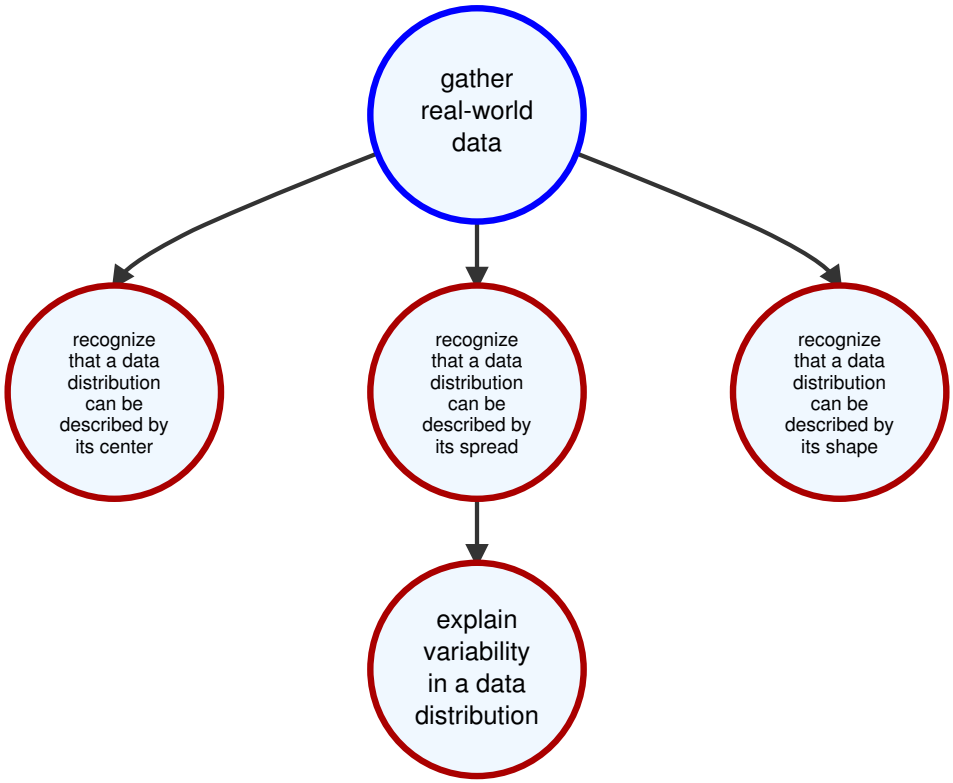
**Learning map model of* 6.SP.1-3

6.SP.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.*



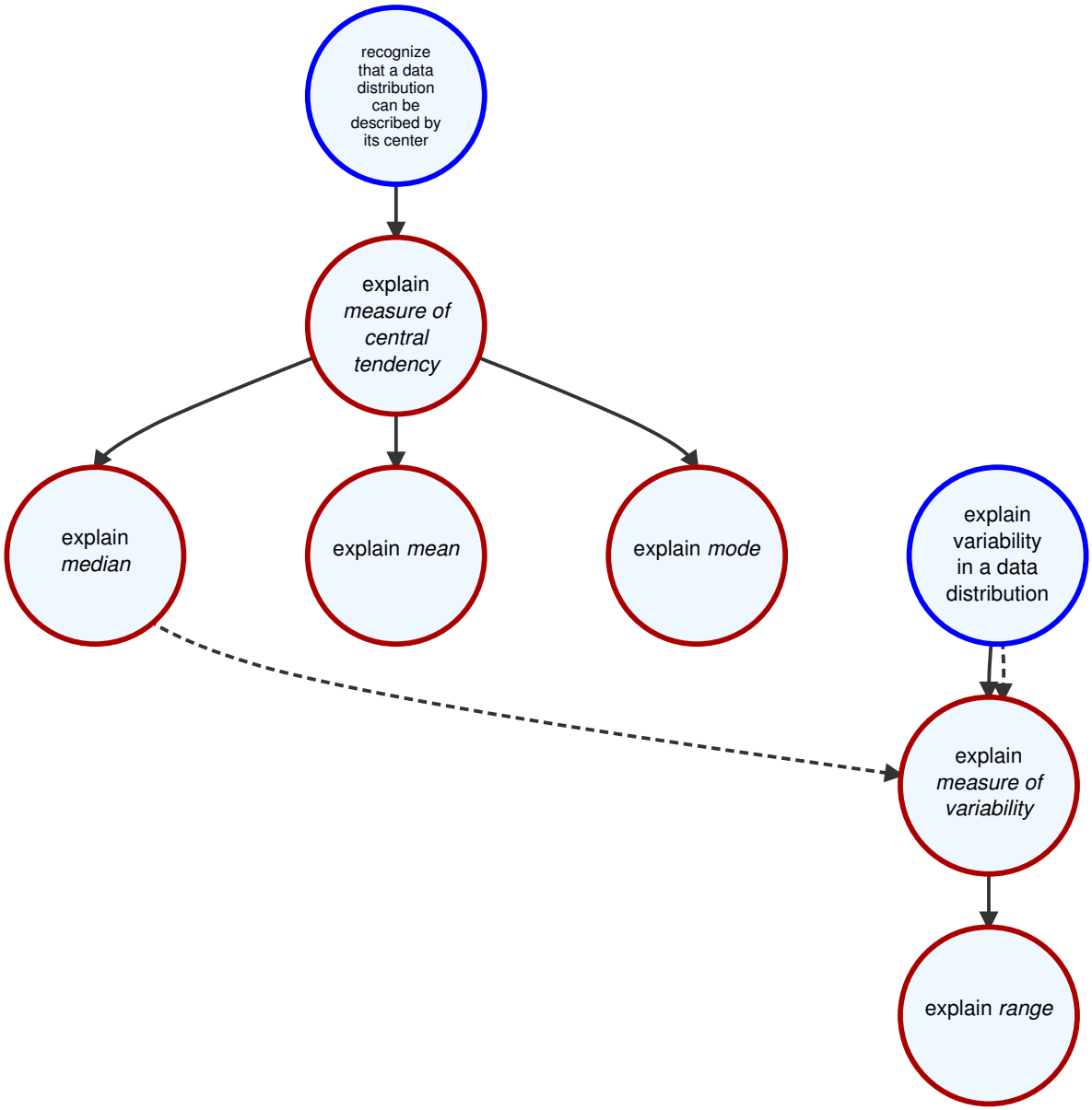
**Learning map model of* 6.SP.1

6.SP.2 Understand that a set of data has a distribution that can be described by its center (mean, median, or mode), spread (range), and overall shape and can be used to answer a statistical question.



**Learning map model of* 6.SP.2

6.SP.3 Recognize that a measure of center (mean, median, or mode) for a numerical data set summarizes all of its values with a single number, while a measure of variation (range) describes how its values vary with a single number.



**Learning map model of* 6.SP.3

Node Name	Node Description
ANSWER QUESTIONS ABOUT REAL-WORLD DATA	Answer questions about a set of real-world data, such as how many are in each category, how many more or less are in one category than another, etc.
ASK QUESTIONS ABOUT REAL-WORLD DATA	Ask questions about real-world data, such as how many are in each category, how many more or less are in one category than another, etc.
CALCULATE *MEAN*	Divide the sum of all data by the number of observations.
CALCULATE *MEAN ABSOLUTE DEVIATION*	Calculate the mean absolute deviation by adding the absolute value of the difference between each data value and the mean and then dividing by the number of data values.
CALCULATE *MEDIAN* FOR AN ODD NUMBER OF DATA	Calculate the median for an odd number of data by identifying the middle datum when a set of data is ordered in numerical value.
CALCULATE *RANGE*	Calculate the range for a given set of data by subtracting the minimum from the maximum.
EXPLAIN *MEAN*	Make known your understanding that the mean is the arithmetic average of a set of data in relation to its context. Students should be able to describe the mean of a data set as a fair share and as a balance point. The mean can only be used for numerical data and is not appropriate for describing categorical data.
EXPLAIN *MEAN ABSOLUTE DEVIATION*	Make known your understanding that the mean absolute deviation is a measure of variation in a set of numerical data which describes, on average, how different the data values are from the mean value.
EXPLAIN *MEASURE OF CENTRAL TENDENCY*	Make known your understanding that measures of central tendency describe how data are clustered around a single value.
EXPLAIN *MEASURE OF VARIABILITY*	Make known your understanding that measures of variability describe how data vary with a single number.
EXPLAIN *MEDIAN*	Make known your understanding that the median is the middle datum when all data is ordered in numerical value in relation to its context. The median can only be used for numerical and is not appropriate for describing categorical data.
EXPLAIN *MODE*	Make known your understanding that the mode is the the data value that appears most in a given set of data. A data value must appear more than once to be the mode, and any set of data can have more than one mode in relation to its context. The mode can be used for numerical data and is also appropriate for describing categorical data.
EXPLAIN *RANGE*	Make known your understanding that the range is the difference between the maximum and the minimum in relation to its context. The range is one way to measure how spread out a set of data is. The range can only be used for numerical and is not appropriate for describing categorical data.
EXPLAIN *STATISTICAL QUESTION*	Make known your understanding that a statistical question anticipates variability in the data.
EXPLAIN VARIABILITY IN A DATA DISTRIBUTION	Make known your understanding whether a set of data is spread out or grouped together.
GATHER REAL-WORLD DATA	Gather real-world data through questions or observations related to a student's immediate environment (e.g., classmates' favorite foods, the number of pets at home, result of flipping a coin, etc.).

IDENTIFY *MODE*	Identify or name the mode for a given set of data by selecting the data value that appears most frequently in the data.
RECOGNIZE *STATISTICAL QUESTION*	Identify or name a question as statistical or not statistical.
RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS CENTER	Identify that a data distribution can be described by the center of the data.
RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SHAPE	Identify that a data distribution can be described by the overall shape of the data.
RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SPREAD	Identify that a data distribution can be described by how spread out the data are.

ADDITIONAL NODES RELATED TO THIS UNIT OF INSTRUCTION

Node Name	Node Description	Related Node
ANALYZE A DATA DISTRIBUTION BY OVERALL SHAPE	Analyze a data distribution based on its overall shape by describing peak(s), outlier(s), symmetry or direction of skew, approximate values of measures of central tendency, etc.	Postrequisite of RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SHAPE

ASKING QUESTIONS AND DESCRIBING DATA

TEACHER NOTES

This unit includes the following documents:

- Learning Map Information
- Instructional Activity (three lessons)
- Instructional Activity Student Handout (for Lessons 2 and 3)
- Instructional Activity Supplement (for Lesson 1)
- Student Activity
- Student Activity Solution Guide

In this unit, students will explore the difference between statistical and nonstatistical questions, ask a statistical question, collect data to describe and analyze, and learn about measures of central tendency (mean, median, and mode) and measures of variability (range and mean absolute deviation).

RESEARCH

In our everyday lives, we encounter a great deal of numerical information presented in various ways. Some of this information appeals to statistical thinking, while other information does not. Sometimes this information is clear and concise, and other times it is misleading. "Statistical literacy is critical to understanding the world around us" (Van de Walle, Bay-Williams, Karp, & Lovin, 2014, p. 325). We must be able to make sense of data to be wise consumers, while making sense of the world requires measuring, observing, experimenting, collecting data, analyzing, and synthesizing (Kader & Mamer, 2008; Ronau & Karp, 2001).

During middle school, students are ready for and should have opportunities to develop their statistical thinking (Groth & Bargagliotti, 2012). Statistics is a priority in middle school in the Common Core State Standards; in sixth grade students explore and describe data, in seventh grade students compare data distributions, and in eighth grade students explore bivariate data, lines, scatter plots, and tables (Van de Walle et al., 2014). When considering information, in the classroom and in our personal lives, it is always important to discuss how the data was collected (Kader & Mamer, 2008).

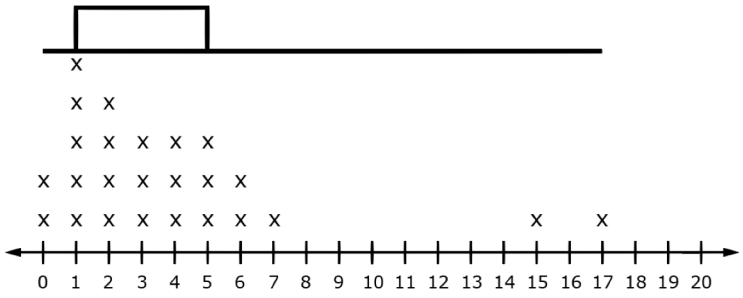
It is also important for students to experience math in context, including instruction in statistics (NCTM, 2000). Scholars (e.g., Van de Walle et al., 2014; Groth & Bargagliotti, 2012) describe statistics is a four-step process that includes formulating questions, collecting data, analyzing data, and interpreting results, and in which students should regularly be involved—from asking the questions to interpreting the results. Students should have opportunities to generate questions that can be answered by statistics, decide on data that is appropriate to answer these questions, and determine methods to collect the data (Van de Walle et al., 2014). Asking a question to conduct a survey helps students recognize variability because respondents answer the question in different ways (Groth & Bargagliotti, 2012). However, it is likely that students will need help writing questions that can be answered with statistics. Using examples and nonexamples of statistical

questions is one way to help students focus on elements of an appropriate statistical question (Van de Walle et al., 2014).

Students should have experience with several data displays by the time they enter the middle grades. Dot plots (or line plots) are introduced in second grade and again in fifth grade (including fractions) and lend themselves to comparison, discussions about shape, and discussions about spread for small data sets (Van de Walle et al., 2014). From dot plots, students should move to histograms and box plots, continuing to answer questions about the center, the spread, and concentration in the data (Kader & Mamer, 2008). The transition from dot plots to box plots could be aided with a "hat plot", which displays individual values, as is the case in dot pots, along with a box (hat) for the middle 50 percent of the data, and a whisker (hat brim) on either side of the box to represent the upper 25 percent and the lower 25 percent of the data (Watson et al., 2008). Hat plots are essentially dot plots with additional lines, indicating the maximum, the minimum, and the middle fifty percent, to begin the transition to box plots. The following is an example of a hat plot and a box plot for the same set of data.

AN EXAMPLE: HAT PLOT

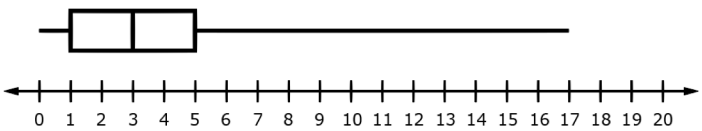
In addition to the traditional dot plot, a "hat" indicating the maximum, the minimum, and the middle fifty percent is drawn.



How Many Pets Do Students Have?

AN EXAMPLE: BOX PLOT

Shown here is a traditional dot plot, displaying the minimum, first quartile, median, third quartile, and maximum.



How Many Pets Do Students Have?

Data can answer questions about the population from which it was gathered, and it is important for students to see data as a whole rather than a collection of numbers in order to reason about the shape of the data (Van de Walle et al., 2014). Seeing a data set as a whole means appreciating the group of numbers together and what they say about a group of individuals. The shape of the data (how the data are spread out or grouped, characteristics of the data set as a whole, and what the data tells us about the population in a global way) is one of the big ideas in the analysis of data and the interpretation of the results (Van de Walle et al., 2014).

In addition to considering the shape of the entire data set, it is important to repeatedly go back and forth between the numerical information and findings and what they mean in terms of the context (Van de Walle et al., 2014). As students are describing data, teachers should ask questions about variability in the data, introduce outliers to the data set and discuss how these values impact the measures of center for the data set, and consider which measure of center makes the most sense for the data set and why (Van de Walle et al., 2014; Foss, 2008). Students don't necessarily have an understanding of advantages and disadvantages associated with mean and median, possibly because they haven't had enough of a chance to make connections among centers and spreads (Zawojewski & Shaughnessy, 2000). Comparing and describing data sets lays the foundation for more advanced statistical reasoning and establishes the need for additional objective measures (e.g., mean absolute deviation) to describe the data (Friel & O'Connor, 1999; Kader, 1999).

During middle school, students should become familiar with measures of center (e.g., mean, median, and mode). While students should continue to view the median as the middle value of the data set, students' conceptions of the mean should evolve during the middle grades. In early grades, students' conceptions of the mean—or the arithmetic average—are as "fair share", but an additional understanding of mean as a "balance point" can be developed in the middle grades (Franklin et al., 2007). The concept of the mean as a balance point is similar to that of the equal sign as a balance point in an equation, which is also emphasized in the middle grades. Interpreting the mean as a balance point helps illustrate why the mean is considered a measure of center (Franklin et al., 2007). Considering distances from the mean and the mean as a balance point leads into the concept of mean absolute deviation as a measure of variability for a data set (Franklin et al., 2007; Kader & Mamer, 2008). As with other measures for a data set, it is important to interpret the mean absolute deviation in the context of the data (Kader, 1999). For middle-grade students, the mean absolute deviation is more natural and more accessible than standard deviation, and it involves similar reasoning (Kader, 1999).

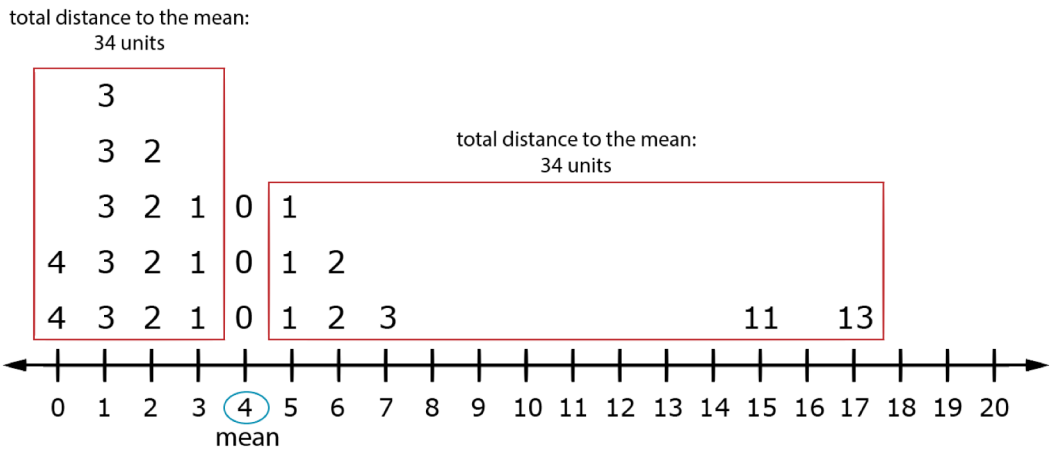
AN EXAMPLE: MEDIAN AS THE MIDDLE VALUE

Ordering the data values from least to greatest and locating the middle value to determine the median helps students understand the median as a measure of center—the median is the central or the middle value.

0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 7, 15, 17

AN EXAMPLE: MEAN AS A BALANCE POINT

Marking the mean on a dot (line) plot, and then determining the sum of the distances to the mean of the data values greater than the mean compared to the sum of the distances to the mean of the data values less than the mean, helps illustrate the mean as a point of balance between the left and right sides of the data. It is important to notice that the mean for this data set is greater than the median, due to the impact outliers have on the mean.



LEARNING MAP INFORMATION

The learning map section for this sequence of activities beings with gathering real-world data, then progresses to asking and answering questions about real-world data and recognizing that data can be described by its shape, its center, and its spread. Students' ability to ask and answer questions about real-world data builds the foundation for explaining and recognizing statistical questions. Understanding that data can be described by its shape and spread allows students first to explain the concept of a measure of variability, then to explain and calculate range and mean absolute deviation. Understanding that data can be described by its center allows students first to explain the concept of a measure of central tendency, then to explain and calculate mean, median, and mode for a data set.

INSTRUCTIONAL ACTIVITIES

The activities in this unit are designed first to establish the difference between a statistical question and a nonstatistical question. Students will consider questions and the possible responses to those questions, identifying statistical questions as those which anticipate variability in the responses. Students will then write their own statistical question and collect data for the question they wrote. In Lesson 2, students consider the

shape and various aspects of a data set displayed in a dot (line) plot as a class, then create a dot (line) plot for the data they collected so they are able to describe the shape of their data. After students have described the shape of their data qualitatively, students will determine measures of center (mean, median, and mode) for the class data set and their own data set, discussing the appropriateness of each measure. Finally, in Lesson 3, students determine measures of variability (range and mean absolute deviation) for the common data set and their own data set, discussing what each measure describes about the data in the context in which it was collected.

REFERENCES

Foss, S. (2008). Teacher to teacher: Literature in the mathematics classroom: Introducing "The Inch Boy" to middle school students. *Mathematics Teaching in the Middle School, 13*(9), 538-542.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). Guidelines for assessment and instruction in statistics education (GAISE) report. *Alexandria: American Statistical Association.*

Friel, S. N., & T O'Connor, W. (1999). Sticks to the roof of your mouth?. *Mathematics Teaching in the Middle School, 4*(6), 404-411.

Groth, R. E., & Bargagliotti, A. E. (2012). GAISEing into the common core of statistics. *Mathematics Teaching in the Middle School, 18*(1), 38-45.

Kader, G. D. (1999). Means and MADS. *Mathematics Teaching in the Middle School, 4*(6), 398-403.

Kader, G., & Mamer, J. (2008). Contemporary curriculum issues: Statistics in the middle grades: Understanding center and spread. *Mathematics Teaching in the Middle School, 14*(1), 38-43.

National Council of Teachers of Mathematics (NCTM). *Principles and Standards for School Mathematics*. Reston, Va.: NCTM, 2000.

Ronau, R. N., & Karp, K. S. (2001). Power over trash: Integrating mathematics, science, and children's literature. *Mathematics Teaching in the Middle School, 7*(1), 26-31.

Van de Walle, J., Bay-Williams, J., Karp, K., & Lovin, L. (2014). Exploring algebraic thinking, expressions, and equations. *Teaching student-centered mathematics: Developmentally appropriate instruction for grades 6-8: Volume III (Second ed.)* (pp. 222-259). Upper Saddle River, NJ: Pearson Education Limited.

Watson, J. M., Fitzallen, N. E., Wilson, K. G., & Creed, J. F. (2008). The representational value of HATS. *Mathematics Teaching in the Middle School, 14*(1), 4-10.

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Mean and median: Are they really so easy?. *Mathematics Teaching in the Middle School, 5*(7), 436-440.

Asking Questions and Describing Data

Overview of Instructional Activities

Lesson	Learning Goal	Nodes Addressed
Lesson 1	Students will establish what it means to ask a statistical question and anticipate the types of responses a particular question would receive.	▶ RECOGNIZE *STATISTICAL QUESTION* ▶ EXPLAIN *STATISTICAL QUESTION* ▶ GATHER REAL-WORLD DATA
Lesson 2	Students will informally and formally discuss data sets, including the meaning and context of the values, particular data values, and the center and the shape of the data. Students will then find quantitative measures of center and consider how well these values describe the data.	▶ ANSWER QUESTIONS ABOUT REAL-WORLD DATA ▶ RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SHAPE ▶ RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS CENTER ▶ EXPLAIN *MEASURE OF CENTRAL TENDENCY* ▶ EXPLAIN *MEAN* ▶ CALCULATE *MEAN* ▶ EXPLAIN *MEDIAN* ▶ CALCULATE *MEDIAN* FOR AN ODD NUMBER OF DATA ▶ EXPLAIN *MODE* ▶ IDENTIFY *MODE*
Lesson 3	Students will first informally discuss how spread out a data set is. Students will then find quantitative measures of variability (range and mean absolute deviation) and consider how these values describe the data.	▶ ANSWER QUESTIONS ABOUT REAL-WORLD DATA ▶ RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SPREAD ▶ EXPLAIN VARIABILITY IN A DATA DISTRIBUTION ▶ EXPLAIN *MEASURE OF VARIABILITY* ▶ EXPLAIN *RANGE* ▶ CALCULATE *RANGE* ▶ CALCULATE *MEAN ABSOLUTE DEVIATION*

ASKING QUESTIONS AND DESCRIBING DATA

INSTRUCTIONAL ACTIVITY

Lesson 1

LEARNING GOAL

Students will establish what it means to ask a statistical question and anticipate the types of responses a particular question would receive.

PRIMARY ACTIVITY

Students will discuss different types of questions and their possible responses in order to establish what it means to ask a statistical question. Students will then consider a variety of questions and classify them based on whether or not they are statistical questions. Students will then practice writing their own statistical question and gathering responses.

OTHER VOCABULARY

Students will need to know the meaning of the following terms:

- Data
- Numeric
- Question
- Response
- Statistical question

MATERIALS

- Scissors
- INSTRUCTIONAL ACTIVITY SUPPLEMENT (Recommend one copy for every two to three students.)

IMPLEMENTATION

Begin the lesson by discussing questions in general.

Ask students to provide examples of questions that could be asked. The type of question is not important, just that students are able to formulate a question.

Record the questions students offer so they can be discussed once a few different type of questions are provided. Try to **elicit** questions that have numeric and non-numeric answers, as well as questions that have a single response and questions that anticipate a variety of responses.

If students do not provide questions that vary, **suggest** a few to include. For example, a question with a single numeric response would be "How many students are in class today?", while a question which anticipates a variety of numeric responses would be "How many pets do you have?". Similarly, a question with a single non-numeric response would be "What color is the door?", while a question which anticipates a variety of non-numeric responses would be "What is your favorite color?".

Once there are approximately ten different questions displayed, **ask** students to consider the response(s) that a person could provide to each question. This could be done as a whole group or the questions could be divided up for small groups to consider.

List the types of responses next to the question and **ask** students to look for patterns in the responses.

Guide students to recognize that some questions are answered with numbers (numeric) while other questions are answered with words (non-numeric). Additionally, some questions have only one answer, while other questions can have many different answers.

Inform students that questions which anticipate or allow for a variety of responses and for which data can be gathered are called *statistical questions*.

Ask students to work in pairs or small groups to determine whether each question displayed is a statistical question or not, and to be prepared to explain their reasoning.

GUIDING QUESTIONS

Elicit student thinking:

- What is a question?
- Why might you ask someone else a question?

Determine if the student can RECOGNIZE *STATISTICAL QUESTION*:

- Is this a statistical question?
- Which of these questions are statistical questions? Which of these questions are not statistical questions?

Determine if the student can EXPLAIN *STATISTICAL QUESTION*:

- Can you explain why this is (or is not) a statistical question?
- What do you look for when deciding whether a question is a statistical question?
- How could this question be answered? How do the possible responses impact whether or not this is a statistical question?

Next, students will work with a new set of questions to decide whether or not they are statistical questions.

Hand out the INSTRUCTIONAL ACTIVITY SUPPLEMENT to pairs or small groups of students.

Require students to cut out each question and sort the questions into a group of statistical questions and a group of non-statistical questions, and to be prepared to explain their thinking.

Ask students guiding questions as they work to scaffold their understanding of statistical questions.

Allow groups of students to compare with each other and discuss any differences they may have and explain their reasoning to each other.

Once groups of students have arrived at decisions regarding which questions are statistical and which questions are not, **discuss** each question as a class.

By the end of the discussion, the class should **establish** that Questions 2, 3, 5, 7, 9, 11, and 12 are statistical questions, while the remaining questions are not.

Remind students that these questions could have different responses and that statistical questions must allow for a variety of answers. The remaining questions will only have one answer and statistics are not required to answer the question.

Now that students have had an opportunity to consider several different questions, students should be required to write their own statistical question which can be answered numerically.

Require students to write a statistical question which they will collect responses for.

Ensure that students write statistical questions with numeric responses in order to use the data in future lessons to discuss center and spread using a dot (line) plot.

Provide students with a number of responses they need to collect. A minimum of 10 responses is recommended, though more may be helpful as students analyze the data.

Students can collect responses to their question by asking other students, family members, friends, or by searching the internet. It may be helpful to provide time in class for students to begin collecting responses to ensure they are on the right track before they leave for the day.

Discuss with students how they plan to collect their data, and make sure they are aware that they must bring a complete set of data for their question to the next class so they can begin analyzing the information they gathered.

GUIDING QUESTIONS

Elicit student thinking:

- How could this question be answered?
- What question would you like to research or ask people? What are you curious about?
- What are you interested in? Can you write a question about that?

Determine if the student is ready to GATHER REAL-WORLD DATA:

- How will you collect responses to this question?

Determine if the student can EXPLAIN *STATISTICAL QUESTION*:

- Can you explain why this is (or is not) a statistical question?
- What do you look for when deciding whether a question is a statistical question?
- How could this question be answered? How do the possible responses impact whether or not this is a statistical question?
- [Point to the question the student wrote.] How could this question be answered? Does that make it statistical or not?

Determine if the student can RECOGNIZE *STATISTICAL QUESTION*:

- [Point to a question in the INSTRUCTIONAL ACTIVITY SUPPLEMENT.] Which of these questions are statistical questions? Which of these questions are not statistical questions?
- [Point to the question the student wrote.] Is this a statistical question?

At the end of the activity, teachers should review the statistical questions students are collecting data for to ensure that they are statistical, appropriately written questions with numeric responses.

Asking Questions and Describing Data

INSTRUCTIONAL ACTIVITY SUPPLEMENT

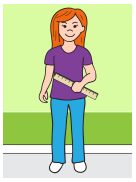
Lesson 1

1. How many pencils do I have?

2. On average, how many pencils do the students in this class have?

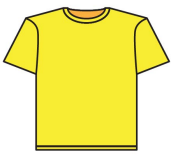
3. What are the favorite fruits among the students in this class?

4. How tall are you?

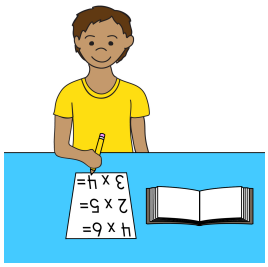


5. How tall are the students in this school?

6. What color is my shirt today?

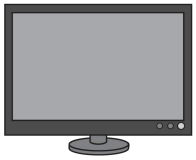


7. How many hours do high school students typically spend doing homework in a week?

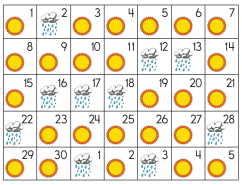


8. How many different species of birds are there?

9. How many hours do middle school students typically spend watching TV in a day?



10. How many days are there in December?



11. How much money do math professors make in a year?

12. How fast do dogs run?

ASKING QUESTIONS AND DESCRIBING DATA

INSTRUCTIONAL ACTIVITY

Lesson 2

LEARNING GOAL

Students will informally and formally discuss data sets, including the meaning and context of the values—particular data values—and the center and the shape of the data. Students will then find quantitative measures of center and consider how well these values describe the data.

PRIMARY ACTIVITY

Students will first consider a data set provided for the whole class. Through a discussion regarding the features of the dot (line) plot, students will come to the understanding that the shape of the data and the measures of center can be described subjectively through words or objectively through quantitative measures. In this lesson, students will build on their understanding of mean as a "fair share" and develop the need for additional measures of center (median and mode). Students will then consider the data set for the question they investigated in a similar fashion.

OTHER VOCABULARY

Students will need to know the meaning of the following terms:

- Center
- Cluster
- Data
- Dot (line) plot
- Mean
- Median
- Mode
- Outlier
- Peak
- Skewed left
- Skewed right
- Symmetric
- Uniform

MATERIALS

- 100 counters or blocks
- Data collected by students at the end of LESSON 1
- Poster paper or several pieces of taped together blank paper
- Ruler
- Sticky notes
- INSTRUCTIONAL ACTIVITY STUDENT HANDOUT

IMPLEMENTATION

> NOTE: This lesson introduces measures of center (i.e., mean, median, and mode). Determining measures of center, particularly mean and median, is included in another sixth grade standard. Therefore, the specific and technical aspects of calculating measures of center are not emphasized in this lesson, but could be included as desired.

Begin the lesson by discussing a hypothetical data set as a class.

Hand out the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT so students can begin considering the dot (line) plot. Students should have some experience with this type of plot from previous grades. Figure 1 shows the dot plot provided in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

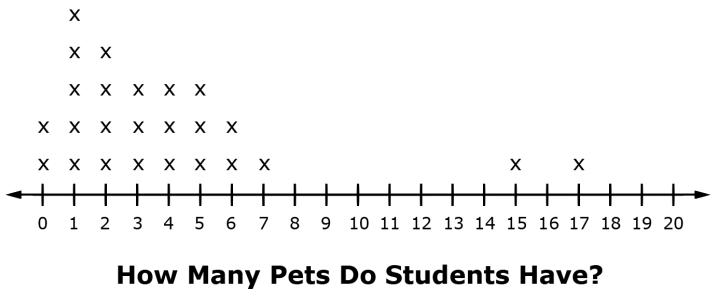


FIGURE 1

Inform students that this information does not reflect the students in their class, but rather it is a hypothetical distribution of a group of students' pets.

Briefly discuss the source of the data and the impact this has on how that data is viewed.

Ask students to consider, briefly, the possible sources of the hypothetical data. For example, the data could have been collected from every student in one class, some of the students in a school, or a random sample (selection) of students across the country.

Discuss that knowing how the data was collected impacts how it is perceived and that this is always something important to consider when students are presented with data of any kind. For example, if this data was collected from the students in their class, we cannot assume that different students in the school, the state, another state, or another country would have the same data. There are several factors that can impact this type of information (e.g., students in urban areas may have fewer pets than students in rural areas). Students will learn more about random samples in seventh grade.

Next, students will consider specific features of the dot (line) plot.

Ask students to write the information they know based on the data display in Question 1 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT. **Allow** students to describe any information they can gather from the display.

Observe what students write, looking for accurate and unique observations students make, then **ask** some students to share their ideas with the rest of the class.

Confirm or **correct** students' observations as they share, discussing them as a class to establish agreement or disagreement.

Include questions during the discussion about how many students are represented in the display, whether there are peaks in the data, whether the data set is clustered around any particular values, where the center/middle of the data appears to be, and whether any points seem to be particularly far away from the rest of the data (outliers).

Note that some descriptions of data can be specific (e.g., the number of data values), while others tend to be more subjective (e.g., clusters in data).

GUIDING QUESTIONS

Elicit student thinking:

- What do you see when you look at this dot (line) plot?
- What information does this display provide?

Determine if the student can ANSWER QUESTIONS ABOUT REAL-WORLD DATA:

- How many students have six pets? How many students have nine pets?
- How many pets are the most common for a student to have in this data set?
- What is the greatest number of pets represented in this data set? What is the least?
- How many pets do the majority of students have based on this display?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SHAPE:

- How would you describe the shape of the data?
- Are there peaks in the data? Are there outliers?
- Are the data clustered in a particular portion of the dot (line) plot?

Next, students will recreate the dot (line) plot using sticky notes on a piece of poster paper in order to consider the different shapes data distributions can have.

Require individual students or groups of students to recreate the dot (line) plot using poster paper (or several pieces of blank paper taped together) and sticky notes for each data value. Students will need to draw and label the number line before placing the sticky notes on the dot (line) plot.

Emphasize that the intervals along the number line must be regular and consistent.

Inform students that because most of the data is grouped together with the exception of a few values to the right, this data distribution is considered to be *skewed (right)*.

Ask students to rearrange the sticky notes to create a different data distribution that is also skewed right.

Observe students' work and ask some students or groups to share their data distribution with the class. **Discuss** how they know the data set is skewed right. Figure 2 shows a data set that is skewed right.

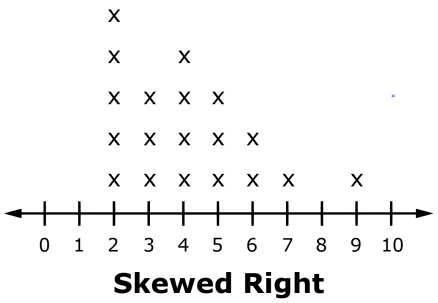


FIGURE 2

Next, **ask** students to create a data distribution that they believe is *skewed left*.

Observe students' work and ask some students or groups who have correctly created a data distribution that is skewed left to share their work with the class. **Discuss** how they know the data set is skewed left. Figure 3 shows a data set that is skewed left.

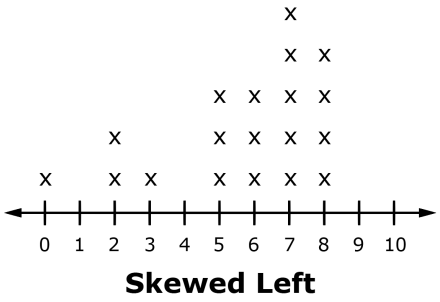


FIGURE 3

Ask students, based on the conversations to this point, what they believe is the definition of *skewed* when it is used in statistics to refer to a data distribution.

Establish, through the discussion, that a *skewed* data set is a way to describe data sets that are *asymmetric* or not symmetric. Either the right or the left side of the distribution can cause asymmetry, which is why data that is skewed can generally be described as skewed right or skewed left.

Discuss, as needed, what it means for a figure or an image to be symmetric or to have a line of symmetry. (Students should be familiar with symmetry from fourth grade.)

Next, **ask** students to create a data distribution that they believe is *symmetric*.

Observe students' work and ask some students or groups who have correctly created a symmetric data distribution to share their work with the class. **Discuss** how they know the data set is symmetric. Figure 4 shows a data set that is symmetric.

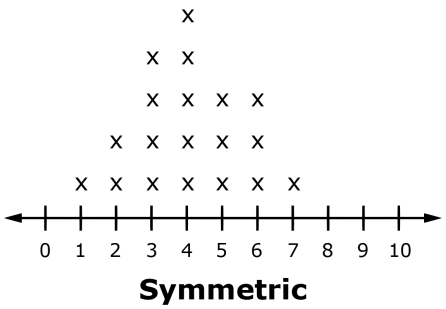


FIGURE 4

Ask students, based on the conversations to this point, what they believe the definition of *symmetric* is when it is used in statistics to refer to a data distribution.

Establish, through the discussion, that a *symmetric* data set has a peak that is approximately in the center of the data, and the shape of the data to the right and the left of the center is approximately the same.

Discuss with students the meaning of the word *uniform*. If students struggle to provide a definition for uniform, help scaffold their understanding of the definition of the word. One approach would be to discuss uniforms for a marching band or an athletic team, noting that uniforms are essentially the same for every person, with allowance for a few differences such as size and number.

Next, **ask** students to create a data distribution that they believe is *uniform*.

Observe students' work and ask some students or groups who have correctly created a uniform data distribution to share their work with the class. **Discuss** how they know the data set is uniform. Figure 5 shows a data set that is uniform.

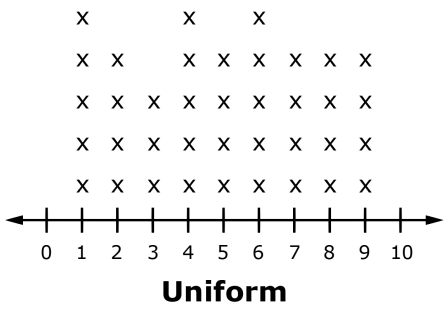


FIGURE 5

Ask students, based on the conversations to this point, what they believe the definition of *uniform* is when it is used in statistics to refer to a data distribution.

Establish, through the discussion, that a *uniform* data set has values that are equally spread across the outcomes.

Now that students have explored dot plots for the hypothetical set of data, students will begin considering the data they have collected.

Require students to use the data they collected to create a dot plot and answer Question 2 in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

Emphasize that the intervals along the number line must be regular and consistent.

GUIDING QUESTIONS

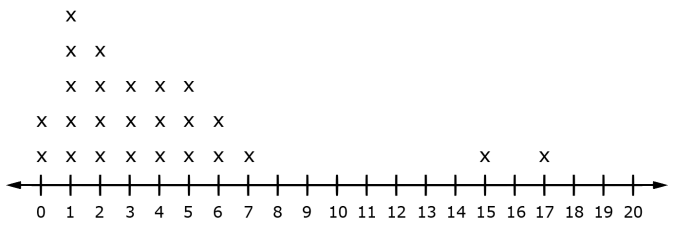
Elicit student thinking:

- What do you notice about the data?
- What information is readily available in this display?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SHAPE:

- How would you describe the shape of the data?
- Are there peaks in the data? Are there outliers?
- Are the data clustered in a particular portion of the dot (line) plot?
- Is this data skewed left, skewed right, symmetric, or uniform? How do you know?
- What does the shape of the data tell you about the information you collected?

Next, students will focus on the center of the data distribution about the number of pets students have.



How Many Pets Do Students Have?

Ask students where they believe the center of the data would be. Because this is a vague question, student responses may vary. Regardless of the response, **require** students to explain their reasoning.

Discuss with students that because the center is difficult to identify, there are measures of central tendency that describe how data are clustered using a single value.

Inform students that one measure of central tendency, the *mean* (or the arithmetic average), can be found by figuring out how many pets each person would have if everyone had the same number of pets. Note that this is a "fair share" perception of the mean.

Ask students to determine or recall how many students are represented in the dot (line) plot. Students should identify that there are 25 students represented in the dot (line) plot.

If there are 25 students in the class, let each student represent one of the students in the dot (line) plot. If there are more than 25 students in the class, some students may need to work in a pair. If there are fewer than 25 students, some students may need to represent two students from the dot (line) plot.

Distribute counters or blocks to students such that each block represents one pet. For example, two students will not get any blocks, five students will get one block each, four students will get two blocks each, etc., so the students in the class represent the students in the hypothetical data display.

Discuss how the number of counters or blocks varies by student and **ask** how many counters or blocks (pets) each student would have if everyone had the same number.

Encourage students to model a redistribution of the counters or blocks by collecting all the counters or blocks from all the students and redistributing them one by one such that each student takes one, then a second, then a third, until all the counters or blocks have been taken.

Ask students how many counters or blocks (pets) they have now. Each student should have four counters or blocks (pets) at this point.

Repeat that this is one measure of central tendency that can be found for a data set and it is referred to as the mean. The mean for this data set is 4, a single number that describes how the data about students and their pets are clustered.

Ask students if they feel the mean accurately represents the center of the data. While it is close to the center of the data, some students hopefully will argue that it seems a little high for the data. If a student does not make this statement, **ask** students if anyone believes 4 is too high or too low to represent the data.

Ask students why the mean could appear to be a little higher than the center of the data.

Scaffold questions to lead students to the realization that the two students who have a large number of pets cause the mean to be higher than it would be without those values.

Note that the mean is one objective measure of central tendency, which combines (adds) each of the values and divides the sum by the number of values there were. However, the mean may not be the best measure to use with every data set.

Require students to practice determining the mean for the data they collected at the end of LESSON 1 on Question 3 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

GUIDING QUESTIONS

Elicit student thinking:

- Where do you think the center of the data is?
- Can you be sure you are identifying the exact center of the data by looking at the dot (line) plot? Why or why not?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS CENTER:

- If we report the mean for a data set, what are we trying to describe about the data?
- In addition to the shape of the data, how else could we describe a data set?

Determine if the student can EXPLAIN *MEASURE OF CENTRAL TENDENCY*:

- What are measures of central tendency?
- What do the measures of central tendency tell you about the data?

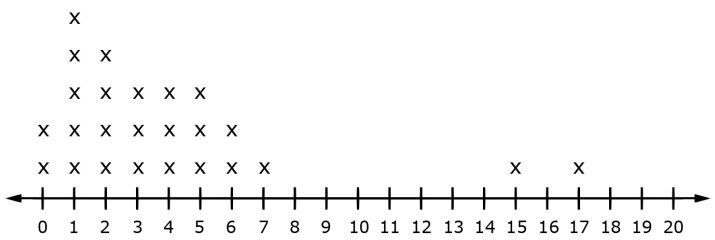
Determine if the student can EXPLAIN *MEAN*:

- What does the mean tell you about the data?
- Could you change some data values and still have the same mean? How?
- Mean is sometimes described as "fair share". Can you explain how that phrase makes sense to describe the mean?
- Can you determine the mean for a set of numerical data? Why or why not?
- Can you determine the mean for a set of categorical data? Why or why not?
- Would the mean be impacted by an extremely large or extremely small value? Explain your reasoning.

Determine if the student can CALCULATE *MEAN*:

- What is the mean for your data set?
- How did you determine the mean for your data?

Now that students have considered the mean for a data distribution, they will be introduced to another measure of central tendency: the median. Students will first work with the data distribution about the number of pets students have, then practice with the data sets individual students have collected.



How Many Pets Do Students Have?

Inform students that there are additional measures of center that can be found for a data distribution.

Explain that another measure of central tendency in the data, the *median*, is also called the middle of the data, or the value that falls in the middle, when all the data values are arranged from least to greatest.

Ask students to order the values in the dot plot from least to greatest and locate the value in the middle of the list. Students should identify 3 as the middle value or the median. **Note** that 12 values are less than (or equal to) 3, and 12 values are greater than (or equal to) 3.

Ask students if they feel the median accurately represents the center of the data. Answers will vary, but students should identify that the median does appear to be close to the center of the data.

Again, **note** that the median is one objective measure of central tendency, which identifies the middle value of the data set when the values are ordered from least to greatest. However, the median may not be the best measure to use with every data set.

Require students to practice determining the median for the data they collected at the end of LESSON 1 on Question 4 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT. **Note** that if students have collected an even amount of values, they will have to find the mean of the middle two values in order to obtain the median.

Now that students are familiar with two measures of central tendency, **encourage** them to evaluate which measure of center better represents the center of their data set and to respond to Question 5 in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

GUIDING QUESTIONS

Elicit student thinking:

- Where do you think the center of the data is?
- Can you be sure you are identifying the exact center of the data by looking at the dot (line) plot? Why or why not?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS CENTER:

- If we report the mean or the median for a data set, what are we trying to describe about the data?
- In addition to the shape of the data, how else could we describe a data set?

Determine if the student can EXPLAIN *MEASURE OF CENTRAL TENDENCY*:

- What are measures of central tendency?
- What do the measures of central tendency tell you about the data?

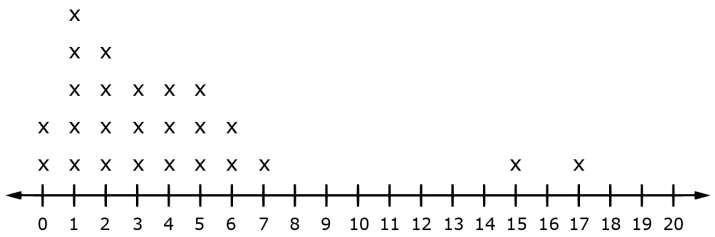
Determine if the student can EXPLAIN *MEDIAN*:

- What does the median tell you about the data?
- Could you change some data values and still have the same median? How?
- The median is sometimes described as the middle of the data. Can you explain how that phrase makes sense to describe the median?
- Can you determine the median for a set of numerical data? Why or why not?
- Can you determine the median for a set of categorical data? Why or why not?
- Would the median be impacted by an extremely large or extremely small value? Explain your reasoning.

Determine if the student can CALCULATE *MEDIAN* FOR AN ODD NUMBER OF DATA:

- What is the median for your data set?
- How did you determine the median for your data?

Students will next consider the final measure of central tendency, the mode, for a data distribution. Students will first work with the data distribution about the number of pets students have, then practice with the data sets they have collected.



How Many Pets Do Students Have?

Explain that the final measure of central tendency for a data set, the *mode*, is also referred to as the value (or values) that occurs most often.

Ask students to identify the mode for the set of data describing students and the number of pets. Students should identify the value that occurs most often is one pet.

Ask students if they feel the mode accurately represents the center of the data. Answers will vary, but students should identify that the mode does not appear to be close to the center of the data, but rather it seems to be a little below the center of the data.

Again, **note** that the mode is one objective measure of central tendency, which identifies the most frequent value of the data set. However, the mode may not be the best measure to use with every data set.

Require students to practice determining the mode for the data they collected at the end of LESSON 1 on Question 6 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT. **Note** that if students have collected data where two or more values occur the same number of times and are the most frequent, then each of these values is considered to be a mode for the data set.

Remind students that they have learned about mean, median, and mode as measures of central tendency for a data set, or measures that describe how data are clustered around a single value.

GUIDING QUESTIONS

Elicit student thinking:

- Where do you think the center of the data is?
- Can you be sure you are identifying the exact center of the data by looking at the dot (line) plot? Why or why not?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS CENTER:

- If we report the mean, median, or mode for a data set, what are we trying to describe about the data?
- In addition to the shape of the data, how else could we describe a data set?

Determine if the student can EXPLAIN *MEASURE OF CENTRAL TENDENCY*:

- What are measures of central tendency?
- What do the measures of central tendency tell you about the data?

Determine if the student can EXPLAIN *MODE*:

- What does the mode tell you about the data?
- Could you change some data values and still have the same mode? How?
- The mode is sometimes described as the most frequent value. Can you explain what that means in terms of the data?
- Can you determine the mode for a set of numerical data? Why or why not?
- Can you determine the mode for a set of categorical data? Why or why not?
- Would the mode be impacted by an extremely large or extremely small value? Explain your reasoning.

Determine if the student can IDENTIFY *MODE*:

- What is the mode for your data set?
- How did you determine the mode for your data?

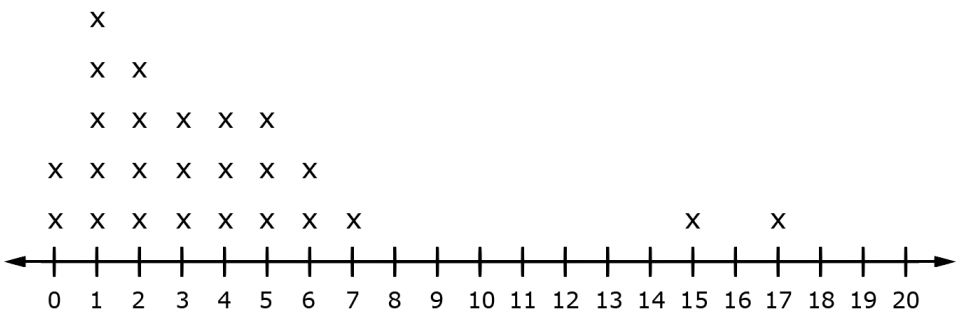
Students should be required to complete the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT thoroughly to provide information describing their understanding of how to describe data by its shape and by measures of central tendency (i.e., mean, median, and mode).

At the end of the activity, provide students with an additional data set and require students to create a dot (line) plot for the data, describe the shape of the data (i.e. skewed left, skewed right, symmetric, or uniform), and describe the center of the data (using the measures of central tendency from this lesson).

Name_____

Asking Questions and Describing Data

Lesson 2



How Many Pets Do Students Have?

1. What information do you know based on this data display?

2. Use the number line to create and label a dot (line) plot for your data, then answer the following questions.

- Does your data have any peaks? If so, where do they occur?

- Where is your data clustered? Why do you think this is the case?

- How would you describe the overall shape of your data? (i.e., skewed left, skewed right, symmetric, or uniform) Explain your reasoning.

3. What is the mean for the data you collected? Provide the mathematical reasoning you used to determine this value.

4. What is the median for the data you collected? Provide the mathematical reasoning you used to determine this value.

5. Does the mean or the median better represent your data set? Explain your reasoning.

6. What is the mode for the data you collected? Provide the mathematical reasoning you used to determine this value.

ASKING QUESTIONS AND DESCRIBING DATA

INSTRUCTIONAL ACTIVITY

Lesson 3

LEARNING GOAL

Students will first informally discuss how spread out a data set is. Students will then find quantitative measures of variability (range and mean absolute deviation) and consider how these values describe the data.

PRIMARY ACTIVITY

Students will first consider a data set provided for the whole class to come to an understanding that how spread out data are can be described subjectively through words or objectively through quantitative measures. Students will first determine the range to describe the variability of the data, then, conceptualizing mean as a balance point, determine the mean absolute deviation and describe this value in terms of the data. Students will then consider the data set for the question they investigated in a similar fashion.

OTHER VOCABULARY

Students will need to know the meaning of the following terms:

- Data
- Dot (line) plot
- Mean absolute deviation
- Measure of variability
- Range
- Spread
- Variability

MATERIALS

- Data collected by students at the end of LESSON 1
- INSTRUCTIONAL ACTIVITY STUDENT HANDOUT

IMPLEMENTATION

> NOTE: This lesson introduces measures of variability (i.e. range and mean absolute deviation). Determining measures of variability, particularly mean absolute deviation, is included in another sixth grade standard. Therefore, mastery of mean absolute deviation is not emphasized in this lesson but could be included as desired.

Begin the lesson by discussing as a class a hypothetical data set, the same data set that was considered in LESSON 2.

Hand out the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT so students can begin considering the dot (line) plot. Figure 1 shows the dot (line) plot provided in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

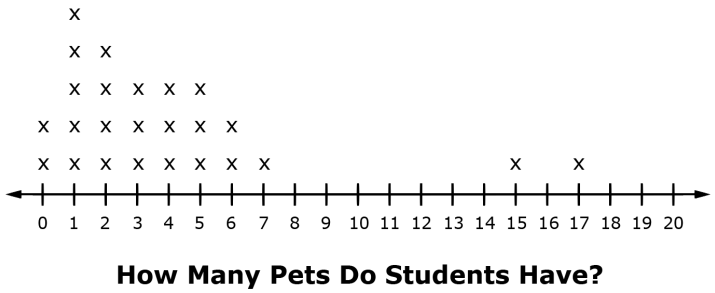


FIGURE 1

Remind students that this information does not reflect the students in their class, but rather it is a hypothetical distribution of a group of students' pets.

Next, students will consider—informally—how spread out the data set is.

Ask students to think about the meaning of the phrase "spread out". If necessary, provide an example such as "In physical education class, the teacher may ask students to spread out before doing jumping jacks."

Define the *spread* of a data distribution as how similar or varied the values in the data distribution are.

Ask students to describe how spread out the data set is and what the spread indicates about students and the number of pets they have on Questions 1 and 2 in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

Observe what students write, looking for accurate and unique observations students make, then **ask** some students to share their ideas with the rest of the class.

Confirm or **correct** students' observations as they share, discussing them as a class to establish agreement or disagreement.

Note that students' descriptions of the spread of the data can be specific (e.g., the highest and the lowest data values), while others tend to be more subjective (e.g., describing how spread out the data are).

GUIDING QUESTIONS

Elicit student thinking:

- What do you see when you look at this dot (line) plot?
- What information does this display provide?

Determine if the student can ANSWER QUESTIONS ABOUT REAL-WORLD DATA:

- What is the greatest number of pets represented in this data set? What is the least?
- Are there values where the data seem to be grouped? Can you describe these values?
- How many values seem to fall "outside" the rest of the data? What are these values?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SPREAD:

- How would you describe the spread of the data?
- Does the data vary significantly? How do you know?
- What does the spread of the data tell you about the information in the data display?

Now that students have explored the spread of the hypothetical data set, students will consider the data they collected at the end of LESSON 1 and worked with in LESSON 2.

Require students to create a dot (line) plot and answer the questions for the data they collected (at the end of LESSON 1) on Question 3 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

Emphasize that the intervals along the number line must be regular and consistent.

Observe student work and ask guiding questions as students work to ensure accuracy and promote understanding.

GUIDING QUESTIONS

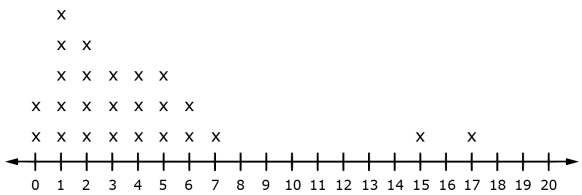
Elicit student thinking:

- What do you notice about the data?
- What information is readily available in this display?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SPREAD:

- How would you describe the spread of your data?
- Does your data vary significantly? How do you know?
- What does the spread of the data tell you about the information you collected?

Next, students will learn specifically about one measure of variability that describes the spread of the data: the range.



How Many Pets Do Students Have?

Redirect students' attention to the dot (line) plot describing students and their pets.

Ask how far apart the lowest and highest values are. Students should see that because the lowest value is zero and the highest value is 17, the lowest and highest values are 17 units apart. **Require** students to explain their reasoning.

Ask students what this difference means in the context of the data. Students should indicate this means the person with the most pets has 17 more pets than the person with the least pets, or an equivalent statement.

Inform students that the distance between the highest and lowest data values, also known as the *range*, is one *measure of variability* in a data set because it describes how the data vary with a single number.

Require students to practice determining and interpreting the range for the data they collected at the end of LESSON 1 on Questions 4 and 5 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

GUIDING QUESTIONS

Elicit student thinking:

- How spread out are the data?
- What words might you use to describe how much the data vary?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SPREAD:

- If we report the range for a data set, what are we trying to describe about the data?
- In addition to the shape and center of the data, how else could we describe a data set?

Determine if the student can EXPLAIN VARIABILITY IN A DATA DISTRIBUTION:

- How much does this data vary? Explain your reasoning.
- How spread out are the values in this data set? Explain your reasoning.
- Would you describe the data as being spread out? Why or why not?
- Would you describe the data as being grouped together? Why or why not?

Determine if the student can EXPLAIN *MEASURE OF VARIABILITY*:

- What is a measure of variability?
- What does a measure of variability tell you about a data set?
- Can you give an example of a measure of variability? Can you explain how you know it is a measure of variability?

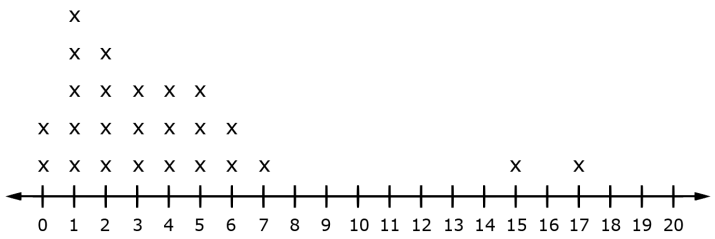
Determine if the student can EXPLAIN *RANGE*:

- What does the range tell you about the data?
- What values are important to determine the range of a data set?
- Can you determine the range for a set of numerical data? Why or why not?
- Can you determine the range for a set of categorical data? Why or why not?

Determine if the student can CALCULATE *RANGE*:

- What is the range for your data set?
- How did you determine the range for your data?

Now that students have considered the range for a data distribution, they will be introduced to a measure of variability that is based on the mean: the mean absolute deviation. Students will work with the data distribution about the number of pets students have.



How Many Pets Do Students Have?

Inform students that there is a measure of variability that is based on the mean. Just as there are different measures of central tendency, there are also different measures of variability (measures that describe how data vary with a single number).

Ask students to recall or recalculate the mean of the data set about students and their pets. Students should arrive at or recall that the mean for the data set is four.

Explain that one measure of variability that is based on the mean, the *mean absolute deviation*, measures how far the data values are, on average, from the mean.

On Question 6 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT, **require** students to replace each mark in the original dot (line) plot with a number describing how far that value is from

the mean. It may be necessary to begin as a class, then allow students to finish individually. Figure 2 models the result of this process.

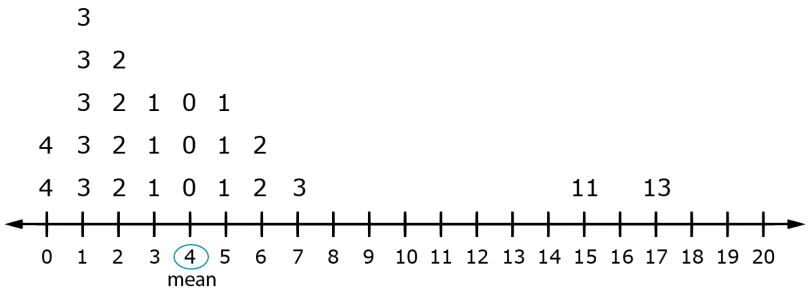


FIGURE 2

Once students have created this model, **require** that they answer Questions 7 and 8 in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT regarding the sum of the distances for the data values that are less than the mean and the sum of the distances for the data values that are greater than the mean.

Once students have completed Questions 7 and 8 in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT, **ask** students what they notice about their responses to each question. Students should notice that the sum of the distances (34) is the same for the values greater than the mean as it is for the values less than the mean.

This discovery can help students begin to interpret the *mean as a balance point* in addition to the fair share conception of the mean they may already hold.

Discuss how the values greater than the mean "balance out" the values less than the mean.

> NOTE: Modeling how data values "balance" works best with a whole number mean. If you wish to provide students with an additional example or many examples to emphasize that the values always balance, ensure the mean is a whole number for the provided data sets. This lesson does not require students to determine the mean absolute deviation for the data they collected, because it is unlikely that students have a whole number mean for their data.

Ask students to answer Question 9 in the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT.

Ensure students have accurately determined the sum of all distances from the mean to be 68 (including the values that are zero units away from the mean).

To determine the mean absolute deviation on Question 10 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT, students should divide the sum of all distances from the mean (68) by the total number of data values (25).

Require students to complete Question 10 of the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT and show their work to determine the mean absolute deviation, or the average distance between each data value and the mean.

Discuss what a mean absolute deviation of 2.72 means in terms of the data. **Emphasize** that, on average, the number of pets students (in this display) have varies by almost three pets from the mean of four pets.

If a student in the class has collected data with a whole number mean (or if you wish to provide students with an additional data set with a whole number mean), students could determine and explain the mean absolute deviation for the data set as additional practice.

Now that students are familiar with two measures of variability, **encourage** them to consider which measure of variability is more useful when considering data sets in the future.

GUIDING QUESTIONS

Elicit student thinking:

- How spread out are the data?
- What words might you use to describe how much the data vary?

Determine if the student can RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SPREAD:

- If we report the range or the mean absolute deviation for a data set, what are we trying to describe about the data?
- In addition to the shape and center of the data, how else could we describe a data set?

Determine if the student can EXPLAIN VARIABILITY IN A DATA DISTRIBUTION:

- How much does this data vary? Explain your reasoning.
- How spread out are the values in this data set? Explain your reasoning.
- Would you describe the data as being spread out? Why or why not?
- Would you describe the data as being grouped together? Why or why not?

Determine if the student can EXPLAIN *MEASURE OF VARIABILITY*:

- What is a measure of variability?
- What does a measure of variability tell you about a data set?
- Can you give an example of a measure of variability? Can you explain how you know it is a measure of variability?

Determine if the student can EXPLAIN *MEAN ABSOLUTE DEVIATION*:

- What does the mean absolute deviation tell you about the data?
- What values are important to determine the mean absolute deviation of a data set?
- Can you determine the mean absolute deviation for a set of numerical data? Why or why not?
- Can you determine the mean absolute deviation for a set of categorical data? Why or why not?

Determine if the student can CALCULATE *MEAN ABSOLUTE DEVIATION*:

- [Point to a value on the dot plot.] How far is this value from the mean?
- What is the mean absolute deviation for the data set?
- How did you determine the mean absolute deviation for the data?

Students should be required to thoroughly complete the INSTRUCTIONAL ACTIVITY STUDENT HANDOUT to provide information describing their understanding of how to describe data by its spread and their ability to determine measures of variability.

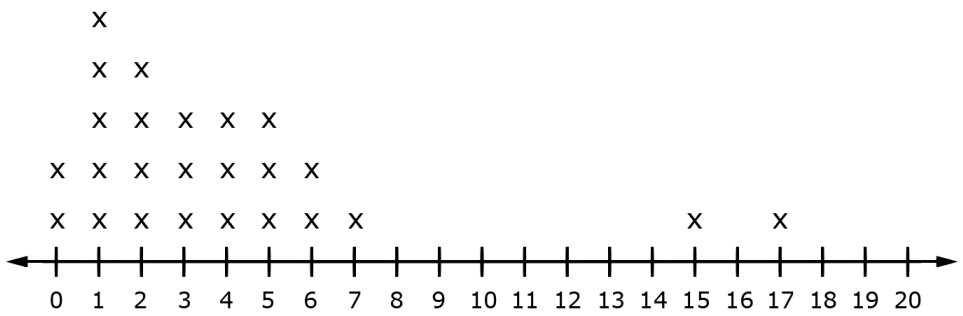
At the end of the activity, provide students with an additional data set and require students to create a dot (line) plot for the data, informally describe the spread of the data, determine the range, and determine the mean absolute deviation. Following is an example data set (with a whole number mean).

Students' scores on the first quiz: 1, 3, 6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10

Name_____

Asking Questions and Describing Data

Lesson 3



How Many Pets Do Students Have?

1. How spread out are the data in this data display?

2. What does the spread in the data tell you about students and their pets?

3. Use the number line to create and label a dot (line) plot for your data, then answer the following questions.

Name_____

- How spread out are the data in your data display?

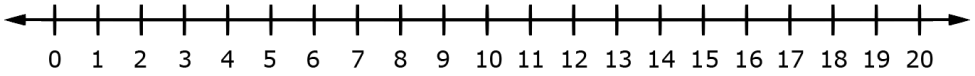
- What does the spread in the data tell you about the question you asked?

4. What is the range for the data you collected? Provide the mathematical reasoning you used to determine this value.

5. Describe the range in terms of the data you collected.

Name_____

6. Use the following number line to label the mean for the set of data about students and their pets, then replace the marks (Xs) on the dot (line) plot with a number indicating how far that value is from the mean.



How Many Pets Do Students Have?

7. What is the sum of the distances (from the mean) of the data values that are less than the mean?

8. What is the sum of the distances (from the mean) of the data values that are greater than the mean?

9. What is the sum of the distances (from the mean) of all data values?

10. The *mean absolute deviation* is a measure of spread determined by dividing the sum of the distances from the mean by the total number of data values. What is the mean absolute deviation for this data? Show your work.

11. Describe the mean absolute deviation in terms of the data you collected.

Name_____

Asking Questions and Describing Data

Lessons 1 – 3

1. Consider each of the following questions. Circle each statistical question and draw a line through each question that is not statistical.

What are the favorite colors among the students in this class?

When is your birthday?

How many states are there in the United States?

How old are the students in this school?

How many hours do middle school students typically spend doing homework in a day?

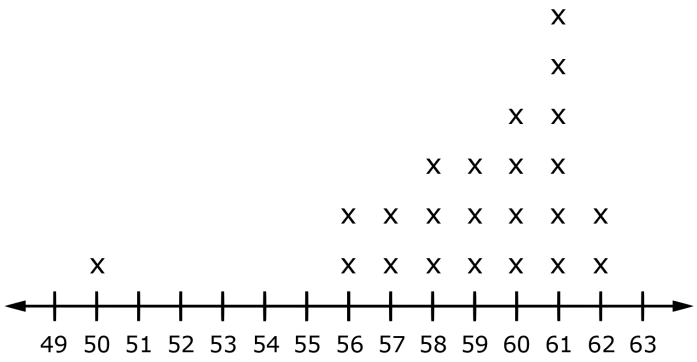
2. Write your own statistical question, then explain how you know the question you wrote is statistical.

3. What does a measure of central tendency describe about a data set?

Name_____

4. What does a measure of variability describe about a data set?

5. Use the data set provided in the dot plot to answer the following questions.



Car Speeds on a Highway (mph)

5.a. Describe the shape and spread of the data. In your description, be sure to include peaks, clusters, outliers, how spread out the data are, and whether the distribution is skewed right, skewed left, symmetric, or uniform.

5.b. Determine the mean in the context of the data. Provide appropriate work to support your answer.

5.c. Determine the median of the data. How well does the median describe the center of the data?

5.d. Determine the mode of the data. How well does the mode describe the center of the data?

5.e. Determine the range of the data. What does this number describe about the data?

5.f. Determine the mean absolute deviation of the data. Provide appropriate work to support your answer. What does this number describe about the data?

6. Which measure(s) of central tendency (i.e., mean, median, and mode) and measures of variability (i.e., range and mean absolute deviation) from this unit can be determined for categorical data? Explain your reasoning.

Name_____

7. One way of thinking about the mean of a data set is as a "fair share". Explain how this description of the mean is related to the process used to calculate the mean.

8. Another way of thinking about the mean of a data set is as a "balance point". Explain how this description of the mean is modeled in a dot plot.

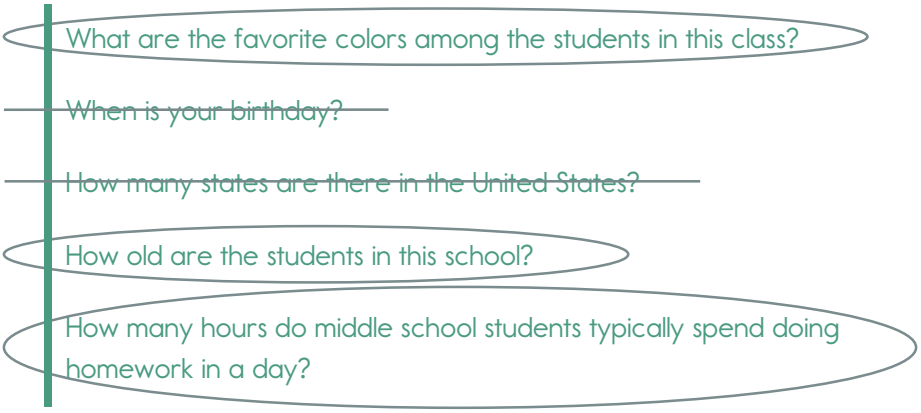
ASKING QUESTIONS AND DESCRIBING DATA

STUDENT ACTIVITY SOLUTION GUIDE

Lessons 1 – 3

1. Consider each of the following questions. Circle each statistical question and draw a line through each question that is not statistical.

CORRECT ANSWER



ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
Student circles only the third, fourth, and fifth questions.	believes statistical questions are defined as questions with numerical responses	RECOGNIZE *STATISTICAL QUESTION* and EXPLAIN *STATISTICAL QUESTION*
Student circles all five questions.	does not know the difference between a statistical question and a question in general	RECOGNIZE *STATISTICAL QUESTION* and EXPLAIN *STATISTICAL QUESTION*

2. Write your own statistical question, then explain how you know the question you wrote is statistical.

CORRECT ANSWER

[*Student responses will vary. Verify that students write a question that anticipates variability in the data.*]

I know this is a statistical question because there are many responses that could be provided as answers to this question.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
Student writes a question with only one possible, numeric response.	believes statistical questions are defined as questions with numerical responses; does not realize statistical questions must anticipate variability in data; may not realize statistical questions can have nonnumeric responses	EXPLAIN *STATISTICAL QUESTION*
Student writes a general question that does not anticipate variability in the data (e.g., "Why is the sky blue?").	does not know the difference between a statistical question and a question in general	EXPLAIN *STATISTICAL QUESTION*

3. What does a measure of central tendency describe about a data set?

CORRECT ANSWER

A measure of central tendency describes how data are clustered around a single value.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
A measure of central tendency tells you the center.	uses context clues to infer that the center is related to a measure of central tendency, but cannot provide more specific information related to the data	EXPLAIN *MEASURE OF CENTRAL TENDENCY*
Student does not respond to the question.	is not familiar enough with this phrase to explain what it means	EXPLAIN *MEASURE OF CENTRAL TENDENCY*

4. What does a measure of variability describe about a data set?

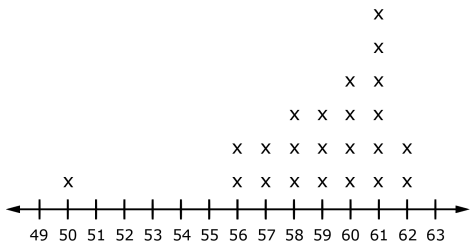
CORRECT ANSWER

A measure of variability describes how data vary with a single number.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
A measure of variability tells you the variation.	uses context clues to infer that variation is related to a measure of variability, but cannot provide more specific information related to the data	EXPLAIN *MEASURE OF VARIABILITY*
Student does not respond to the question.	is not familiar enough with this phrase to explain what it means	EXPLAIN *MEASURE OF VARIABILITY*

5. Use the data set provided in the dot plot to answer the following questions.



Car Speeds on a Highway (mph)

5.a. Describe the shape and spread of the data. In your description, be sure to include peaks, clusters, outliers, how spread out the data are, and whether the distribution is skewed right, skewed left, symmetric, or uniform.

CORRECT ANSWER

These data have a peak at 61 miles per hour and are clustered between 56 and 62 miles per hour. There seems to be an outlier of 50 miles per hour. Most of the data are grouped together. The data set appears to be skewed left.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
Student does not discuss the peaks, clusters, outliers, or whether the distribution is skewed, symmetric, or uniform.	does not know how to describe a data distribution by its shape	RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SHAPE
Student does not discuss the spread of the data in the dot plot.	does not know that a data distribution can be described by its spread or does not know how to describe a data distribution by its spread	RECOGNIZE THAT A DATA DISTRIBUTION CAN BE DESCRIBED BY ITS SPREAD
The data are spread out.	may be focusing on the value at 50 miles per hour without considering that the rest (the majority) of the data are very grouped together	EXPLAIN VARIABILITY IN A DATA DISTRIBUTION
The data set appears to be skewed right.	identifies the skew of the data using the location of the majority of the data values rather than the direction the data (*NOTE: students are not expected to have mastered the direction of skew in a data distribution in sixth grade. Rather, it is more important that a student realizes the data is skewed.*)	ANALYZE A DATA DISTRIBUTION BY OVERALL SHAPE

5.b. Determine the mean in the context of the data. Provide appropriate work to support your answer.

CORRECT ANSWER

Sum of the data values: 1,357 (student may show the addition)
Number of data values: 23
1,357 ÷ 23 = 59
The mean of the data is 59 miles per hour.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
The mean is 60 miles per hour.	confuses mean and median	EXPLAIN *MEAN* and EXPLAIN *MEDIAN*
The mean is 61 miles per hour.	confuses mean and mode	EXPLAIN *MEAN* and EXPLAIN *MODE*
Student provides a decimal that is close to 59 but is not the mean.	made a mistake determining the sum of the data values or made a mistake when counting the number of data values	CALCULATE *MEAN*

5.c. Determine the median of the data. How well does the median describe the center of the data?

CORRECT ANSWER

The median of the data is 60 miles per hour. This value describes the center of the data well, because when looking at the dot (line) plot, the center of the data appears to be approximately 60.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
The median is 59 miles per hour.	confuses mean and median	EXPLAIN *MEAN* and EXPLAIN *MEDIAN*
The median is 61 miles per hour.	confuses median and mode	EXPLAIN *MODE* and EXPLAIN *MEDIAN*
The median is 56 miles per hour.	determines the middle value between 50 and 62 (or 49 and 63); determines the median of the values labeled on the number line rather than the actual data values	EXPLAIN *MEDIAN* and CALCULATE *MEDIAN* FOR AN ODD NUMBER OF DATA
Student determines the median incorrectly (such that it does not describe the center of the data), then states the median does not describe the center of the data.	incorrectly determines the median for a data set	CALCULATE *MEDIAN* FOR AN ODD NUMBER OF DATA

5.d. Determine the mode of the data. How well does the mode describe the center of the data?

CORRECT ANSWER

The mode of the data is 61 miles per hour. This value is on the higher end of the data and seems to be greater than the center of the data.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
The mode is 59 miles per hour.	confuses mean and mode	EXPLAIN *MODE* and EXPLAIN *MEAN*
The mode is 60 miles per hour.	confuses median and mode	EXPLAIN *MODE* and EXPLAIN *MEDIAN*
The mode is 62 miles per hour.	interprets the mode as the greatest value in the data set rather than the value that appears most often	EXPLAIN *MODE* and IDENTIFY *MODE*
The mode is in the center of the data.	assumes the mode must be in the center of the data because it is a measure of central tendency; does not recognize that 61 miles per hour is greater than the majority of the data	EXPLAIN *MEASURE OF CENTRAL TENDENCY*

5.e. Determine the range of the data. What does this number describe about the data?

CORRECT ANSWER

> The range of the data is 12 miles per hour. This means the greatest (fastest) and least (slowest) data values, or car speeds on the highway, are 12 miles per hour apart.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
The range is 50 to 62 miles per hour.	provides the greatest and least values in the data set but does not know to determine the difference between the two in order to calculate the range	CALCULATE *RANGE* and EXPLAIN *RANGE*
The range is 6 miles per hour.	ignores the outlier at 50 miles per hour and determines the range for the remaining values	CALCULATE *RANGE* and EXPLAIN *RANGE*
The range is 14 miles per hour.	subtracts 49 (the least number labeled on the number line) from 63 (the greatest number labeled on the number line)	CALCULATE *RANGE* and EXPLAIN *RANGE*

5.f. Determine the mean absolute deviation of the data. Provide appropriate work to support your answer. What does this number describe about the data?

CORRECT ANSWER

Total distance of the data values from the mean: 44 (student may show the addition)
Number of data values: 23
$44 \div 23 \approx 1.913$
The mean absolute deviation of the data is approximately 1.913 miles per hour. This means that, on average, the data values vary from the mean of 59 miles an hour by approximately 1.913 miles per hour.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
The mean absolute deviation is 44 miles per hour.	determines the total amount that the values vary or differ from the mean but does not divide by the number of values	CALCULATE *MEAN ABSOLUTE DEVIATION*
The mean absolute deviation is 2.2 miles per hour.	divides the total amount that the values vary or differ from the mean by 20 instead of 23 (does not count the three values that are zero units from the mean in the total number of values)	CALCULATE *MEAN ABSOLUTE DEVIATION*
The mean absolute deviation is 12 miles per hour.	confuses the range and the mean absolute deviation	EXPLAIN *RANGE* and EXPLAIN *MEAN ABSOLUTE DEVIATION*
Student does not describe or inaccurately describes what the mean absolute deviation explains about the data.	does not understand the intent of the mean absolute deviation; may only be able to follow a procedure to determine the mean absolute deviation	EXPLAIN *MEAN ABSOLUTE DEVIATION*

6. Which measure(s) of central tendency (i.e., mean, median, and mode) and measures of variability (i.e., range and mean absolute deviation) from this unit can be determined for categorical data? Explain your reasoning.

CORRECT ANSWER

Only the mode can be determined for categorical data. The mode describes the data value(s) that occurs most frequently, which could be a number or a category (word). The remaining measures of central tendency require addition or ordering values from least to greatest, which cannot occur for categorical data. Similarly, the measures of variability (range and mean absolute deviation) require addition or determining the greatest and least values in order to subtract, which cannot occur for categorical data.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
Student identifies the mean as a measure of central tendency that can be determined for categorical data.	may not realize categorical data are nonnumeric and therefore cannot be added together or divided to determine the mean, or does not know how to calculate the mean	EXPLAIN *MEAN* and CALCULATE *MEAN*
Student identifies the median as a measure of central tendency that can be determined for categorical data.	may not realize categorical data are nonnumeric and therefore cannot be ordered from least to greatest to determine the median, or does not know how to calculate the median	EXPLAIN *MEDIAN* and CALCULATE *MEDIAN* FOR AN ODD NUMBER OF DATA
Student identifies the range as a measure of variability that can be determined for categorical data.	may not realize categorical data are nonnumeric and therefore do not have a least or greatest value to determine the range, or does not know how to calculate the range	EXPLAIN *RANGE* and CALCULATE *RANGE*
Student identifies the mean absolute deviation as a measure of variability that can be determined for categorical data.	may not realize categorical data are nonnumeric and therefore cannot be added together or divided to determine the mean, which is the measure the mean absolute deviation is based on, or does not know how to calculate the mean absolute deviation	EXPLAIN *MEAN ABSOLUTE DEVIATION* and CALCULATE *MEAN ABSOLUTE DEVIATION*
Student identifies the mode as the only measure that can be determined for categorical data, but cannot explain why it can be determined for categorical data (and the other measures cannot).	knows only the mode can be determined for categorical data, but cannot explain why this is the case	EXPLAIN *MODE*

7. One way of thinking about the mean of a data set is as a "fair share". Explain how this description of the mean is related to the process used to calculate the mean.

CORRECT ANSWER

To determine the mean, all data values are added together, then divided by the number of values there are. If you think of a situation where multiple people have different amounts of candy and you are trying to share the candy fairly or evenly among each person, you could combine all the candy (add the candy together) and then redistribute (divide) the candy among the people who are there so everyone gets the same amount.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
Fair share is if you add all the numbers up and divide by how many numbers there are.	only can describe how to find the mean	EXPLAIN *MEAN*
Student cannot explain how mean as a "fair share" relates to the process to calculate the mean.	is not able to describe how to find the mean or how the process relates to the concept of mean as a fair share	EXPLAIN *MEAN*
NOTE: *This question requires students to explain mean in an extended, more sophisticated manner. Students who are able to correctly describe the mean as a fair share have a relatively advanced perception of and ability to explain the mean of a data set.*		

8. Another way of thinking about the mean of a data set is as a "balance point". Explain how this description of the mean is modeled in a dot plot.

CORRECT ANSWER

When a data set is modeled on a dot plot and the mean has been determined, the mean can be thought of as a balance point because the sum of the distances to the mean of the data values less than the mean equals the sum of the distances to the mean of the data values greater than the mean. In other words, the mean "balances" the values that are less than the mean with the values that are greater than the mean.

ERRORS, MISCONCEPTIONS, AND MISSING KNOWLEDGE

Example Error	Misconception	Missing Knowledge
The mean is a balance point because it is a measure of central tendency.	cannot specifically describe how the values less than the mean balance the values greater than the mean; does not relate the mean to the distances to greater and lesser values on a dot plot	EXPLAIN *MEAN*
The mean is a balance point because it is in the center of the data.	cannot specifically describe how the values less than the mean balance the values greater than the mean; does not relate the mean to the distances to greater and lesser values on a dot plot	EXPLAIN *MEAN*
Student cannot explain how mean as a "balance point" relates to a dot plot for the data set.	is not able to describe how the mean is a balance point for a data set modeled on a dot plot	EXPLAIN *MEAN*
NOTE: *This question requires students to explain mean in an extended, more sophisticated manner. Students who are able to correctly describe the mean as a balance point have an advanced perception of and ability to explain the mean of a data set.*		