Table of Contents

CHAPTER 4: STATISTICS (6 WEEKS) **2**
4.0 Anchor Problem: The Electoral College *8*
SECTION 4.1: MEAURES OF SHAPE 11
4.1a Class Activity: What is a Statistical Question? *12*
4.1a Homework: What is a Statistical Question? *17*
4.1b Class Activity: Creating and Analyzing Dot Plots *20*
4.1b Homework: Creating and Analyzing Dot Plots *27*
4.1c Class Activity: Creating and Analyzing Histograms *31*
4.1c Homework: Creating and Analyzing Histograms *37*
4.1d Self Assessment: Section 3.1 ..... *42*
SECTION 4.2: MEAURES OF CENTER 47
4.2a Class Activity: How Much is a Handful? *48*
4.2a Homework: Mean as a Fair Share *51*
4.2b Class Activity: Finding the Mean *52*
4.2b Homework: Finding the Mean *57*
4.2c Class Activity: Mean as a Balance Point *60*
4.2c Homework: Mean as a Balancing Point *66*
4.2d Class Activity: Median and Mode *69*
4.2e Class Activity: Choosing the Best Measure of Center *73*
4.2e Homework: Choosing the Best Measure of Center *79*
4.2f Self Assessment: Section 3.2 ..... *82*
SECTION 4.3: MEAURES OF VARIABILITY 85
4.3a Class Activity: Variability and Mean Absolute Deviation *86*
4.3a Homework: Variability and Mean Absolute Deviation *96*
4.3b Class Activity: The Interquartile Range *102*
4.3b Homework: The Interquartile Range *108*
4.3c Class Activity: Box Plots *112*
4.3c Homework: Box Plots *118*
4.3d Class Activity: Analyzing Box Plots *122*
4.3d Homework: Analyzing Box Plots *127*
4.3e Self Assessment: Section 3.3 ..... *128*
SECTION 4.4: INTREPRET AND DRAW CONCLUSIONS ABOUT DATA 132
4.4a Class Activity: Make, Interpret and Draw Conclusions About Data Part 1 ..... *133*
4.4a Homework: Make, Interpret, and Draw Conclusions About Data Part 1 ..... *136*
4.4b Class Activity: Make, Interpret, and Draw Conclusions About Data Part 2 ..... *138*
4.4b Homework: Make, Interpret, and Draw Conclusions About Data Part 2 ..... *144*
4.4c Self Assessment: Section 3.4 ..... *146*

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Chapter 4: Statistics (6 Weeks)

Common Core Standard(s)

- 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.1)
- 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. (6.SP.2)
- 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. (6.SP.3)
- Display numerical data in plots on a number line, including dot plots, histograms, and box plots. (6.SP.4)
- Summarize numerical data sets in relation to their context, such as by: (6.SP.5)
    a) Reporting the number of observations.
    b) Describing the nature of the attribute under investigation, including how it was measured and its unit of measurements.
    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.
    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.

Vocabulary: statistical question, numerical data, categorical data, data distribution, attribute, dot plot, frequency table, histogram, shape, cluster, gap, peak, skew, symmetric, outlier, measures of center, mean, absolute deviation, median, mode, measures of variability, spread, mean absolute deviation (MAD), maximum (max), minimum value (min), range, quartiles, 1st quartile (lower quartile), 3rd quartile (upper quartile), the 5-number summary, interquartile range (IQR)

This chapter is very heavy on vocabulary. Consider making a word wall, vocabulary cards, Frayer models or foldables to provide extra support for learning these terms.

Chapter Overview:

In this chapter students build on their knowledge and experience in data analysis developed in previous grades. They begin the chapter by determining "What is a statistical question?" They learn that statistical questions involve variability in the data collected to answer them. Investigating statistical questions helps students begin to develop an understanding of statistical variability. As students continue to expand their knowledge in data analysis they begin to characterize data distributions by measures of shape, center, and spread. These measures guide the investigation of data analysis throughout the chapter. Working with counts or measurements, students display data with dot plots, histograms, or box plots in order to analyze the shape of the data. They use terms such as cluster, peak, gap, symmetric, skew, and outlier to describe the shape of the data in these graphical displays. In 6th grade, students develop well-defined numerical measures that describe the center and spread. The center is measured by the mean, median or sometimes mode. The variability or spread is measured by the mean absolute deviation (MAD), range, or interquartile range (IQR). Once students become more familiar with the components of shape, center, and spread they use them together to summarize and describe distributions, often comparing two or more data sets. They take into account the context in which the data is presented and answer statistical questions.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

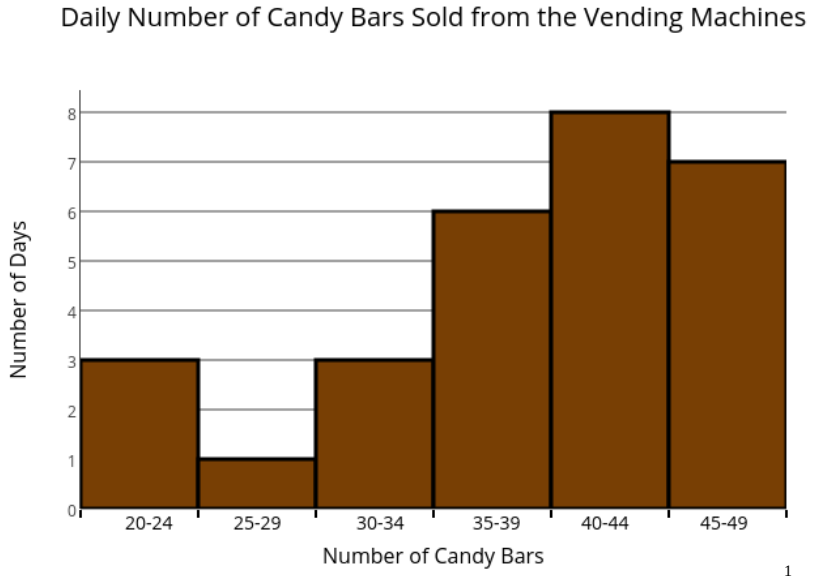
Connections to Content:

Prior Knowledge: In previous grades, students use their work with data to build the foundation for statistical study in 6th grade and beyond and to strengthen and apply their study of arithmetic. As early as kindergarten, students begin to classify and sort objects into categories by count. Throughout elementary school they organize, represent, and interpret categorical data and measurement data (See Table 1 in the K-3, Categorical Data; Grades 2-5, Measurement Data Progressions document). Their work with data is closely related to the number line and fractions. They ask and answer questions about the total number of data points within each category and make comparisons. They display their data on line plots (essentially the same thing as dot plots), picture graphs, and bar graphs and answer questions about data given in these plots as well. This investigative process is continued and extended in a more formal statistical approach in 6th grade.

Future Knowledge: In Grade 7, students move from concentrating on analysis of data to production of data. They understand that good statistical answers depend on a well-developed plan for collecting data. They investigative random sampling, and in turn, concepts related to probability. Until 8th grade, almost all encounters with data analysis have been with univariate data (collections of counts with one variable or characteristic). In 8th grade, students extend their knowledge of shape, center, and spread to the analysis of bivariate data, (collection of counts with 2 variables or characteristics) as related to their work with linear functions. In high school statistics, students enhance their understanding of data analysis to give more precise answers to deeper statistical questions. For example, they use standard deviation as a measure of variability that is based on many of the same principals as the mean absolute deviation.

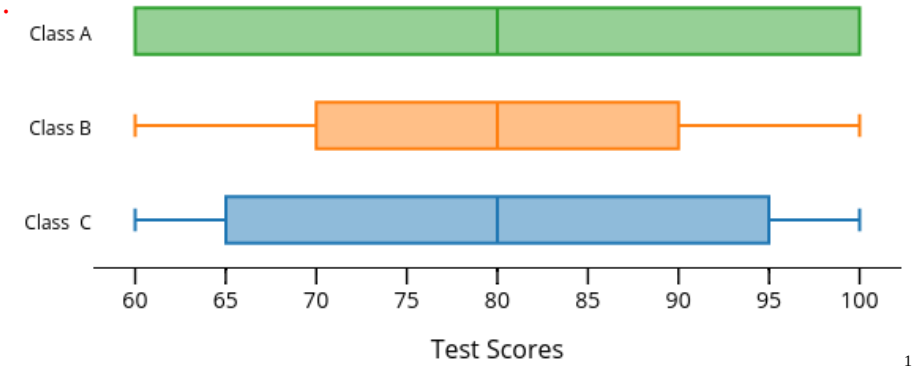
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

MATHEMATICAL PRACTICE STANDARDS

	Make sense of problems and persevere in solving them.	Roman thinks that his school needs to get another vending machine in the school cafeteria because there is always a long line to use the machine at lunch time. The principal has told everyone that they cannot get another vending machine unless they can show that on average at least 40 candy bars are sold each day. Roman conducts a survey and recorded his results in the histogram below.  [1] What possible arguments could Roman give to his principal to convince him that on average at least 40 candy bars are sold from the vending machine each day? *The example problem given above shows how students must make sense of practical problems and turn them into statistical investigations. They must make sense of what statistical arguments can be made about the data. They must determine what statistical measures might be used to support their arguments and how to go about finding them. Throughout the solving process, students must stop and evaluate their progress. Once they have arrived at a desired solution they can look back and evaluate if their outcome makes sense given the context in which the data is presented.*
	Reason abstractly and quantitatively.	A family of five goes out to lunch. The numbers below show the cost of each item that a person ordered for lunch. They decide to split the bill so that everyone pays the same amount. What is the average cost per person for dinner? $8.25 $8.70 $7.50 $6.95 $9.35 Discuss how to find the mean for any given set of data. *As you discuss how to find the mean for any given set of data students are abstracting the solving process. They must understand that for any data set they can sum the data and divide by the number of data values to get the mean. They come to this understanding by reasoning through the problems and repeatedly creating "fair shares".*

1 evatipp. "Daily Number of Candy Bars Sold from Vending Machines" Plotly Technologies Inc., https://plot.ly/516/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

	Construct viable arguments and critique the reasoning of others.	The three box plots below represent the test scores for three different classes. Examine each plot and then discuss the questions that follow.  [1] a. What is the same about these box plots and what is different? b. Find the IQR for each plot and use them to compare the variability of each set of class scores. c. Make an argument for each class that supports the claim that this class performed the best on the test. *The study of statistical analysis requires students to construct arguments that are based off of statistical measures. As students compose these arguments they use stated assumptions, definitions, and previously established results. In many instances the interpretation and analysis of statistical information is subjective and well-constructed and viable arguments can be made for many cases. In the example above, students must make an argument that supports the success of each class's test score data. They can see that such arguments depend on preference and the interpretation of success for this test.*
	Model with mathematics.	Below are the ages of 20 people that sing in a choir together. 16, 21, 24, 29, 30, 30, 32, 32, 32, 34, 34, 35, 39, 42, 42, 43, 45, 47, 52, 61 a. Describe the distribution of ages for people in this choir. Be sure to describe shape, center and variability. b. What is the (arithmetic) averag age for people in this choir? State what kind of numerical measure you used to answer this question, explain your choice. b. Are most of the people in the choir around the same age? State what kind of numerical measure you used to answer this question, explain your choice. *Modeling with mathematics is a major theme throughout this chapter. Not only do students need to know how to interpret data that is displayed graphically but they also need to be able to construct their own models of data. They use dot plots, histograms, and box plots to analyze and draw conclusions about the distribution of data in a variety of contexts.*

1 evatipp. "Test Scores Box Plot (2)" Plotly Technologies Inc., https://plot.ly/447/~evatipp/ (Last Accessed: March 2016)

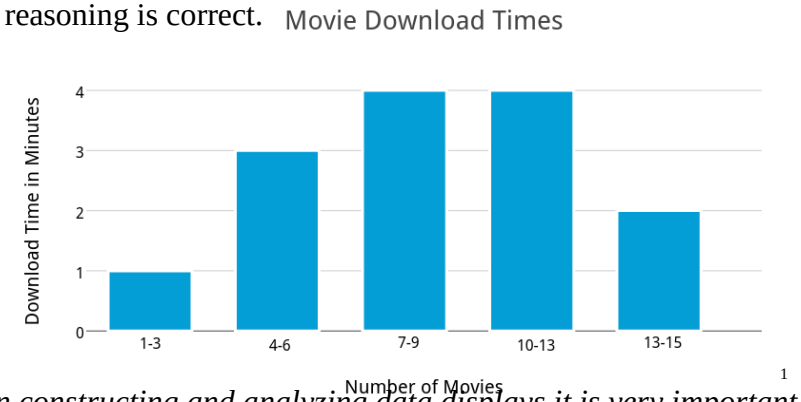
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Attend to precision.

The table given below shows the time that it takes to download different movies.

Time to download a movie (minutes)	1 – 3	4 – 6	7 – 9	10 – 12	13 – 15
Number of movies	1	3	4	4	2

Carmen made a histogram of the data and has made many mistakes. Find at least two of her mistakes, fix them on the histogram and justify why your reasoning is correct.



[1]

When constructing and analyzing data displays it is very important to pay attention to the scale as this can greatly affect how the data is viewed. This is especially important when students interpret the shape of the data or compare two or more plots. As students attend to precision in constructing data displays they will be able to clarify the correspondence that is occurring within that data distribution that data distribution.

Look for and make use of structure.

The most recent test scores for Mr. Petrov's science class are shown in the table.

Mr. Petrov's Test Scores					
79	91	72	89	84	77
71	82	80	76	91	83
80	79	82	81	78	78
82	72	86	81	80	64
83	79	82	84	81	81
92	80	80	87	83	81

a. Make a dot plot of the data. Be sure to label your number line and give it a title.
b. How many students are in Mr. Petrov's science class?
c. Which test score was earned by the largest number of students?
d. Describe any peaks, clusters, or gaps in the data by marking them on the plot.
e. What is the overall shape of the data? Justify your answer.
f. Mr. Petrov asks, "What is the most typical score for this test?" Use the distribution of data to answer this question.

Students use structure when analyzing the shape of a data distribution. While identifying peaks, clusters, gaps, skewness, symmetry, and outliers students infer more knowledge about the characteristics of the data. While looking at shape is not numerical structure it is graphical structure and being able to identify what that structure infers about the data is imperative in its analysis.

1 evatipp. "Movie Download Times" Plotly Technologies Inc., https://plot.ly/160/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

	Use appropriate tools strategically.	The high scores for a popular video game are shown below. Li wants to know what the typical high score for this video game is. 90 78 62 84 92 66 77 6 65 53 79 71 71 59 51 69 21 35 44 52 27 38 35 50 105 71 60 48 57 42 22 68 56 45 41 30 64 63 65 60 Li begins to make a dot plot for this data but soon realizes that a dot plot is not the best graphical display for this data. Discuss why a dot plot may not be best for this data and see if you can come up with a better way to represent the data. *Throughout this chapter students will use graphical displays to interpret the shape, center, and spread of the data. These graphical displays are mathematical tools. It is important to choose a graphical display that best suits your data, and your desired measurement. In the example above, the range of the data and the sample space is large, so a dot plot may not be as desirable as a histogram where you can use intervals that summarize the values. At other times a box plot might be preferred because this shows a "snap shot" of the 5-number summary and a visual picture of the variability within the data distributions. Students must also choose appropriate numerical measures of center (mean, median, mode) and variability (MAD, Range, IQR) when interpreting data.*
	Look for and express regularity in repeated reasoning.	What does the shape of the data distribution tell you about which measure of center to use to summarize the data? *After students have analyzed the shape and center of a series of data distributions they begin to understand through repeated reasoning that generally if the shape of the data is fairly symmetrical then the mean is a good measure of center. If the shape of the data is skewed or there are significant outliers, then a good measure of center is the median. Similarly, it is through repeatedly analyzing the shape, center, and spread of several data distributions that they begin to understand how to interpret data, make meaningful conclusions, and answer statistical questions.*

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.0 Anchor Problem: The Electoral College

(Illustrative Mathematics Task)

This anchor problem should be revisited throughout the chapter as students gain more knowledge of how to analyze and interpret the shape, center, and spread of a data distribution. It could even be used as a culminating assessment for the chapter. There is a lot of information about the Electoral College listed below; you may want to discuss or paraphrase how the Electoral College works in the United States with your students in a class discussion. Consider the reading level of your students and provide scaffolding and/or modifications as needed. There is obviously a great cross-curricular connection between this problem and social sciences/civics.

Unlike many elections for public office where a person is elected strictly based on the results of a popular vote (i.e., the candidate who earns the most votes in the election wins), in the United States, the election for President of the United States is determined by a process called the Electoral College. According to the National Archives, the process was established in the United States Constitution "as a compromise between election of the President by a vote in Congress and election of the President by a popular vote of qualified citizens." (*http://www.archives.gov/federal-register/electoral-college/about.html* accessed September 4, 2012).

Each state receives an allocation of electoral votes in the process, and this allocation is determined by the number of members in the state's delegation to the US Congress. This number is the sum of the number of US Senators that represent the state (always 2, per the Constitution) and the number of Representatives that the state has in the US House of Representatives (a number that is directly related to the state's population of qualified citizens as determined by the US Census). **Therefore the larger a state's population of qualified citizens, the more electoral votes it has.** Note: the District of Columbia (which is not a state) is granted 3 electoral votes in the process through the 23rd Amendment to the Constitution.

The following table shows the allocation of electoral votes for each state and the District of Columbia for the 2012, 2016, and 2020 presidential elections. (*http://www.archives.gov/federal-register/electoral-college/allocation.html* accessed September 4, 2012).

State	Electoral Votes	State	Electoral Votes	State	Electoral Votes
Alabama	9	Kentucky	8	North Dakota	3
Alaska	3	Louisiana	8	Ohio	18
Arizona	11	Maine	4	Oklahoma	7
Arkansas	6	Maryland	10	Oregon	7
California	55	Massachusetts	11	Pennsylvania	20
Colorado	9	Michigan	16	Rhode Island	4
Connecticut	7	Minnesota	10	South Carolina	9
Delaware	3	Mississippi	6	South Dakota	3
District of Columbia	3	Missouri	10	Tennessee	11
Florida	29	Montana	3	Texas	38
Georgia	16	Nebraska	5	Utah	6
Hawaii	4	Nevada	6	Vermont	3
Idaho	4	New Hampshire	4	Virginia	13
Illinois	20	New Jersey	14	Washington	12
Indiana	11	New Mexico	5	West Virginia	5
Iowa	6	New York	29	Wisconsin	10
Kansas	6	North Carolina	15	Wyoming	3

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

If desired consider adding additional discussion questions that will help students analyze the distribution of this data. Some optional discussion questions are in the teacher notes below.

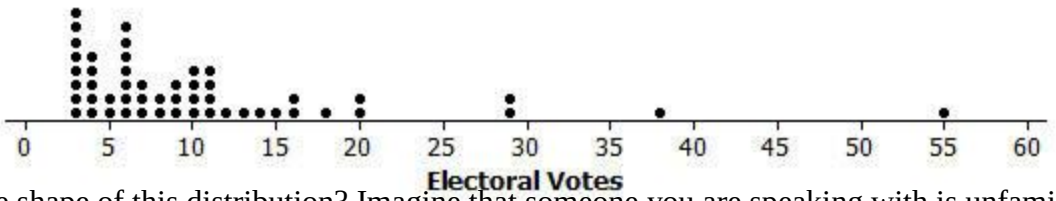
1. Which state has the most electoral votes? How many votes does it have?
   California has 55 electoral votes, this is the maximum value.

Which state has the least electoral votes? What is the smallest value in a data set called? What is the largest value in a data set called?

2. Based on the given information, which state has the second highest population of qualified citizens?
   The larger the state's population the more electoral votes it has. The second highest number of electoral votes is 38. This state, Texas, would be associated with the second highest population.

3. Here is a dot plot of the distribution.

In previous grades, a dot plot was called a line plot. Ask students to recall how to interpret a line plot.



What is the shape of this distribution? Imagine that someone you are speaking with is unfamiliar with these shape terms. Describe clearly and in the context of this data set what the shape description you have chosen means in terms of the distribution.

Student answers will depend on if students are encountering this question before learning the concepts in this chapter or if you are using it as a review or assessment at the end of the chapter. If this is their first encounter with shape ask them what they think is meant by shape.

This distribution is skewed right. Most states have a small number of electoral votes and so most of the data is clustered between the values of 3 and 15. The data has a "tail" the represents a few states with a higher number of electoral votes.

A common mistake for students is to assume that the direction of the "skew" in a data display corresponds with where the data is clustered or peaks. Rather the direction of the "skew" refers to the tail of the data.

If your students are encountering the shape of the data for the very first time, ask them to discuss what they think is meant by the terms skewed left, skewed right, and symmetric.

Additional topics of discussion could be to address any peaks, clusters, or gaps in the dot plot.

4. Does the dot plot lead you to think that any states are outliers in terms of their number of electoral votes? Explain your reasoning, and if you do believe that there are outlier values, identify the corresponding states.
   California at 55 electoral votes and Texas at 38 votes should be listed as outliers based on the visible gaps in the dot plot and/or based on a numerical argument that their values are very far away from the cluster of the other observations. Many students may also say that New York and Florida are outliers, because of the gap of 9 electoral votes between these observations and the "3 to 20" votes cluster.

If students are encountering outliers for the first time ask them to describe what they think is meant by the term outlier.

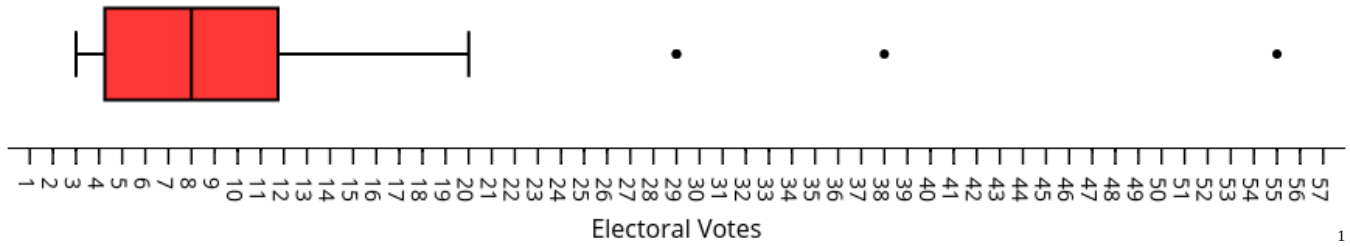
5. What numerical measure or number would you use for describing the "middle" of this data set? Why did you choose this number?
   Because the data distribution is skewed to the right and there are outliers, the median would be a better choice to describe center for this data set. If students have not yet dealt with measures of center discuss that a numerical measure of center describes the number of electoral votes that represent the center or middle of the data.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

6. Determine the value of the median for this data set (electoral votes).
   The median is 8.

7. Find the 5-Number Summary values for this data set.
   $Max = 55, min = 3, median = 8, Q1 = 4, Q3 = 12$

8. Make a box plot for this data set. Use the plot and the IQR to discuss the variability of the data set.



[1]

The $IQR = 12 - 4 = 8$. This indicates that 50% of the data surrounding the median falls within 8 votes of each other. There is little variability for the data if you exclude the outliers.

If desired you can also have students make a histogram of the data.

1 evatipp. "Electoral Votes Final" Plotly Technologies Inc., https://plot.ly/561/~evatipp/(Last Accessed: April 2015)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Section 4.1: Meaures of Shape

Section Overview:
Students begin this section by learning that a statistical question is a question that anticipates variability in the data related to the question and accounts for it in its answers. It is the variability of the data that we try to summarize in statistical analysis. In this section, students will focus on interpreting the shape of the data. To do so they interpret and construct dot plots and histograms of the data. They analyze key features of these graphical displays by looking for clusters, peaks, gaps, and outliers. They also look for whether the data appears to be skewed or symmetric. They answer questions about the data from the data displays including questions that relate to the number observations, the nature of the attribute under investigation, and the context in which the data was gathered.

Concepts and Skills to Master in this Section:

By the end of this section, students should be able to:

1. Recognize a statistical question as one that anticipates variability in the data.
2. Identify if the set of data used to answer a statistical question is numerical or categorical.
3. Display numerical data in a dot plot, use key terms to describe its shape, and interpret the distribution of data.
4. Display numerical data in a histogram, use key terms to describe its shape, and interpret the distribution of data.
5. Summarize a numerical data set in relation to its context by reporting on the number of observations, identifying how the data was measured, and its units of measurement.

Note: Statistics can be very subjective and how questions and data are interpreted may not be as black and white as one might expect for a math topic. Allow students to discuss their reasoning and interpretations. Explain to them that their mathematical interpretation sometimes relies upon how they look at something. This perspective might completely change your answer or how you understand an outcome.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.1a Class Activity: What is a Statistical Question?

Jada, Marco, Ben, Audrey, and Beau each got a different Lego set for Christmas. They each dump out all of their pieces onto the floor to get a better look at what kind of Legos their set includes. As they each sort through their Legos they begin to think about some questions. Their questions are listed below.

Be sure to read the instructions on the next page. These answers are meant to be included as a class discussion and not necessarily written as a student response.

<u>Lego Questions</u>

1. How many Legos are in each set?
   Statistical- Not all sets will have exactly the same number of Legos. There would be variability in the data collected to answer this question. This question could be answered by recording the number of Legos in each set.

2. How many red Legos are in each set?
   Statistical- Not all sets will have the same number of red Legos. There would be variability in the data collected to answer this question. This question could be answered by recording the number of red Legos in each set.

3. How much time is needed for each person to complete the Lego set?
   Statistical- This question is answered by collecting data that varies. It could be answered by recording the amount of time it takes each person to complete their set; not everyone would complete their set in the same amount of time.

4. How much did Marco's Lego set cost?
   Not Statistical –This question is answered by finding the exact cost of this particular Lego set. This produces a single value. However if question were phrased as "What is the typical cost of Marco's Lego set?" it would be a statistical question because this question would be answered by recording the cost of the Lego set from several different stores. There would most likely be variability in the price of the set from store to store.

5. How many Lego pieces have 4 studs (circular bumps on top of the Lego) on them in each set?
   Statistical-This question is answered by collecting data that varies. It would be answered by recording the number of pieces with 4 studs for each set.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

You can use statistics to answer most of these questions. **Statistics** is the science of collecting, analyzing, and interpreting data that answer questions. In order to use statistics to answers questions the question posed must be a statistical question.

> **A *statistical question* is question that generates a variety of answers rather than a single answer. This means that the data related to the question varies or has *variability*.**

Part 1: Identify which of the questions on the previous page are statistical questions. For each question, explain why it is or is not a statistical question.

As students discuss whether each question is statistical, ask them to argue why and justify themselves. Ask other students it their reasoning makes sense.

Part 2: Add two more statistical questions related to the Lego sets. Sample answers are given.

6. How many tiles (no studs on top of the Lego) are in each set? How many blue Legos are in each set? What is the most popular Lego set among 6th graders at my school?

7. How many Lego people are included with each set? How many different Lego sets does each person own?

Directions: Given each pair of questions below, circle the question that is a statistical question and then explain why. If students struggle identifying statistical questions ask them to identify if there could be more than one answer for the question.

8. a. How many siblings do you have? b. How many siblings does each student in your class have? The number of siblings that each student has could be different or vary.	9. a. How many homes were sold each day this month? b. How many homes were sold this month? The number of homes sold each day could be different from day to day so there is variability in this answer.
10. a. On a scale from 1 to 10 how much pain do you feel when you get a shot at the doctor's office? b. On a scale from 1 to 10 how much pain do your classmates feel when they get a shot at the doctor's office? Each student in the class may express a different number to express the amount of pain felt when they get a shot. There is variability in the answer.	11. a. What model of car is most popular for parents to drive in your city? b. What model of car do your parents drive? There is variability in the answer for the model of car that is most popular for parents to drive in your city. The answer could differ from parent to parent.

On the next page students will learn to distinguish between numerical or quantitative data and categorical data. In previous grades, students have made bar graphs that represent categorical data. In 6th grade we only focus on numerical data. However, it is important for students to understand that a statistical question can still generate categorical data.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

In order to answer statistical questions you must collect data. There are two types of data, *numerical* and *categorical*. Suppose you were interested in the Star Wars movies, and collected data on how much money each of the eight (so far) Star Wars movies made. This would be numerical data because each of the eight data values would be a numerical value. If you collected data on who your favorite character was in each movie this would be categorical data. Each of the seven data values would not be a number but a name of a character, which you could think of as a category.

> **There are two types of data, *numerical* (sometimes called quantitative) and *categorical* (sometimes called qualitative). In a numerical data set every value is a number that represents a count or measurement. In a categorical data set every value is non-numerical.**

Together as a class brainstorm other examples of numerical and categorical data.
Numerical Data: Height, Age, Temperature, Number of Pets, # of times you have seen Star Wars, etc
Categorical Data: Eye Color, Month you were born, favorite pet, favorite movie, etc.
Caution students to watch out for some categorical values that are numbers that represent a category, such as, zip codes, phone numbers, room numbers, etc.

12. Identify each of the following data sets as categorical (C) or numerical (N).

If your students struggle, ask them to think of possible data values that would be generated by this data set. If the data values are numbers that are counts or measures, then the data set is numerical. If the data values are words or categories then the data set is categorical.

a. How far can a 6th grader jump? N

b. How many minutes a day does each person in our class read? N

c. What did each person in our class eat for lunch? C

d. How many pencils are in each person's desk? N

e. What is the favorite football team of each of the teachers at our school? C

f. How many letters are in each of our last names? N

g. Can you roll your tongue? C

h. In what zip code do students in our class live? C

13. For each of the statistical questions below, determine if the data is categorical (C) or numerical (N). List three possible data values for each question.

a. How much do boxes of cereals cost at the grocery store?
N; $3.50, $4.00, $5.00

b. What is the most popular kind of cereal sold at the grocery store?
C; Cheerios, Kix, Chex,

c. At which grocery store do people in your neighborhood like to buy cereal?
C; Walmart, Target, Smiths

d. How many boxes of Cheerios were sold at the store each day in the month of January?
N; 50, 45, 60

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Directions: Discuss why each question given below is, or is not, a statistical question. Write "yes" if it is statistical or "no" if not. For the questions that are not statistical, change them so that they can be statistical questions. Then determine if the data collected to answer each question is categorical (C) or numerical (N). You could provide each of these questions on a separate slip of paper and have students sort them into Statistical and Non-Statistical piles. You could also have them sort them into Categorical and Numerical questions.

14. How long does it take you to travel to school every day? Yes; numerical	15. How many boxes of cookies did each Girl Scout sell? Yes; numerical
16. What is your resting heart rate at this moment? No; What is the typical resting heart rate for a 6th grader? numerical	17. How many jumping jacks can a kindergarten student do in 30 seconds? Yes; numerical
18. Which brand of bubble gum holds its flavor the longest? Yes; categorical	19. What is the typical number of cars that each family owns on your street? Yes; numerical
20. How many text messages did you send yesterday? No; what is the average number of text messages that a 6th grader sends per day?; numerical	21. Which video game do 6th graders like to play? Yes; categorical
22. How many cars in your town were sold this month? No; How many cars were sold each month this year?; numerical	23. How many days have people in my class been absent from school so far? Yes; numerical
24. How many cars were sold each day this month? Yes; numerical	25. How many pets does each of the students in my class have? Yes; numerical
26. How old is Caden's pet? No; What is the typical age for pets of kids in our class?; numerical	27. What did you eat for lunch yesterday? No; What did 10 of my friends eat for lunch yesterday?; categorical

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

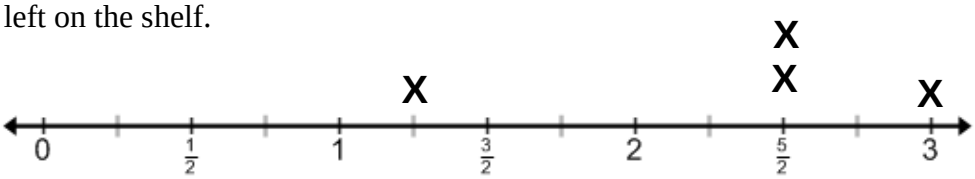
28. What pizza toppings do the people in my class like? Yes; categorical	29. What is your teacher's favorite type of pizza? No; What is the most liked pizza of all the teachers at my school?; categorical
30. How many pockets do I have on the clothes I am wearing today? No; what is the typical number of pockets on people's clothes in my class today?; numerical	31. What is the speed of cars driving in front of our school? Yes; numerical

32. Your class is having an ice cream party. Create two statistical questions that could help make your ice cream party a success. Then determine if the data collected to answer each question is categorical or numerical data.

Answers may include the following;
What toppings do the students in my class like on their ice cream? Categorical
What is the most popular flavor of ice cream in my class? Categorical
How many scoops of ice cream do people in my class want to eat? Numerical
How much do different brands of ice cream cost? Numerical

Spiral Review

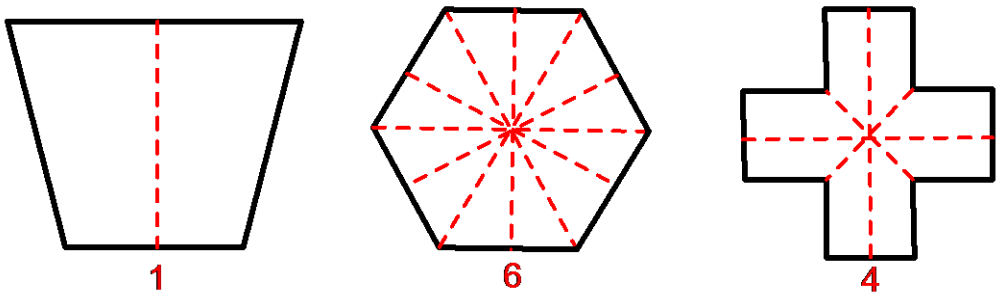
1. Barbara is measuring several objects to line up on a small shelf in her room. The shelf is 10 inches long. She has recorded the length of each object on the line plot below. Use the information given in the line plot to determine if Barbara's shelf is long enough to fit all of the objects, and if so, how much room will she have left on the shelf.



Length of Objects on Shelf (inches)

The length of all the objects together is 9 ¼ inches. She has ¾ of an inch left on the shelf.

2. Find and draw all lines of symmetry for the following figures. Write the number of lines of symmetry that you found next to each figure.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.1a Homework: What is a Statistical Question?

Yesterday, Ruth and Carl invited 10 friends to go out to lunch. The questions below came up during the meal. Decide whether or not each question is a statistical question, and justify your decision.
Some of these questions are subjective. Allow students to justify their reasoning based on their own interpretation of the question.

1. How much does each person's meal cost?
   Statistical- This questions would be answered by collecting data on the amount of the bill for each person attending the lunch. It is unlikely that each person ordered items that all cost the same amount.

2. How long did it take for each person to get to the restaurant?
   Statistical- This question would be answered by collecting data on the amount of time that it took for each person to travel to the restaurant. Unless they all left from the same place and traveled together there would be variability in the answer.

3. Would Carl rather have burgers or pizza?
   Not Statistical- This question refers to a single person and so it would be answered based on a single response and not on data that vary.

4. What percent of the bill did each person leave for a tip?
   This could be a non-statistical question if, for example, each friend always leaves the same fixed percent of the bill for a tip or if the restaurant automatically adds on a tip for large groups. On the other hand, each person may base the percentage off of their dining experience taking into consideration how much they enjoyed the meal and service. This would make it a statistical question.

5. What was the most frequently ordered dish in the restaurant during lunch time?
   Not Statistical- This is not a statistical question if it refers to a particular day. There is a single unambiguous answer with no variability expected. On the other hand, if the question had been "Is pizza the most popular dish for lunch on Saturday?" this would could be considered to be a statistical question. It would be answered by looking at data for many Saturday lunches and there would be variability in the data.

6. Add one more statistical question of your own that relates to this situation. Justify why it is a statistical question.
   Sample Answer: How many people were sitting at each table in the restaurant?
   Like the previous question, this is a statistical question. It would be answered by collecting data on the number of people at a table from many tables and there would be variability in the data collected.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Directions: Given each pair of questions below, circle the question that is a statistical question and then explain why.

7. a. On a scale from 1 to 10 how much do you like math? b. On a scale from 1 to 10 how much do the people in your class like math? Each student in the class may choose a different number to express how much they like math. There is variability in the answer.	8. a. How long will it take to drive to the campground? b. How long did it take you to drive to the campground on Saturday? There is variability in the answer for how long it will take you to drive to the campground because of variables such as stoplights, the route taken, the amount of traffic, etc. The answer could differ from driver to driver and day to day.

9. Identify each of the following data sets as categorical (C) or numerical (N).

   a. How old are the grandmas for people in our class? N

   b. On which day of the month were you born? N

   c. How many pieces of paper are in each person's backpack in our class? N

   d. What color of shoes are people in my class wearing today? C

   e. What is the favorite subject in school for people in your class? C

   f. How many marshmallows can you eat in 30 seconds? N

   g. Can you do a flip on a trampoline? C

   h. What is the area code for people's cell phone numbers in your class? C

10. Create a statistical question that could be answered by collecting numerical data.
    Answers will vary

11. Create a statistical question that could be answered by collecting categorical data.
    Answers will vary

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Directions: Discuss why each question given below is, or is not, a statistical question. Write "yes" if it is statistical or "no" if not. For the questions that are not statistical, change them so that they can be statistical questions. Then determine if the data collected to answer each question is categorical (C) or numerical (N).

12. How much does a gallon of gas cost? Yes; numerical	13. Who is Mr. Reed's favorite NFL football player? No; Who is the most popular NFL football player for teachers at our school?; categorical
14. How many laps can a 6th grader run during P.E. time? Yes; numerical	15. How many songs do you have on your MP3 player? No; what is the average number of songs on a 6th grader's MP3 player?; numerical
16. What is the typical number of TVs that each person in your class has at home? Yes; numerical	17. Which brand of AAA batteries lasts the longest? Yes; categorical
18. How much does your backpack weigh? No; what is the average weight of people's backpacks in our class? numerical	19. How many cans of orange soda did the grocery store sell each day this week? Yes; numerical

20. You are starting a small business mowing lawns in your neighborhood for the summer. Create two statistical questions that will help make your lawn mowing business a success. Then determine if the data collected to answer each question is categorical or numerical.

Sample answers are given.
What kind of gasoline makes my lawnmower run the best? Categorical
How many gallons of gasoline will it take to mow the typical yard in my neighborhood? Numerical
How much do different kinds of gasoline cost? Numerical
How big are the yards in my neighborhood? Numerical
How long will it take me to mow a typical yard in my neighborhood? Numerical

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.1b Class Activity: Creating and Analyzing Dot Plots

In this lesson students are introduced to a tool that can be used to help them answer a statistical question. This tool is a dot plot, which is a model of the data set. Being able to utilize modeling tools helps in the problem solving process.

Each member of your class is going to make and fly a paper airplane using the instructions given below or your own method. Write one statistical question that you could ask in relation to the paper airplanes. Be ready to share your question with the class.

1. My statistical question about the paper airplane is, _____

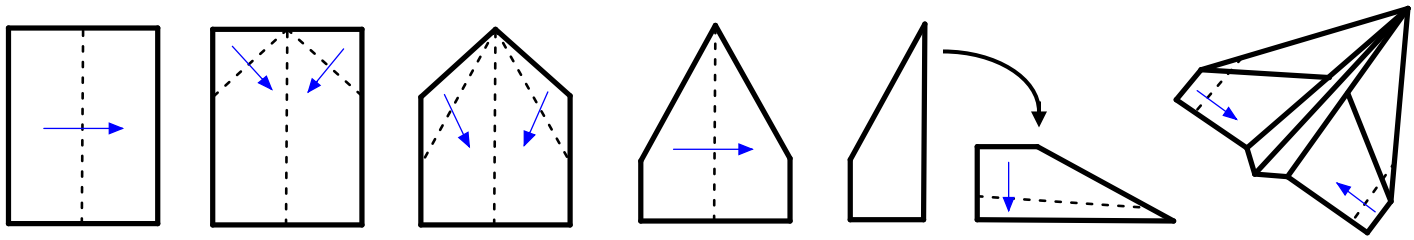
_____

Possible questions may include: How long does it take a 6th grader to make a paper airplane? How far will a paper airplane fly? How many paper airplanes can our entire class make each day this week? Which paper folding method will make the airplane fly the farthest distance? Have students share their questions with the class. Be sure to discuss how each statistical question suggested anticipates variability in the data related to the question and its answer. Also take time to discuss how each non-statistical question suggested does not anticipate variability.

This is great opportunity for students to critique the reasoning of others as you discuss the different questions suggested and determine as a class whether they are statistical questions or not. Ask the students to explain why or why not the questions offered are statistical questions. Be sure to provide examples of non-statistical questions if students do not bring them up on their own.

Instructions for folding a Paper Airplane

Make an airplane from a piece of paper that measures 8 ½ by 11 inches.



1. Fold paper in half and then unfold.
2. Fold the top two corners.
3. Fold corners again.
4. Fold paper in half again.
5. Rotate your paper 90 degrees.
6. Fold wings out on both sides.
7. Fold wing tips up.

We are going to try and answer the statistical question of,

"How far will a paper airplane fly?"

Predict how far you think the typical paper airplane made in your class will fly.

2. Prediction:_____

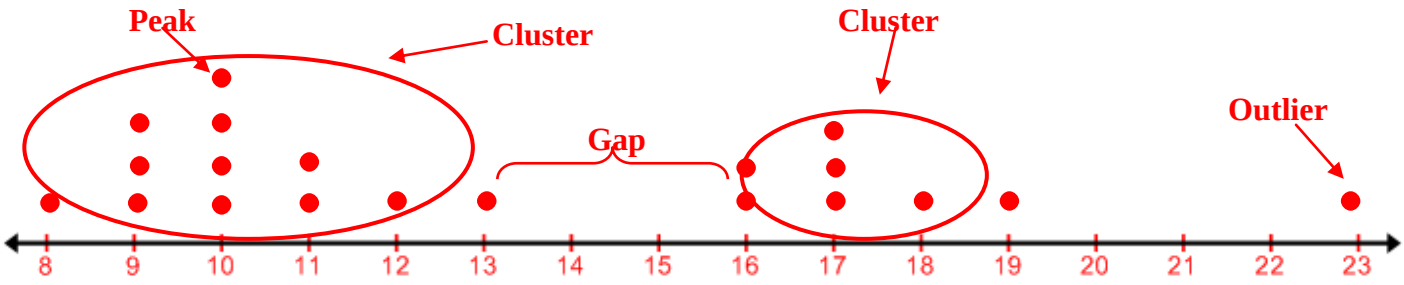
3. What units of measurement do you think the class should use and why?
   Have a discussion about what measurements would best be used to collect the distance that the airplanes fly. Students might suggest inches, feet, or yards. All of these units will work; however for the sake of discussion decide on one unit of measurement as a class. You can use this opportunity to review ratios by converting between inches and feet if desired.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Follow the instructions from your teacher for gathering the data.

Choose an appropriate location for students to perform this experiment. Have each student throw their airplane and have a partner measure the distance that it travels. Ask them to write the distance for their plane on a sticky note (tell them to write it really big) and stick it on the board. You can then have a discussion about how to organize the data. Students might suggest a table, sorting, or even a dot/line plot (they have done line plots for plotting fractions in previous grades). Organize the data by making a dot plot; begin by scaling the number line by asking for the smallest (min) and biggest number (max). Label the number line and have students place their sticky note above the appropriate number. Repeated numbers should be stacked above each other. Be sure to label the dot plot and give it a title.

4. Copy the dot plot made by the class below by using a dot to represent each data value.
   A sample plot is shown with notes from discussion questions given at the bottom of the page.



Distance of Paper Airplanes in Feet

This problem will be referred to in later sections when discussing measures of center. You may want to take a picture of your class's data for easy reference.

A **dot plot** uses a number line to show the number of times each value in a data set occurs. A dot plot shows the ***distribution*** of the data set. Distribution describes how the data is related and organized; it can be described by its ***center***, ***spread (variability),*** and overall ***shape***. We are going to discuss how to interpret the shape of the data set using words like **peak, gap, cluster, symmetrical, skewed,** and **outlier**.

While students investigate the shape of a model, they are looking for what the structure of the data values tells us about the distribution.

Lead a discussion about the distribution of the airplane data in the dot plot. You could discuss the following questions. Encourage students to mark their dot plot with features that are addressed in the questions below.

1. How can you determine the total number of students that participated in the experiment from the dot plot?
2. What does each dot on the plot represent?
3. What does it mean when dots are stacked on top of each other?
4. Choose a dot of a particular student and ask the class who that dot represents.
5. How far did most of the airplanes fly?
6. Did more airplanes fly more than 15 ft or less than 15 ft?
7. Describe any peaks, clusters, or gaps in the data. Have students mark these on their plot.
8. Describe any outliers in the data.
9. Describe the shape of the data.
10. Is the data skewed?
11. How does the data help us to answer our question of "How far will a paper airplane fly?"

At this point accept arguments that just include discussion about the shape of the data. Note that some students may begin to formalize ideas about *measures of center* (mean, median, and mode) and *measures of spread/variability* to answer the statistical question. Encourage them to describe their reasoning and thinking, which should be very informal at this point.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

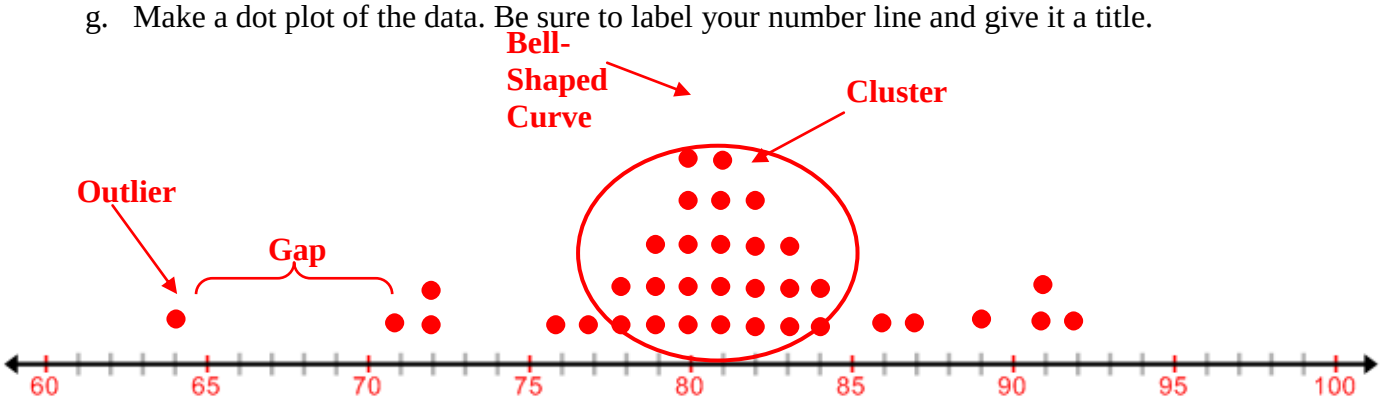
5. How close was your prediction?

Use the blank space on the previous page in the student manual for students to make notes and given explanations for key vocabulary words discussed during the lesson, such as; peaks, clusters, gaps, outliers, shape, symmetrical, bell-shaped curve, skewed left, skewed right. A common mistake for students is to assume that the direction of the "skew" in a data display corresponds with where the data is clustered or peaks. Rather the direction of the "skew" refers to the tail of the data.

6. The most recent test scores for Mr. Petrov's science class are shown in the table. Make a dot plot of the data. Be sure to label your number line and give it a title.

Mr. Petrov's Test Scores					
79	91	72	89	84	77
71	82	80	76	91	83
80	79	82	81	78	78
82	72	86	81	80	64
83	79	82	84	81	81
92	80	80	87	83	81

g. Make a dot plot of the data. Be sure to label your number line and give it a title.



Test Scores in Mr. Petrov's Science Class

h. How many students are in Mr. Petrov's science class?
36

i. Which test score was earned by the largest number of students?
80 and 81 are the most common scores.

j. Describe any peaks, clusters, or gaps in the data by marking them on the plot.
The dot plot shows most of the data clustered around 80. There is a gap between the scores of 64 and 71 and the point 64 appears to be an outlier.

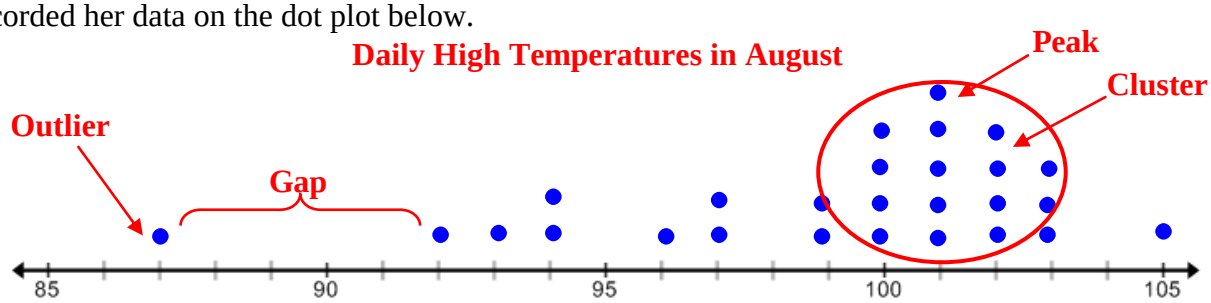
k. What is the overall shape of the data? Justify your answer.
The shape of this data is somewhat symmetrical or bell-shaped.

l. Mr. Petrov asks, "What is the most typical score for this test?" Use the distribution of data to answer this question.
Based off of the features in the dot plot the most typical score is around 80.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

When creating data displays it is so important to attend to the precision of labeling your model so that all the information that the model presents is clearly understood.

7. Marta records the high temperatures for each day she goes swimming in the month of August. She has recorded her data on the dot plot below.



Temperature in degrees Fahrenheit

a. Marta did not label her dot plot with units or a title. Determine the appropriate units for this data and how the data was collected. Then give the dot plot an appropriate title.
The most appropriate units for this plot are degrees Fahrenheit. Marta most likely used a thermometer to collect the temperature data. See title on plot above.

b. How many days did Marta attend the pool in August?
Marta went to the pool 27 days in August.

c. Mark any clusters, gaps, or peaks that you observe on the graph.
See plot above.

d. What is the overall shape of the data?
The data is skewed left. Most of the data is clustered around the right side of the graph. As the temperature increases the number of data entries also increases.
The skew of the data is often counter-intuitive to students. The skew refers to the "tail" of the data. Since the data in the plot above has a tail on the left, it is skewed left.

e. Write and answer a statistical question related to the data shown in the plot.
What is the typical high temperature at Marta's pool in August? The typical high temperature is between 100-102 degrees Fahrenheit. Again, students are very informally reasoning about measures of center and spread as they answer this question.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

8. The numbers below represent the amount of time in hours that several 6th grade students spent doing homework last week.

5, 15, 10, 11, 11, 6, 8, 0, 6, 13, 11, 7, 7, 12, 7, 12

a. What attribute is being measured in the data above?
The attribute being measured is time.

b. What is the unit of measurement for the attribute?
Hours

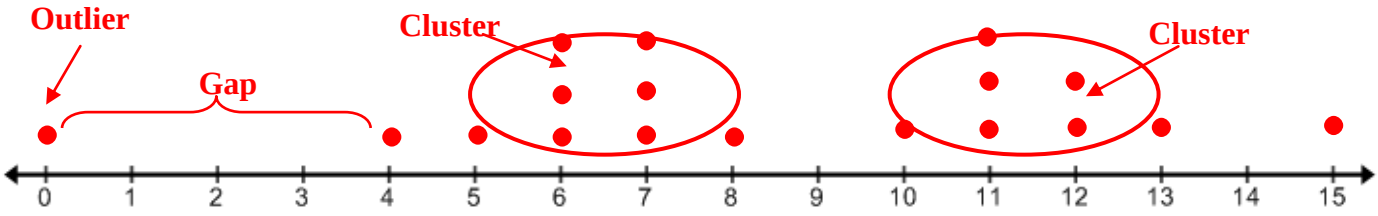
c. How could this data have been collected?
Most likely a survey

d. From how many people do we have data for?
16 people were surveyed.

e. Create a statistical question that could be answered using this data.
How much time did 6th grade students spend doing their homework last week?

The questions above directly correspond to standard 6.SP5 a and b. It is important for students to describe the nature of the attribute under investigation, including how it was measured and its unit of measurement.

f. Plot the data on a dot plot, be sure to label the units and give it an appropriate title.



Hours spent studying per week for 6th grade students

g. Mark any clusters, gaps, peaks or outliers that your observe on the plot.
See plot above. Be sure to mention that not all data sets have outliers.

h. What is the overall shape of the data?
This data does not have a common shape (symmetrical, skewed, uniform, etc.). It appears that the numbers are clustered between 5 and 8 hours and 10 and 13 hours.

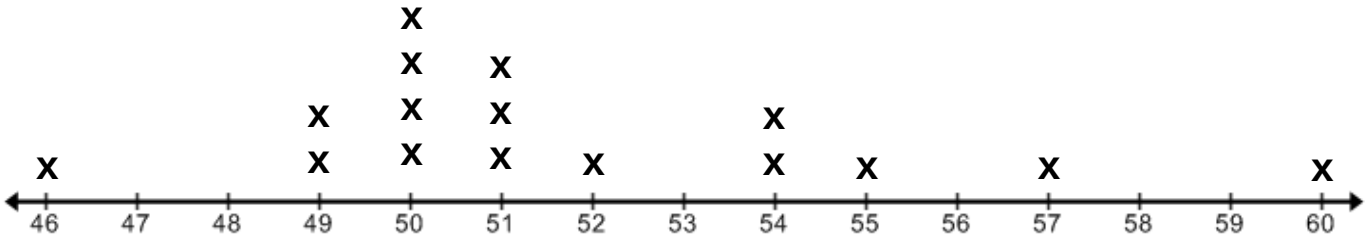
i. Use the dot plot to answer the question "What is the typical amount of time that a 6th grader spends on homework each week?"
The typical 6th grade students spends anywhere from 6 to 12 hours studying per week. For this question we are beginning to see the need for a measure of center. The variability in this plot makes it difficult to really answer this question without more in-depth investigation, which will happen in later sections.

j. If you increased the number of students surveyed how would the graph change?
There will be more dots on the graph. Be sure to discuss that most likely the scale will not change; based off of the data analyzed already a likely response will fall within the scale. However, there could be an extreme outlier that could change the scale.

k. This data represents 6th graders. Do you think the answer to our question would change if we included high school students? Why or why not?
The answer would most likely change because the amount of homework that students do in high school increases.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

9. Which statistical question best matches the line plot given below? (Note: A line plot is just like a dot plot but instead of using dots for each data entry it has an **x**.) Once you have chosen a question, determine an appropriate unit to measure the data related to this question.



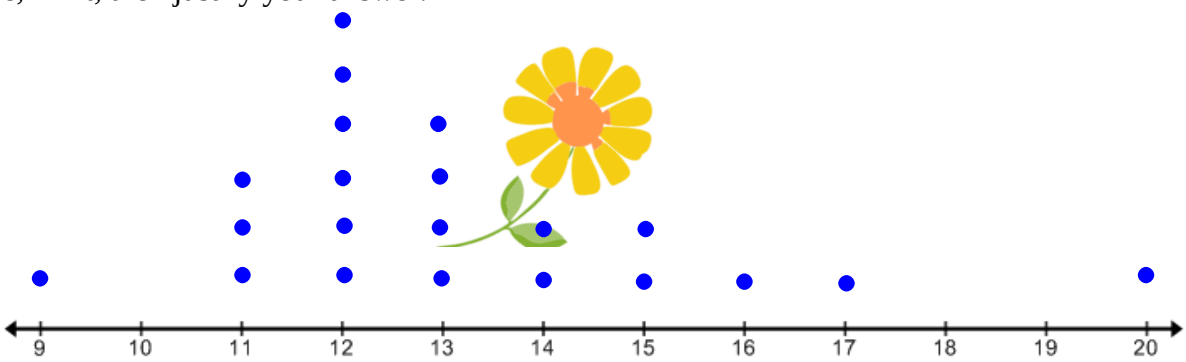
Speed of Cars on Highway in MPH

a) How many glasses of water do each member of our class drink a day?
b) How many minutes does it take to eat a Popsicle?
c) How fast do cars travel down the highway in our town? Units: Mile per hour.
d) How many gallons of water does a 6th grader use when showering?

10. Write a title for this plot based off of the statistical question that you chose in number 9.
See plot

11. Use the shape of the plot to answer your chosen statistical question.
Much of the data is clustered around 50 and 51 mph. This means that many people drive around 50 mph down the highway. The plot is skewed right, this means that people are more likely to drive faster than 50 mph than slower than 50 mph as shown by the tail of the graph.

12. Find, Fix, and Justify
The dot plot below is about the number of yellow flowers that are in each vase at a wedding. Penelope is interpreting the data and makes the following statements. For each statement given below, find her mistake, fix it, then justify your answer.



Number of Yellow Flowers in Vases

a. "The dot plot shows that 13 vases have 4 yellow flowers each."
Penelope thinks that each dot represents one yellow flower and the quantity on the number line is the number of vases. The correct statement is that 4 vases have 13 yellow flowers. On the plot, each dot represents the number of vases and each quantity on the number line is the number of flowers.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

b. "The data appears to be skewed left."
Penelope thinks that the data is skewed left because the dots on the plot are clustered and peak on the left side of the number line. The correct statement is that the data is skewed right because there is a "tail" on the right side of the number line.

c. "The dots above the number 12 are outliers because there are more of them compared to the other numbers."
Penelope thinks that an outlier is the number that occurs most often. The outlier in this data is 20 because it is an extreme value compared to the rest of the numbers in the data set.

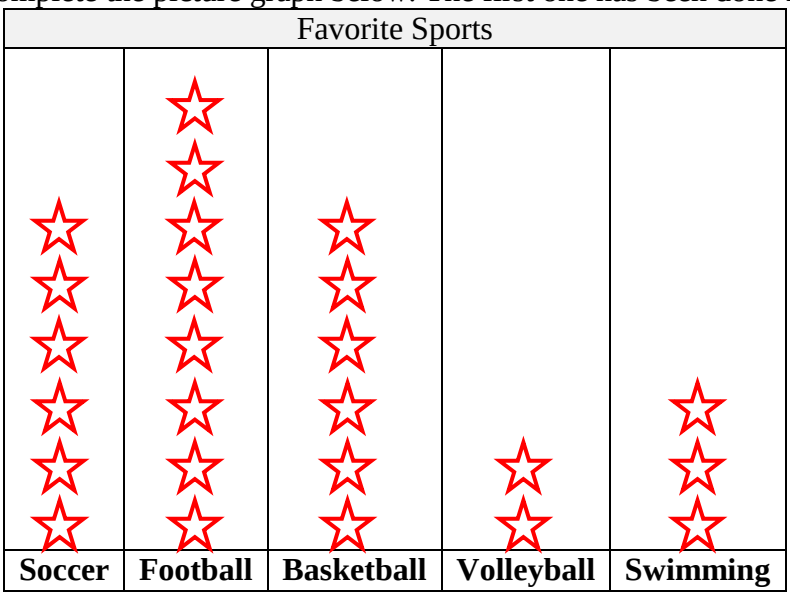
Be sure to allow time for students to collect data from class members for number 1 on their homework assignment. You may find it valuable to brainstorm ideas for statistical questions together as a class before asking students to create their own statistical question.

Spiral Review

The tally chart shows a survey of students' favorite sports.

Sport	Tally
Soccer	𝍸 |
Football	𝍸 |||
Basketball	𝍸 |
Volleyball	||
Swimming	|||

1. Use the chart to complete the picture graph below. The first one has been done for you.



2. How many students were surveyed? 25

3. The same number of students picked_____soccer_____ and _____basketball_____ as their favorite sport.
4. How many students picked volleyball as their favorite sport? 2

5. How many more students chose soccer than swimming as their favorite sport? 3

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.1b Homework: Creating and Analyzing Dot Plots

See teacher note on previous page.

1. Create a statistical question of your own that you can answer by surveying people in your class. Then ask at least 15 people in your class your question and record your data below.

a. Write your question here:____Answers will vary_____

b. Record your data in the table below.

Student	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Response															

c. What is the attribute that is being measured from your question?
Answers will vary.

d. What is the unit of measurement for your attribute?
Answers will vary.

e. Display your data in a dot plot. Be sure to label your number line and give it a title.
Answers will vary.

f. Mark any clusters, gaps, peaks or outliers that your observe on the plot.
See plot above.

g. What is the overall shape of the data?
Answers will vary.

h. Use the distribution of the data in your dot plot to answer your question.
Answers will vary.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. The table given shows how much time it took for 3rd graders to read the same book.

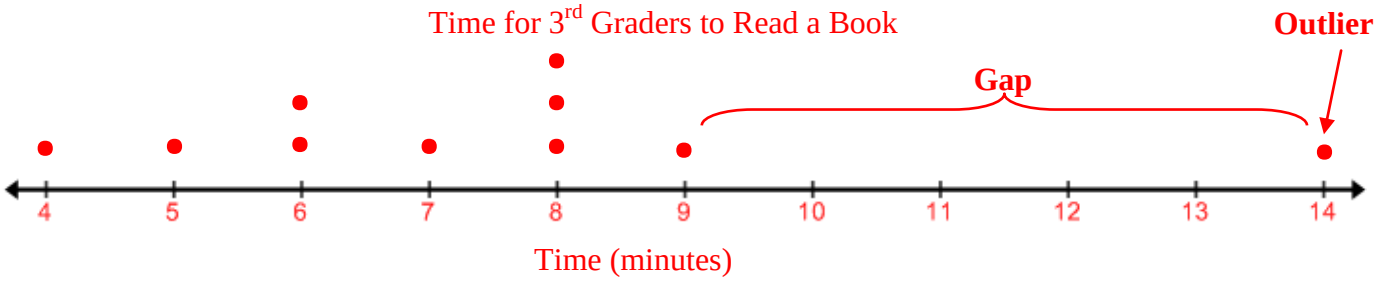
Student	Time (minutes)	Student	Time (minutes)
1.	4	6.	6
2.	7	7.	8
3.	8	8.	5
4.	14	9.	8
5.	6	10.	9

a. What attribute is being measured in the table.
Time

b. What is the unit of measurement for this attribute?
Minutes

c. How was this data measured?
Clock or stop watch.

d. Display the data in a dot plot, be sure to label the number line and give it a title.

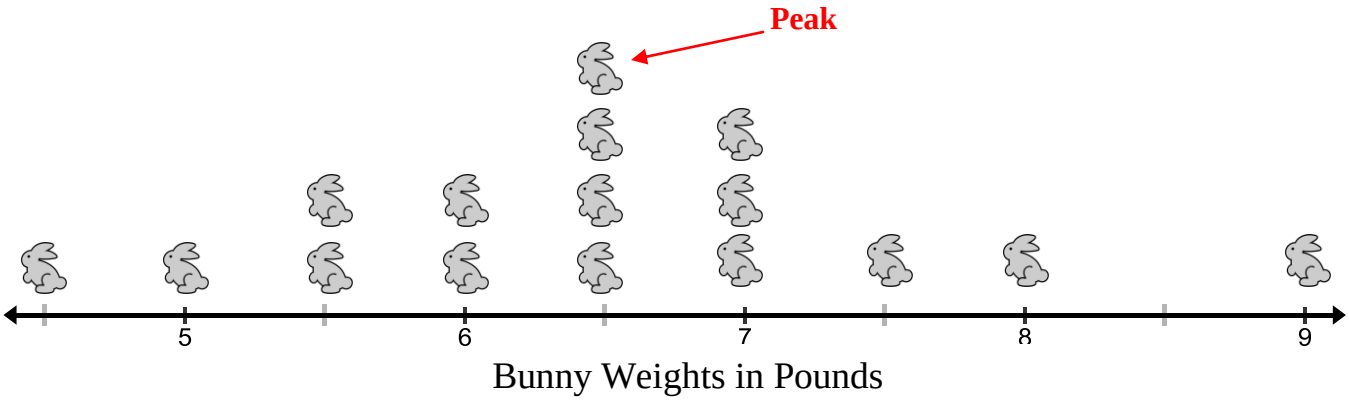


e. Mark any clusters, gaps, peaks or outliers that you observe on the plot.
See plot above.

f. What is the overall shape of the data?
The plot does not have a definitive shape.

g. What is the most typical time that it takes for a 3rd grader to read this book?
It takes between 6 and 8 minutes for a 3rd grader to read this book.

3. Use the graph below to answer the following questions.



a. How many observations are there?
16

b. What attribute is being measured in the dot plot?
The weight of a bunny

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

c. What is the unit of measurement for this attribute?
Pounds

d. How was the attribute measured?
With a scale

e. Mark any clusters, gaps, peaks or outliers that your observe on the plot.
See plot above. Note that the data point at 9 is not an outlier because there is not a significant gap between it and the other values.

f. What is the overall shape of the data?
The shape of the data is fairly symmetric. As the weight increases from 4.5 pounds to 6.5 pounds the number of bunnies increases. Then as the weight increases from 6.5 to 9 pounds the number of bunnies decreases.

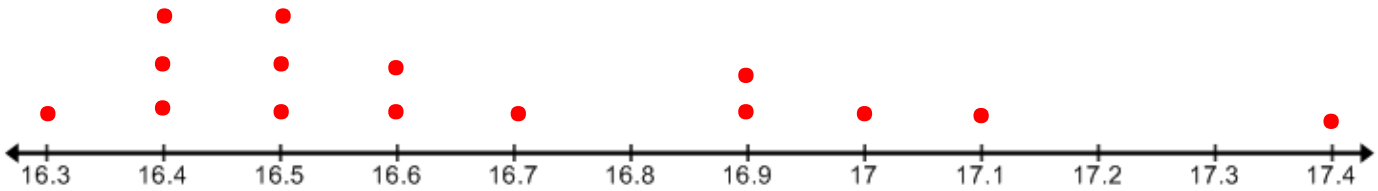
g. What is the typical weight for this kind of bunny?
Around 6.5 pounds

4. Use the data given below to answer the questions that follow.

100 Yard Dash Times (seconds)

16.3	16.7	16.4
16.4	16.9	16.4
17.1	17.4	16.6
16.5	16.9	17.0
16.5	16.6	16.5

a. Display the data above in a dot plot.



100 Yard Dash Times (seconds)

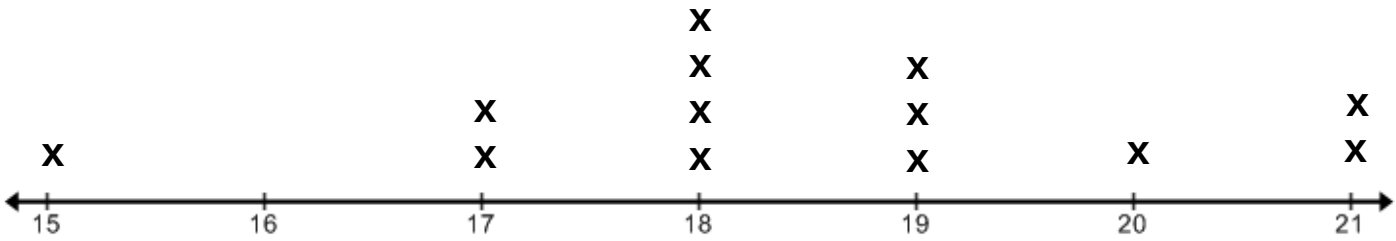
b. How many people ran in the race?
15 people

c. Circle the statement below that **does not** accurately reflect the data above.

- The data in the plot is skewed right.
- The attribute being measured is distance.
- The typical time for runners in this race was around 16.5 seconds.
- There is a cluster around the values of 16.9 seconds through 17.1 seconds.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. Which statistical question best matches the line plot given below? (Note: A line plot is just like a dot plot but instead of using dots for each data entry it has an x.) Once you have chosen a question determine an appropriate unit to measure the data related to this question.



Height of Flowers in Garden (inches)

a) How many siblings do students in my class have?
b) What is the typical height of flowers in my garden?; Units:Inches
c) How many hours of sleep do students in my class get each night?
d) How much money do people in my city spend on rent each month?

6. Write a title for this plot based off of the statistical question that you chose in number 5.
See plot

7. Use the shape of the plot to answer your chosen statistical question.
The distribution of the data is fairly symmetrical. Most flowers in the garden are between 18 and 19 inches tall.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.1c Class Activity: Creating and Analyzing Histograms

The high scores for a popular video game are shown below. Li wants to know what the typical high score for this video game is.

81	81	90	78	62	84	92	66	77	6	65	53	51	60
19	10	79	71	71	59	51	69	21	35	44	52	50	58
39	33	27	38	35	50	105	71	60	48	57	42	47	45
27	66	22	68	56	45	41	30	64	63	65	60	59	59

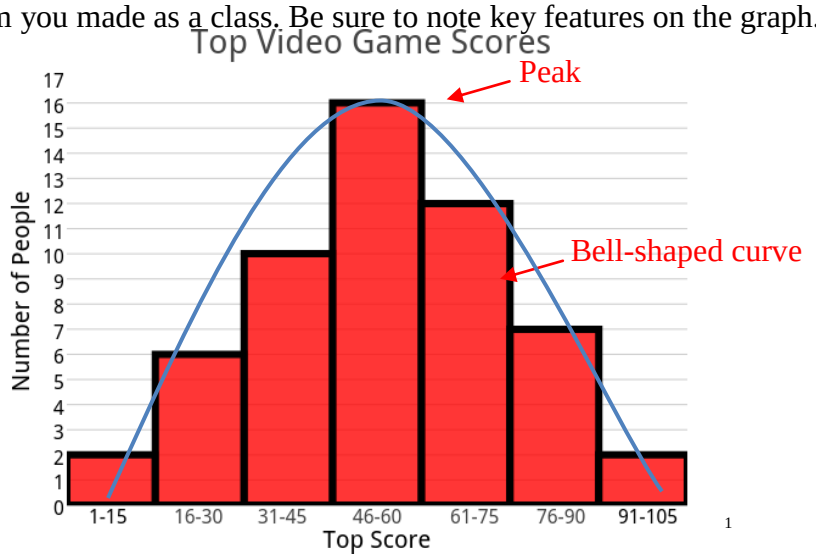
Li begins to make a dot plot for this data but soon realizes that a dot plot is not the best graphical display for this data. Discuss why a dot plot may not best for this data and see if you can come up with a better way to represent the data. The "range" of the data is very large. The smallest entry is 6 and the largest is 105. Unless you used intervals on the dot plot the number line would be very big. Also there is a very large "sample space", this means that there would be a lot of dots to plot. This is an example of choosing the appropriate tool or graphical display to represent your data.
Ask students for ideas of an alternative way to graph this data. They might suggest a bar graph or breaking the number line values into intervals. Talk to them about how a histogram is kind of like a bar graph but each bar represents more than one quantity. Begin by talking about possible intervals you could use for this data and make a frequency table/tally chart.

1. Copy the frequency table you made as a class below.

High Score Intervals	Frequency/Tally	Total
1 – 15	||	2
16 – 30	𝍸 |	6
31 – 45	𝍸 𝍸	10
46 – 60	𝍸 𝍸 𝍸 |	16
61 – 75	𝍸 𝍸 ||	12
76 – 90	𝍸 ||	7
91 – 105	||	2

A **frequency table**, or tally chart, groups data into intervals. The *frequency* is the number of values in an interval.

2. Copy the histogram you made as a class. Be sure to note key features on the graph.



1

1 evatipp. "Top Video Game Scores" Plotly Technologies Inc., https://plot.ly/79/~evatipp/ (Last Accessed: Feb 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

A **histogram** is a frequency display that uses bars to show the distribution of data in a set. The height of the bar shows the frequency of the values in that interval.

Students will be using the histogram model to interpret the shape of the data.

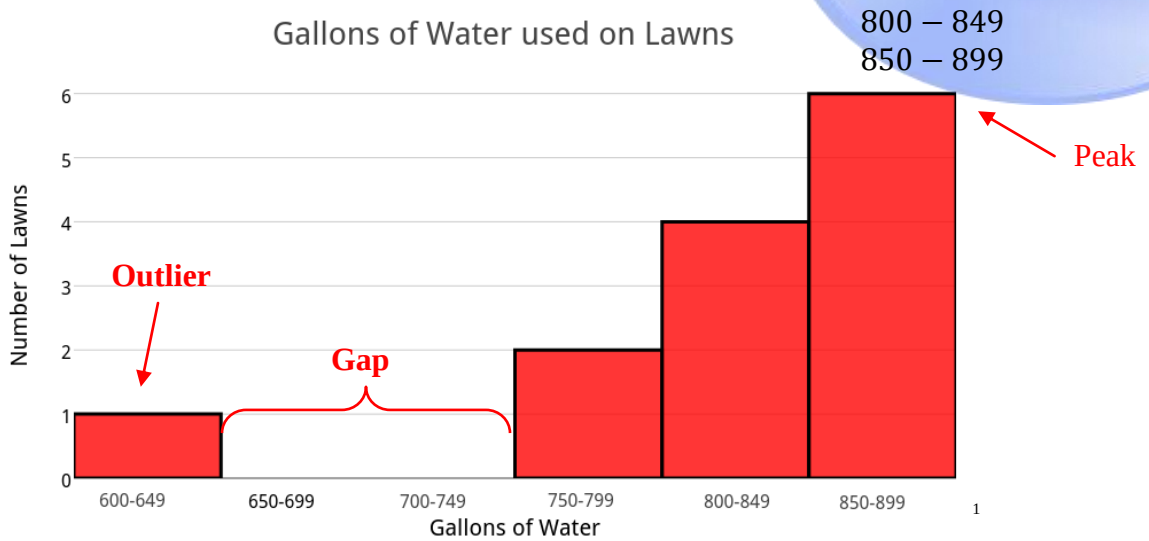
Lead a discussion about the distribution of the data in the histogram. You could discuss the following questions. Encourage students to mark their histogram with features that are addressed in the questions below.

1. How can you determine the total number of students that participated in the experiment from the histogram?
2. What does each bar in the histogram represent?
3. Is it possible to find the data point for a particular score on the histogram?
4. Did more people have a top score greater or less than 60?
5. Describe any peaks, clusters, or gaps in the data? Have students mark these on their histogram.
6. Describe any outliers in the data.
7. Describe the shape of the data.
8. Is the data skewed?
9. How does the data help Li to answer her question of "What is the typical top score for this video game?"

At this point, accept arguments that just include discussion about the shape of the data. Note that some students may begin to formalize ideas about *measures of center* (mean, median, and mode) and *measures of spread/variability* to answer the statistical question. Encourage them to describe their reasoning and thinking, which should be very informal at this point.

3. The graphic below shows amount of water used on a particular day for people's lawns in the same neighborhood.
   a. What is the attribute being measured?
      Water
   b. What is the unit of measurement for this attribute?
      Gallons
   c. How would you measure this data?
      Water meter
   d. Display the data in a histogram?

Gallons of Water Used	Number of Lawns
600 – 649	1
650 – 699	0
700 – 749	0
750 – 799	2
800 – 849	4
850 – 899	6



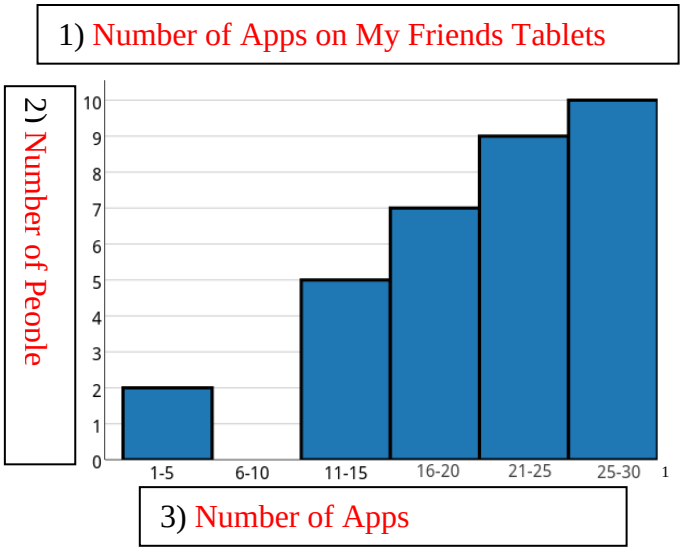
[1]

   e. Mark any clusters, gaps, peaks or outliers that your observe on the plot.
      See plot above.
   f. What is the overall shape of the data?
      The data is skewed left.
   g. What is the typical amount of water used on a lawn in this neighborhood?
      Most lawns use between 800 and 899 gallons of water each day.

1 evatipp. "Gallons of Water Used on Lawns" Plotly Technologies Inc., https://plot.ly/~evatipp/216/gallons-of-water-used-on-lawns/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. The graph below shows the number of Apps that my friends have on their electronic tablets.



[1]

It is a common mistake to create intervals that are not the same size. Make sure that you create intervals that are all the same size. For example if you wish to have intervals of 5 and you start at 0 then your intervals would be $0-4, 5-9, 10-14$, etc. In the problem above the intervals are of 5 but they start at 1. Note how the difference between the low number (and high number) for each pair of consecutive classes is the same (ie 6 – 1 = 5, 20 – 15 = 5, etc.)

a. Write in appropriate labels for boxes 1 through 3.
   See histogram above.

b. How many of my friends have between 21 and 25 apps on their tablets?
   9

c. How many of my friends have between 6 and 10 apps on their tablets?
   0

d. What does the shape of the distribution tell you?
   The data appears to be skewed left; with most of the friends having between 21 and 30 apps on their tablets.

e. How many of my friends were surveyed?
   33

f. Mandy has 17 apps on her tablet. She asks where her particular data value is on the graph. Is it possible to show her this?
   No, with a histogram you can only determine the interval in which a certain value falls.

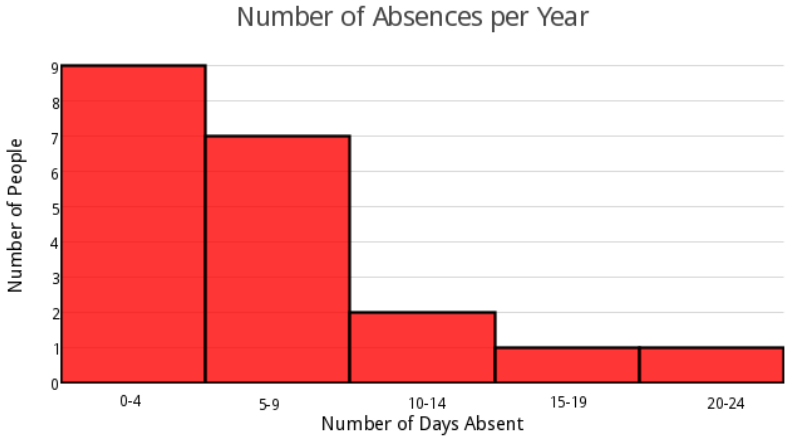
1 evatipp. "Number of Apps on Tablets" Plotly Technologies Inc., https://plot.ly/120/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. Marissa is interested in the number of days that students in her class are absent each year. On the last day of school, she asks several students in her class how many times they missed school for the entire year. Her data is given below.

6	12	0	2	16	5	4	0	5	5
1	7	4	21	3	13	2	3	9	5

a. Make a histogram of this data. (If needed, make a frequency table first.)



[1]

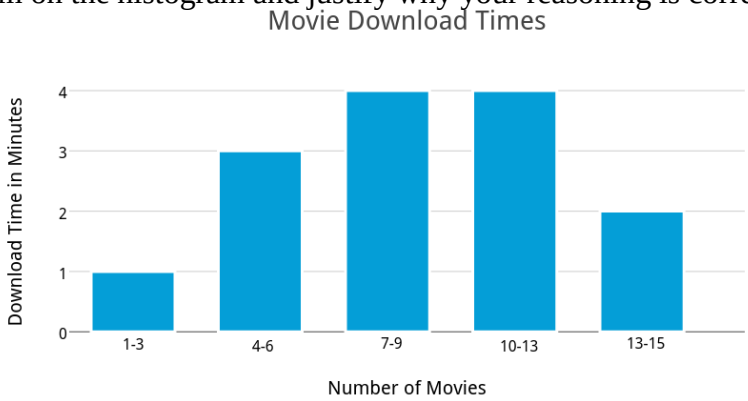
b. Circle the statement below that cannot be made about the data displayed in your histogram?

One student was absent at least 20 days and at most 24 days.
The data in the histogram is skewed left
Marissa surveyed 20 students.
Less than ¼ of the class was absent more than 10 days.

6. Find, Fix, and Justify.
The table given below shows the times that it takes to download different movies.

Time to Download a Movie in Minutes	1 – 3	4 – 6	7 – 9	10 – 12	13 – 15
Number of Movies	1	3	4	4	2

a. Carmen made a histogram of the data and has made many mistakes. Find at least two of her mistakes, fix them on the histogram and justify why your reasoning is correct.



[2]

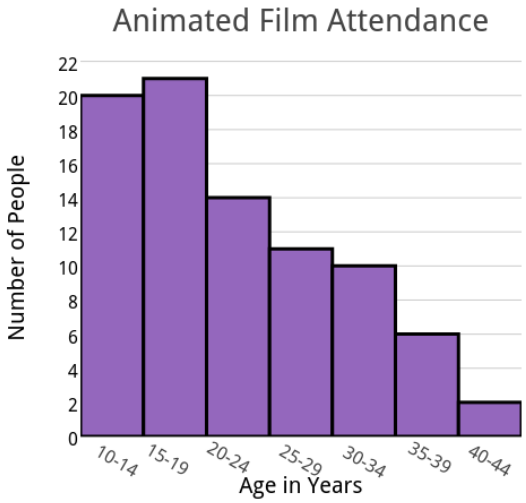
Carmen has her labels for the *x* and *y* axis mixed up. The bars on a histogram should not have a space between them they should be touching. The interval labeled 10-13 should be labeled 10-12.

b. Carmen has just purchased a movie online based on the data in the graph, what would you expect the typical download time to be for this movie?
Anywhere between 7 and 13 minutes

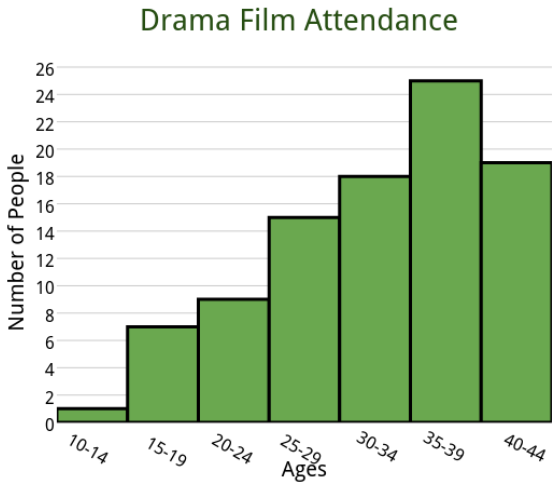
1 evatipp. "Number of Absences per year" Plotly Technologies Inc., https://plot.ly/150/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Movie Download Times" Plotly Technologies Inc., https://plot.ly/160/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

7. The two histograms given below show the ages of people watching two different types of movies at the same multi-screen movie theater.



1



2

a. Describe the shape of the distribution for each type of movie.
The data is skewed right for the animated film and skewed left for the drama film.

b. How many people attended the drama?
96 people

c. What percentage of the people attending the drama film were younger than 20?
Approximately 8.33%

d. What percentage of people attending the animated film were 30 or older?
Approximately 21.4%

e. Suppose you work for a marketing agency that is going to advertise these movies. Which type of movie would you advertise to a younger audience? Justify your answer.
You should market the animated movie to the younger audience because the data distribution of the animated movie suggests that people of a younger age are more likely to attend this movie rather than a drama.

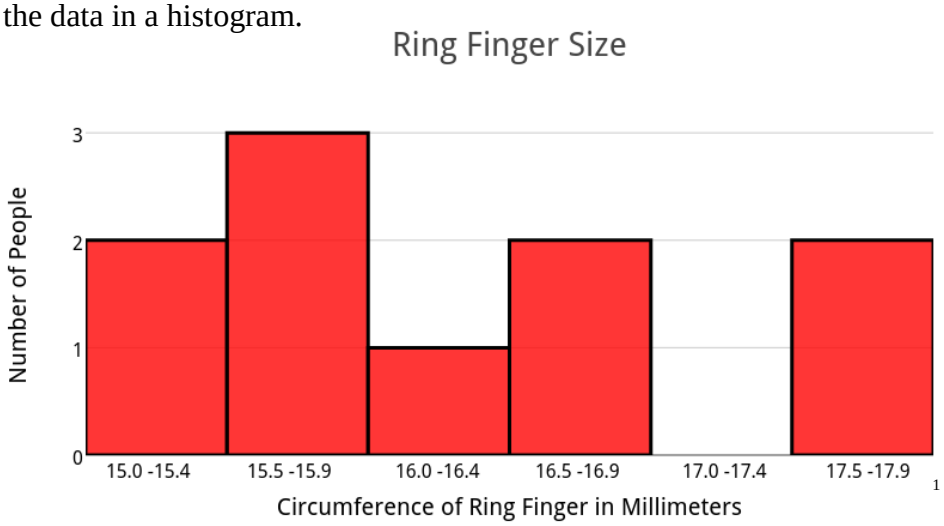
1 evatipp. "Animated Film Attendance" Plotly Technologies Inc., https://plot.ly/177/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Drama Film Attendance" Plotly Technologies Inc., https://plot.ly/171/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

8. The data in the table below shows the ring finger circumference of 10 people.

Ring Finger Circumference in Millimeters									
15.4	17.6	16.8	16.1	17.8	15.7	15.3	15.7	16.9	15.6

a. Display the data in a histogram.



[1]

a. Which interval includes the most data values?
15.5 – 15.9

b. How many people have a ring finger circumference less than 16.5 mm?
6 people

c. What percentage of people have a ring finger circumference greater than 17 mm and less than 18 mm?
Approximately 20%

d. Describe the shape of the distribution.
There is not a common shape for this data, suggesting that the ring sizes are either very random or more people need to be surveyed.

Be sure to allow time for students to collect data from class members for number 1 on their homework assignment. You may find it valuable to brainstorm ideas for statistical questions together as a class before asking students to create their own statistical question.

Spiral Review

1. Use long division to find each solution. Round your answer to the nearest tenth.
   a. $25 \div 4 =$
   6.3
   b. $515 \div 6 =$
   85.8
   c. $732 \div 20 =$
   36.6

2. How would your share…
   a. 12 pieces of licorice evenly between 5 friends?
   Each friend will get 2.4 pieces of licorice.

   b. 58 jelly beans evenly between 7 friends?
   Each friend will get about 8.3 jelly beans.

1 evatipp. "Ring Finger Size" Plotly Technologies Inc., https://plot.ly/197/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.1c Homework: Creating and Analyzing Histograms

1. Create a statistical question of your own that you can answer by surveying people in your class. Then ask at least 15 people in your class your question and record your data below.

   a. Write your question here:____Answers will vary_____

   b. Record your data in the table below.

Student	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Response															

   c. What is the attribute that is being measured from your question?

   d. What is the unit of measurement for your attribute?

   e. Display your data in a histogram, be sure to label your number line and give it a title.

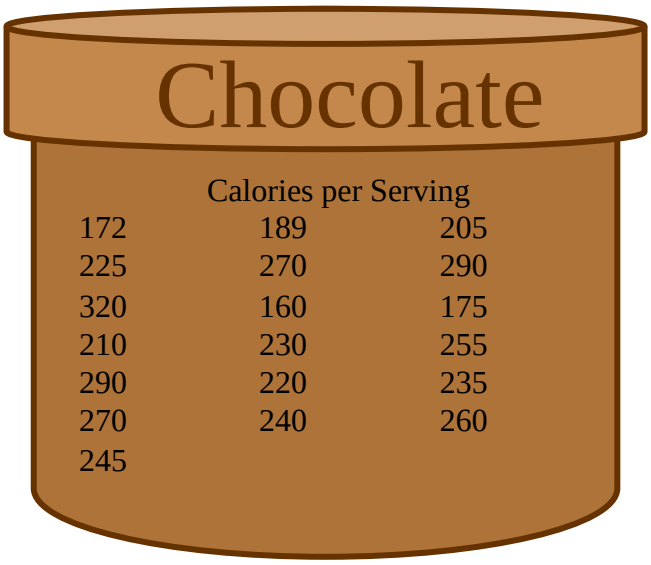
   f. Mark any clusters, gaps, peaks or outliers that you observe on the histogram.

   g. What is the overall shape of the data?

   h. Use the distribution of the data in your histogram to answer your question.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. A nutritionist is interested in the typical number of calories in a serving of chocolate ice cream. He has collected data for calories per serving for several different brands of ice cream. His data is shown below.

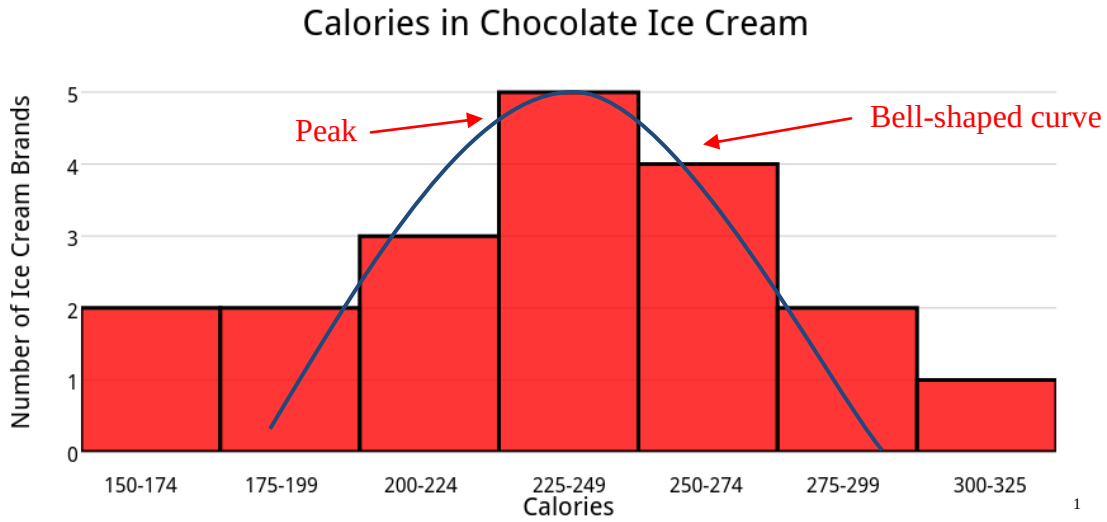


a. What attribute is the nutritionist measuring?
Calories

b. Make a frequency table of the data.

Calorie Intervals	Frequency	Total
150 – 174	II	2
175 – 199	II	2
200 – 224	III	3
225 – 249	~~IIII~~	5
250 – 274	IIII	4
275 – 299	II	2
300 – 325	I	1

c. Make a histogram of the data.



[1]

d. Mark any clusters, gaps, peaks or outliers that your observe on the histogram.
See plot above.

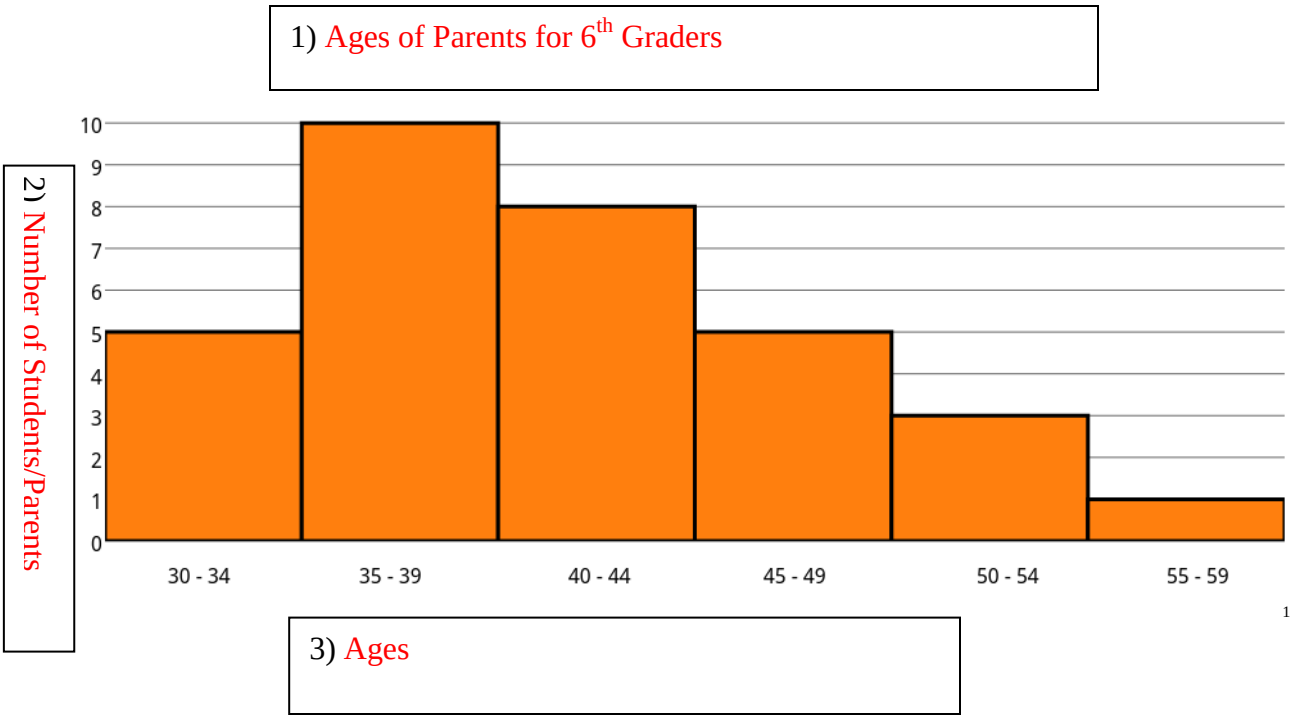
e. What is the overall shape of the data?
The data is fairly symmetrical. As the calories increase, the number of brands of ice cream increases until the data peaks between 225 – 249 calories. After the data peaks, the number of brands of ice cream begins to decrease as the calories continue to increase.

f. What is the typical number of calories in a serving of chocolate ice cream?
Between 225 – 249 *calories*

1 evatipp. "Calories in Chocolate Ice Cream" Plotly Technologies Inc., https://plot.ly/231/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. The histogram below shows the ages of parents for students in Carlotta's 6th grade class.



[1]

a. Write in appropriate labels for boxes 1 through 3.
See histogram above.

b. How many students have a parent between the ages of 45 and 49?
5

c. How many of students have a parent older than 49?
4

d. What does the shape of the distribution tell you?
The data appears to be skewed right; with most of the students having a parent between 35 and 44 years old.

e. How many students were surveyed?
32

f. Carlotta wants to know how many parents are younger than 37, is it possible to determine this from the histogram? Explain.
No, since the intervals count by 5's and begin at 30 we can only determine that there are at least 5 parents younger than 37. But we cannot determine how many parents in the interval of 35-39 are younger than 37.

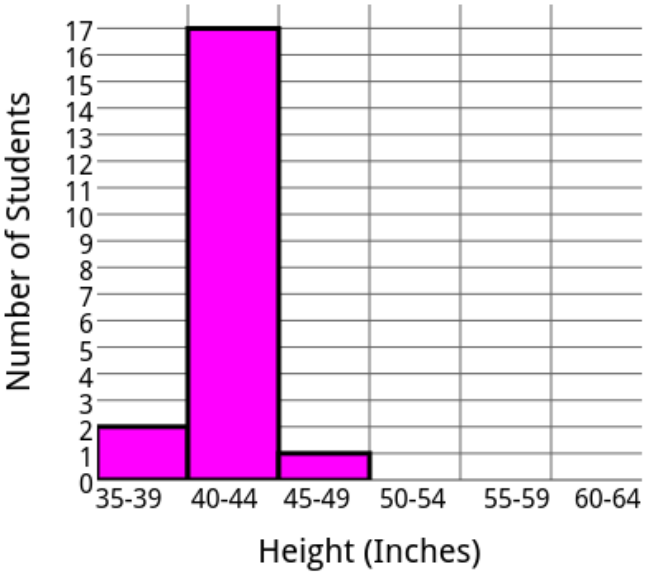
g. Create your own question that could be answered using the histogram. Then answer your question.
Answers will vary.

1 evatipp. "Ages of Parents for 6th Graders" Plotly Technologies Inc., https://plot.ly/238/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

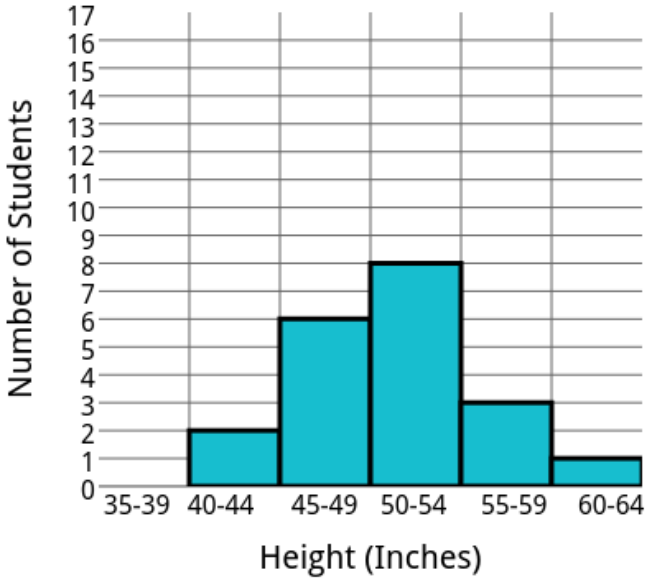
4. Ricardo and Sven are interested in the heights of students at their school. At the end of the day Ricardo stands at the front door of the school and collects the heights of 20 students leaving the school. Sven collects the heights of 20 students leaving the kindergarten classroom. They each make a histogram of the data they collected.

Title: Height of Kindergarten Students



[1]

Title: Height of Students at School



[2]

a. Determine which histogram belongs to each boy by giving each histogram a title.
See plot above (Sven is the plot on the left because the heights are smaller)

b. What attribute is being measured in the histograms.
Height

c. What unit of measurement for this attribute?
Inches

d. How was this data measured?
Most likely with a tape measure or ruler

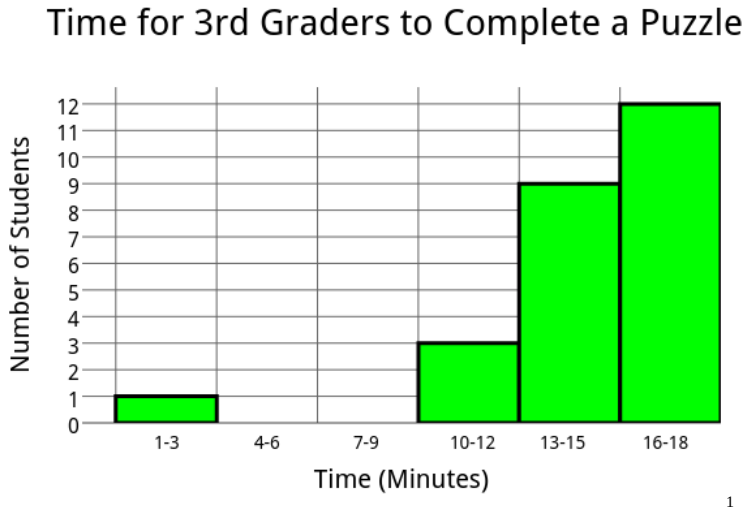
e. What percent of kindergartener's are less than 40 inches tall?
10%

f. What percent of students are at least 50 inches tall?
60%

1 evatipp. "Height of Kindergarten Students" Plotly Technologies Inc., https://plot.ly/262/~evatipp (Last Accessed: March 2016)
2 evatipp. "Height of Students at School" Plotly Technologies Inc., https://plot.ly/268/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. Use the histogram below to answers the questions that follow.



[1]

a. What attribute is being in the histogram, what is the unit of measurement for this attribute, and how was this data measured?
Time; minutes; timer or clock.

b. Circle all the statements below that are true about the distribution of the data in the histogram.

i. 25 third graders were surveyed.

ii. The data is skewed left.

iii. It took between 16 and 18 students 12 minutes to complete the puzzle.

iv. It took one student between one and three minutes to complete the puzzle. This student is an outlier.

v. It takes the typical 3rd graders between 13 and 18 minutes to complete the puzzle.

c. Suppose you gave a class of 6th graders the same puzzle and recorded how much time it took each person to complete the puzzle. Describe how the 6th grade histogram would be different than the 3rd grade histogram.
This histogram would probably be more symmetrical or may even be skewed right because it is likely that a 6th grader would be able to complete the puzzle more quickly.

1 evatipp. "Time for 3rd Graders to Complete a Puzzle" Plotly Technologies Inc., https://plot.ly/273/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.1d Self Assessment: Section 3.1

Consider the following skills/concepts. Rate your comfort level with each skill/concept by checking the box that best describes your progress in mastering each skill/concept. Corresponding sample problems, referenced in brackets, can be found on the following page.

Skill/Concept	Minimal Understanding 1	Partial Understanding 2	Sufficient Mastery 3	Substantial Mastery 4
1. Recognize a statistical question as one that anticipates variability in the data.	I can recognize a statistical question in some contexts.	I can recognize a statistical question in a variety of contexts.	I can recognize a statistical question in a variety of contexts and create statistical questions on my own.	I can recognize a statistical question in a variety of contexts and create statistical questions on my own. I can rephrase questions that are not statistical into statistical questions.
2. Identify if the set of data used to answer a statistical question is numerical or categorical.	I can identify if a statistical question will generate numerical data. I struggle to identify questions that will generate categorical data.	I can identify if a statistical question will generate numerical or categorical data.	I can identify if a statistical question will generate numerical or categorical data. I can suggest possible data values for a statistical question.	I can identify if a statistical question will generate numerical or categorical data. I can suggest possible data values for a statistical question. I can create questions that generate categorical and statistical questions.
3. Display numerical data in a dot plot, use key terms to describe its shape, and interpret the distribution of data.	I can display data in a dot plot but often don't know how to label it appropriately. I don't know how to describe the shape or interpret the distribution of data.	I can display data in a dot plot that is accurately labeled. I can identify some of the terms used to describe its shape. I struggle to interpret the distribution of data.	I can display data in a dot plot that is accurately labeled. I can identify and use terms such as peaks, clusters, gaps, skewed, symmetrical and outliers to describe the shape. If the data is skewed I cannot define its direction.	I can display data in a dot plot that is accurately labeled. I can identify and use terms such as peaks, clusters, gaps, skewed, symmetrical and outliers to describe the shape and interpret the distribution of data.
4. Display numerical data in a histogram, use key terms to describe its shape, and interpret the distribution of data.	I can display data in a histogram but often don't know how to label it appropriately. I don't know how to describe the shape or interpret the distribution of data.	I can display data in a histogram that is accurately labeled. I can identify some of the terms used to describe its shape. I struggle to interpret the distribution of data.	I can display data in a histogram that is accurately labeled. I can identify and use terms such as peaks, clusters, gaps, skewed, symmetrical and outliers to describe the shape. If the data is skewed I cannot define its direction.	I can display data in a histogram that is accurately labeled. I can identify and use terms such as peaks, clusters, gaps, skewed, symmetrical and outliers to describe the shape and interpret the distribution of data.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Skill/Concept	Minimal Understanding 1	Partial Understanding 2	Sufficient Mastery 3	Substantial Mastery 4
5. Summarize a numerical data set in relation to its context by reporting on the number of observations, identifying how the data was measured and its units of measurement.	When summarizing a data distribution I can identify the number of observations.	When summarizing a data distribution I can identify the number of observations and offer ideas on how the data was measured	When summarizing a data distribution I can identify the number of observations, offer ideas on how the data was measured, and identify appropriate units of measurement.	When summarizing a data distribution I can identify the number of observations, offer ideas on how the data was measured, and identify appropriate units of measurement. I look back and account for the context of a data distribution.

Sample Problems for Section 4.1

Square brackets indicate which skill/concept the problem (or parts of the problem) aligns to.

1. Evan and Somer are selling cookies to the kids at their school during lunchtime for their school band fundraiser. They have a variety of cookies that each cost a different amount. The questions below were discussed as they worked on the fundraiser. [1] [2]

   For each question decide if it is a statistical question and justify your decision. Then state whether the data collected to answer the question will be categorical or numerical data.

   a. What is the price of cookies for the fundraiser?
   Statistical; the cost for a cookie varies depending on the type of cookie; numerical

   b. How long did it take us to sell out of cookies on Tuesday?
   Not statistical; there was an exact amount of time that it took for them to sell out of cookies on Tuesday; numerical

   c. How many cookies were sold each day of the fundraiser?
   Statistical, there were most likely a different number of cookies sold each day of the fundraiser; numerical

   d. What is Evan's favorite cookie?
   Not statistical; there is only one answer for this; categorical

   For each of the questions above that are not statistical, rephrase them so that they are a statistical question. Write them below. List three possible data values for each question.

   e. Question 2: How long did it take us to sell out of cookies each day of the fundraiser?
   Possible data values: 15 min, 30 min, 50 min

   f. Question 4: What is the favorite cookie for kids at our school?
   Possible data values: Chocolate Chip, Snicker Doodle, Peanut Butter

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Create two statistical questions of your own, one categorical and one numerical, which are related to the cookie fundraiser.

g. Numerical: Answers will vary, sample answers are given below.
How long will it take us to make each batch of cookies?
How many people are standing in line each day of the fundraiser?
How many cookies did we sell to teachers each day of the fundraiser?
How much money did we make each day of the fundraiser?

h. Categorical: Answers will vary, sample answers are given below.
What the favorite cookie for teachers at our school?
What class of students bought the most cookies each day of our fundraiser?

2. Beckham is interested in the hand span, the length from the tip of the pinky finger to the tip of the thumb, for people in his class. He asks the statistical question, "How long is the hand span for people in my class?" He decides to conduct a survey to answer his question. His data is shown below. [3][5]

Length of hand span in inches: 6, 6.5, 9, 7.5, 7.5, 7, 6.5, 7, 7, 9.5, 6, 8, 5.5, 7, 6.5

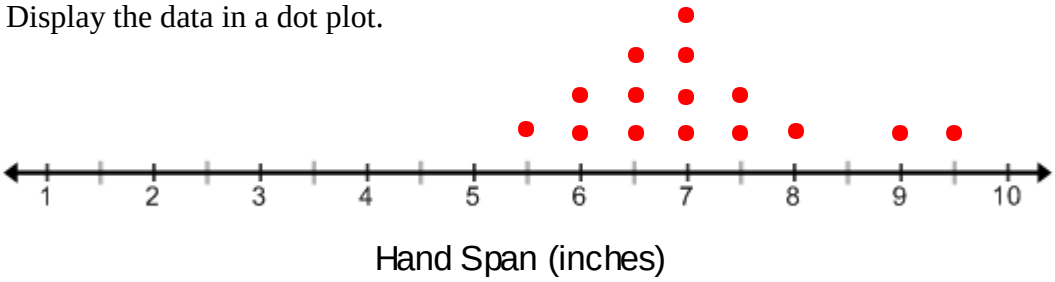
a. What is the attribute that is being measured?
Length

b. What is the unit of measurement for this data?
Inches

c. What tools do you think Beckham used to obtain these measurements?
Measuring tape or ruler

d. For how many people did Beckham measure the hand span?
15

e. Display the data in a dot plot.



f. What is the shape of the data? Be sure to use key terms in your description.
The shape is fairly symmetrical and peaks around 7 inches. Most of the data is clustered between 6 and 7.5 inches.

g. Use the shape of the data to interpret the distribution of data and answer Beckham's question.
Based off of the shape of the data most of the people in Beckham's class have a hand span of about 7 inches.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. The time that it takes for several 6th graders to eat their lunch is shown below. [4][5]

How long does it take you to eat your lunch? (minutes)					
26	24	21	25	19	26
26	24	16	13	15	21
27	27	14	27	22	20
24	4	25	29	27	22
20	17	28	22	26	25

a. What statistical question could be asked that relates to the data.
How long does it take 6th graders to eat their lunch?

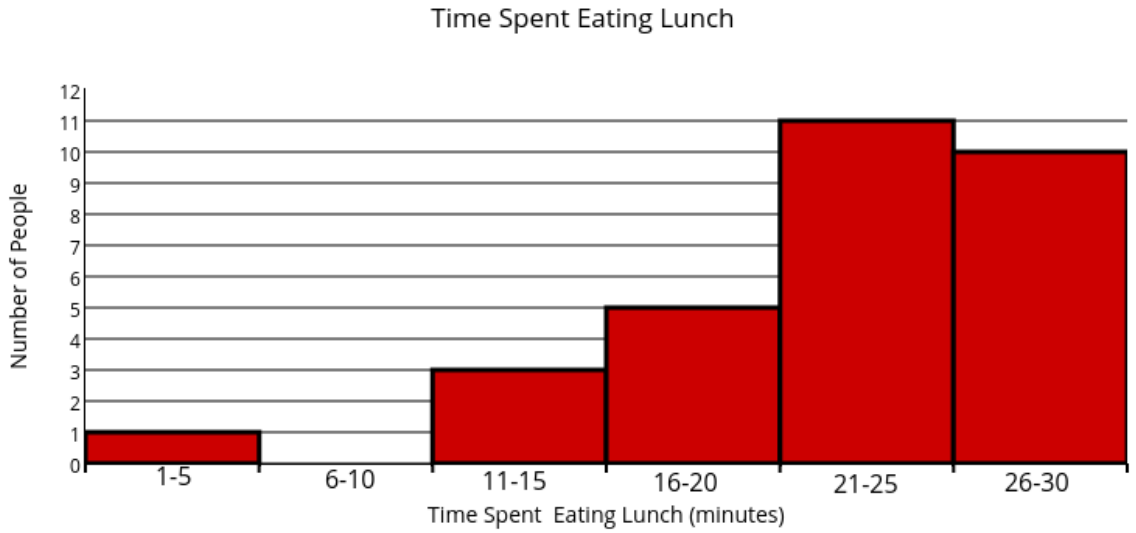
b. What is the attribute that is being measured?
Time

c. What is the unit of measurement for this data?
Minutes

d. What tools do you think Beckham used to obtain these measurements?
Clock or stopwatch

e. How many people were surveyed?
30

f. Display the data in a histogram.



[1]

g. How many people spent 16-20 minute eating lunch?
5 people

1 evatipp. "Time Spent Easting Lunch" Plotly Technologies Inc., https://plot.ly/534/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

h. What percentage of people spent 15 minutes or less eating lunch?
About 13.3 percent

i. What is the shape of the data? Be sure to use key terms in your description.
The data appears to be skewed left with most of the data clustered between 16 to 30 minutes. The data values peak around 21-25 minutes. It appears as though the data value between 1 and 5 may be an outlier.

j. Use the shape of the data to interpret the distribution of data and answer your statistical question.
Based off of the shape of the data most students take between 21 to 30 minutes to eat their lunch.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Section 4.2: Meaures of Center

Section Overview:

In this section, the focus is on finding and interpreting measures of center. To be useful, measures of center must be characterized by numerical descriptions. The first lesson has students begin to understand the mean is a measure of center that levels out the data in the sense of a unit rate. In this "leveling out" sense the mean, (often referred to as the "arithmetic average" or simply "average") is often thought of as a fair share. They abstract the act of gathering all the data and equally redistributing it to summing all the values in the data set and then dividing by the total number of items. In addition to looking at the mean as a fair share, students will experiment with the mean as a balancing point of the distribution. This prepares students to better understand mean absolute deviation when they analyze variability in the next section. Students practice finding the mean from several different data displays and begin to recognize that the mean can be affected by extreme values. Next, students learn about the median and mode and how to find them from a given set of data. They soon realize that they can choose a measure of center that best represents the data by analyzing the shape of the data distribution and the context in which the data is gathered.

Concepts and Skills to Master:

By the end of this section, students should be able to:

1. Understand the mean as a "leveling out" of data or a fair share. Also understand how the mean can be interpreted as the "balancing point" for the data values.
2. Find the mean from a set of data and interpret its meaning for a given context.
3. Find the median or mode from a set of data and interpret their meaning for a given context.
4. Determine the best measure of center for a given data set and justify why it is the best measure of center.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2a Class Activity: How Much is a Handful?

Directions: Grab a handful of linking cubes from your teacher. Count the number of cubes in your hand and write the number on a sticky note.

Provide a large quantity of linking cubes for your students to grab. Ask students to count the number of cubes in their hand and write their number on a sticky note. Review how to make a dot plot and/or histogram by using the sticky notes on the board. After making the dot plot or histogram, review with students that the data display helps us to understand the shape of the data.

Recall that the distribution of a data set can be described by its *center*, *spread (variability)*, and *shape*. In this section we are going to investigate how to describe the **center** of a data set.

1. What do you think is the center of the class's linking cube data?

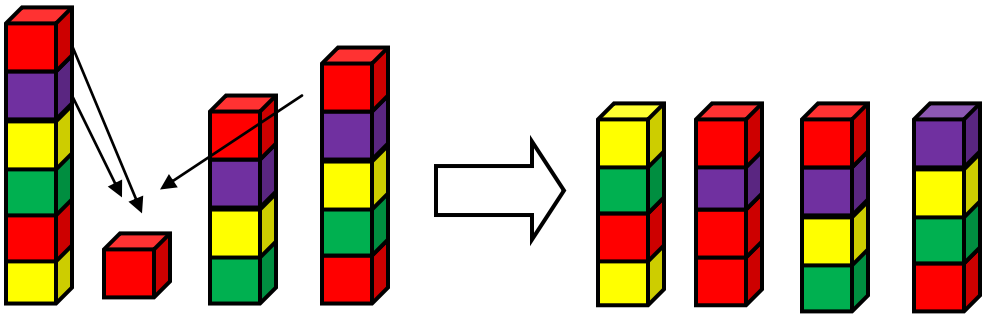
This question is meant to get students thinking about obtaining a number that can characterize a single feature of this data distribution, in this case "the center". It may be useful to ask students what they think is meant by "center"? Maybe "typical", "middle", "average", "occurring most of the time"? It may be natural for students to suggest that the mode or median describe the center of this data. That is great; however, they will not be discussed in this lesson.

2. What if we wanted everyone in the class to have a fair share of cubes, how many cubes would there be per person?

Discuss this question as a class. In previous grades, students have solved problems where they have been asked to gather or sum measurement amounts in order to redistribute items or amounts equally. See **5.MD.2** These questions are meant to get students to understand the *mean* as a "fair share" or "leveling out" in the sense of a unit rate. As your discussion progresses asks students to clarify what is meant by the phrases, *fair share* or *leveling out*.

Work in your group to answer the question above, be ready to discuss your ideas with the class.

It might be more manageable to have students work in groups. Have each group answer the questions above for their own group by use their linking cubes; several strategies might be used. One strategy is to have each student link their own cubes into a rod and then line the rods up in a row. They can distribute cubes from the longer rods to the shorter rods until everyone has the same amount of cubes per rod (leveling out the data).



Another strategy is to combine everyone's cubes together and then redistribute the cubes so everyone has the same amount. This strategy can be directly correlated to the algorithm for finding the mean, i.e., finding the sum of all the numbers (combining the cubes) and then dividing by the number of people in the group (splitting the cubes evenly between everyone in the group). It might be helpful to relate this redistribution to unit rate by talking about *the number of cubes per person.*

Some groups may not be able to redistribute their cubes "evenly", i.e. they may have some cubes left over. Ask them what they should do with the remaining cubes. Discuss how they will have to split the remaining cubes between everyone in the group, meaning that each member will receive a "partial" cube. Have them determine exactly how much of the cube each member gets. Have students share their strategies and ideas with the class and to write the answers to the questions above.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

This number that describes the data as a "fair share" or a "leveling out" of the data is called the arithmetic mean or just **mean**. It is often referred to as the **average**. It is a single number that describes the center of the data.

3. At the arcade it costs one token to play a game of Skeeball. The table below shows the number of tokens that a group of 6 friends brings with them to the arcade.

Tokens					
Kira	Tom	Jade	Braden	Nan	Lonny
5	6	3	3	2	5

They decide that they want each person to have the same number of turns playing the game. Determine the number of times that each person can play the game. In other words, what is the **mean** or **average** number of times that each person can play the game? Use your linking cubes to represent the tokens if needed.

At this point begin to focus on the strategy of combining all the tokens and then redistributing them so each person in the group gets the same number of tokens. Ask students to show this mathematically.

Combine all the tokens together → $5 + 6 + 3 + 3 + 2 + 5 = 24$

Split the tokens "evenly" between all the members of the group → $\frac{24}{6} = 4$

Each person will get 4 tokens.

4. A family of five goes out to lunch. The numbers below show the cost of each item that a person ordered for lunch. They decide to split the bill so that everyone pays the same amount. In other words, what is the average cost per person for lunch?

$8.25 $8.70 $7.50 $6.95 $9.35

If students have not yet arrived at the algorithm for calculating the mean, talk them through it. To find the average cost or mean:

Combine all the costs together → $8.25 + 8.70 + 7.50 + 6.95 + 9.35 = 40.75$

Spilt the total cost in equal shares for each member of the family → $\frac{40.75}{5} = 8.15$

Each person will pay $8.15.

Discuss how to find the mean for any given set of data. As you discuss how to find the mean for any given set of data, students are abstracting the solving process. They must understand that for any data set they can sum the data and divide by the number of data values to get the mean. They come to this understanding by reasoning through the problems above and repeatedly creating "fair shares".

The *mean* or *average* is a *measure of center* for a data set. The mean is found by taking the sum of all the data values and then dividing by the number of data values.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. Find the average for each set of data.

a. Time to run a mile in minutes:
9.5, 8.7, 9.3, 10, 10.4, 7.5, 8.4, 9.1
Average: 9.1125 minutes

b. Number of movies attended this summer:
4, 3, 9, 5, 4, 0, 7, 5, 6, 2, 3, 4
Average: approximately 4.3 movies

c.

Number Of Posts On A Social Media Site Per Week	
JJ 25	Kaci 48
Charlie 34	Mark 38
Andrew 56	Nicole 52
Quinn 34	Seth 46
Cindy 45	Roxy 30

The average number of posts per week is 40.8.

d.

Number of Pets Owned	
Zara	III
Alina	I
Todd	~~IIII~~
John	~~IIII~~ I
Aarav	III

Average: 3.6 pets

A common mistake is for students to leave out the data value of 0. For example, in part b above, students might argue that the 0 will not change the sum of all the data value so it can be discarded. However, it still represents the number of movies that 1 person saw so it must be counted as a data value; i.e. you must divide the sum by 12 and not 11. You may want to discuss that of course it is not really possible to have 3.6 pets. Rather we would say that the average number of pets owned is between 3 and 4. Many of these numerical measures do not make complete sense in relation to the context but they still give us an understanding of this measure of center.

Spiral Review

1. Out of a class of 25 people, 14 of them own a dog. What percentage of the class owns a dog? Round your answer to the nearest percent.
56%

2. 6 out of 7 days a week Lola works out. What percentage of the week does Lola work out? Round your answer to the nearest percent.
86%

3. Carson got 16 out of 15 problems correct on a quiz. What percentage of the problems did he get correct? How could this be possible? Round your answer to the nearest percent.
107%, there may have been bonus or extra credit problems on the quiz.

4. Three friends earned $21.75 doing jobs for a neighbor. How much money did each person earn individually if they divide the total earnings equally?
$7.25

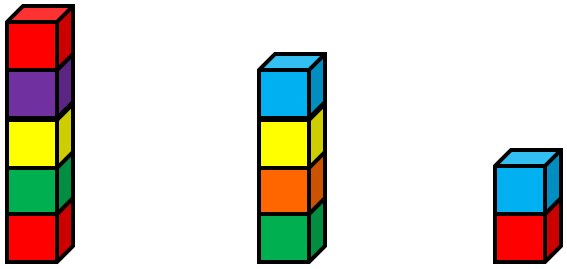
5. How many $\frac{1}{3}$ cup servings are in 3 cups of nuts?
9

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2a Homework: Mean as a Fair Share

Find, Fix, and Justify

Tina and Arnold are trying to find the average number of blocks per stack for the linking cubes shown below. They have both made a mistake in their reasoning. Find their mistake and fix it, then justify why your thinking is correct.



1. Tina knows that the she can think of an average or mean as every group having a fair share or the same number of blocks. She begins to rearrange the blocks and concludes that there are too many cubes for 3 groups of 3 and not enough cubes for 3 groups of 4, so there must not be a mean.
   Tina has made the mistake of thinking that the average must be a whole number. She can begin by making 3 groups of 3 cubes each. Then she needs to split the remaining 2 cubes between the 3 groups. This gives an additional $2/3$ of a cube to each group. Thus the average number of cubes per group is 3 and $2/3$ of a cube.

2. Arnold knows that he can find the average by adding up the total number of blocks and then divide by the number of groups. When he does this he gets an answer of $3\frac{2}{3}$. He claims that this does not seem like the right answer because the original set of numbers does not include $3\frac{2}{3}$.
   Arnold has made the mistake of thinking that the average must be one of the original set of numbers. The mean may or may not be one of the original set of numbers. Arnold's number is the correct mean because $5 + 4 + 2 = 11$ and $\frac{11}{3} = 3\frac{2}{3}$.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2b Class Activity: Finding the Mean

Recall that the mean or the average is a measure of center that summarizes a data set.
Students will compare their numerical measure of center in this section to the conclusion that they made about this data in earlier sections when they analyzed its shape. As they do this, they should begin to understand that by looking at the shape, center, and spread together for every data set they can begin to form a good analysis and draw conclusions about the data distribution. As the chapter progresses, through repeated reasoning, students will become more familiar with each of these statistical measures and should get better at using them in conjunction with one another.

The *mean* or *average* is a *measure of center* for a data set; (a number that best describes the center of a data set). It found by taking the sum of the data values and then dividing by the number of data values.

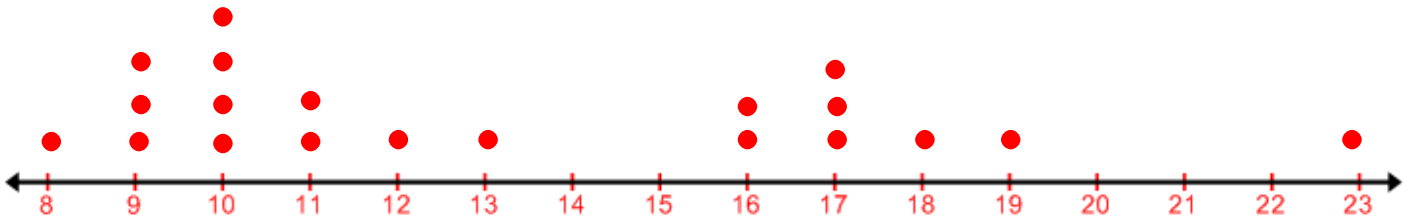
Note that understanding the mean as a fair share as discussed in the previous section and mean as a numerical measure of center as discussed in this section are two different conceptual ideas.
The most recent test scores for Mr. Petrov's science class are shown in the table.

Mr. Petrov's Test Scores					
79	91	72	89	84	77
71	82	80	76	91	83
80	79	82	81	78	78
82	72	86	81	80	64
83	79	82	84	81	81
92	80	80	87	83	81

Have students use a calculator to find the mean.

1. Find the mean test score for the class.
   Approximately 80.86
2. What does this tell you about how well the students in Mr. Petrov's class did on the test? Compare this answer with the conclusions you made about the dot plot that you made for this data in **Section 4.1b**.
   In section 4.1b, the dot plot suggested that most of the scores were around 80. This coincides with the mean average of 80.86.

3. Find the mean of your paper airplane data in the dot plot from **Section 4.1b**. Compare this average to your conclusions about how far a typical paper airplane will fly based on the distribution of the data in the dot plot.
   Sample plot and average is given below.

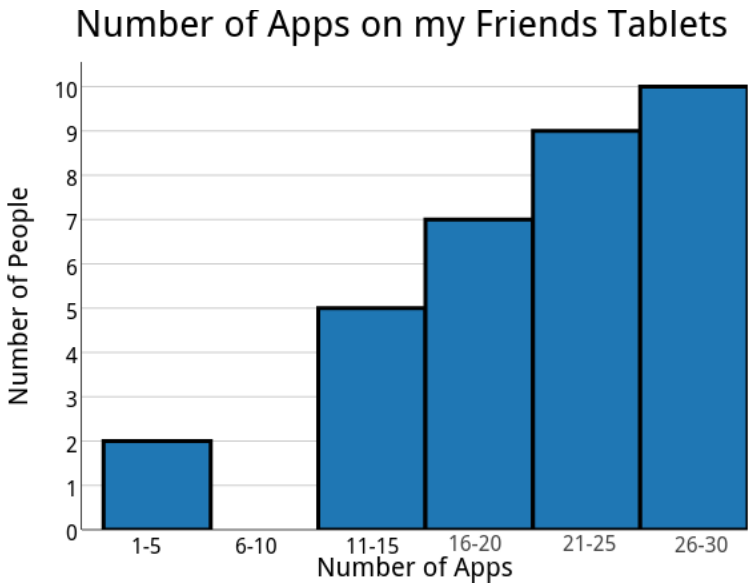


Distance Paper Airplanes Can Fly (feet)

The average distance that a paper airplane flies is 13.25 feet. Based off of the distribution of the dot plot we stated that the typical distance would be between 10 and 17 feet. The mean is within that range.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. The histogram below shows the number of apps that a large group of friends have on their electronic tablets. Is it possible to find the exact mean from the data presented in a histogram? Why or why not?

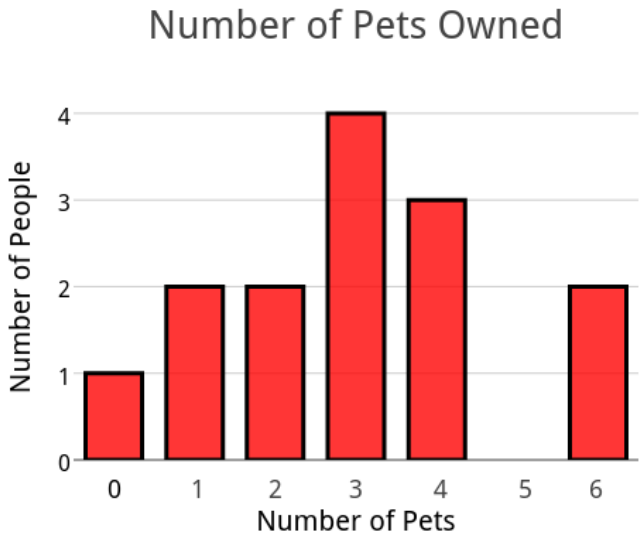


[1]

It is not possible to find the mean from a histogram because a histogram only shows the number of values within a certain interval and not exact values.

5. The **bar graph** below shows the number of pets that fifteen 6th graders have owned in their lifetime. Find the mean number of pets owned from the information given in the graph.

You may want to discuss the difference between and histogram and **bar graph**. A histogram displays the data values within a given interval. It is not possible to know the exact value of a particular piece of data from a histogram, only the interval in which it falls. Also the "bins" or bars on a histogram touch each other; they do not on a bar graph.



[2]

It may be challenging for students to know how to pull the exact data values out of a bar graph. It may be helpful to have them think about it as a dot plot. You can do this by putting dots for each value inside of the bars.

The mean number of pets owned is 3.

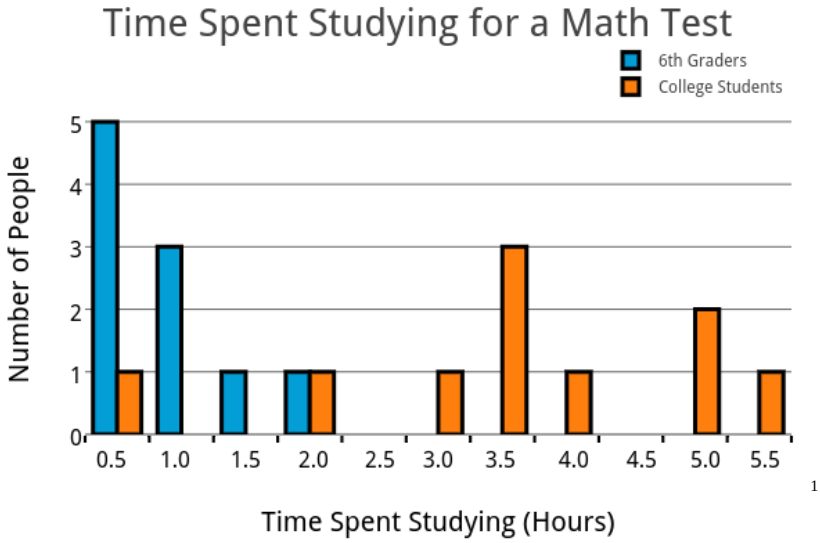
6. In your class 8 people do not get any weekly allowance, 3 people get $5, 5 people get $7, 7 people get $10, 6 people get $12, 1 person gets $14 and 1 person gets $15. Find the average amount of weekly allowance for the class.

The average amount for weekly allowance is $7.13.

1 evatipp. "Number of Apps on My Friends Tables" Plotly Technologies Inc., https://plot.ly/120/~evatipp (Last Accessed: March 2016)
2 evatipp. "Number of Pets Owned" Plotly Technologies Inc., https://plot.ly/307/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

7. The double bar graph shows the time spent doing homework each night for a class of 6th graders and a class of college students.

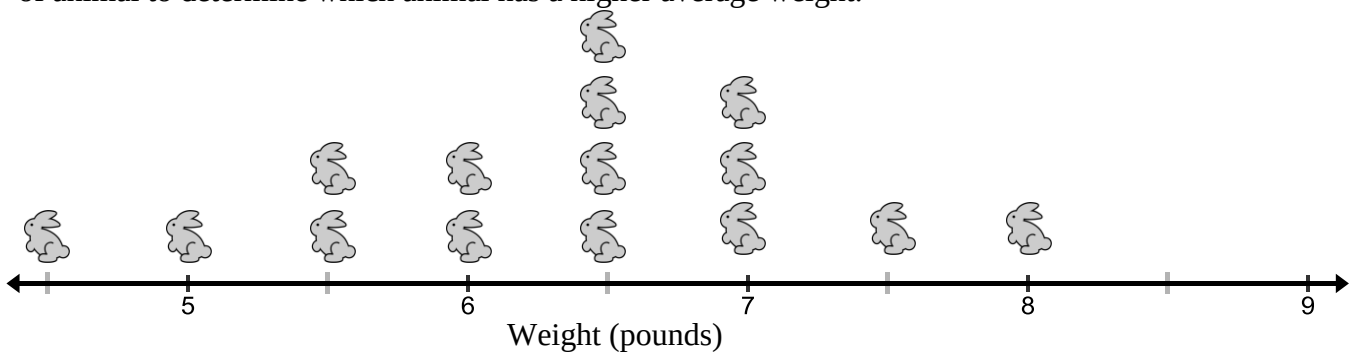


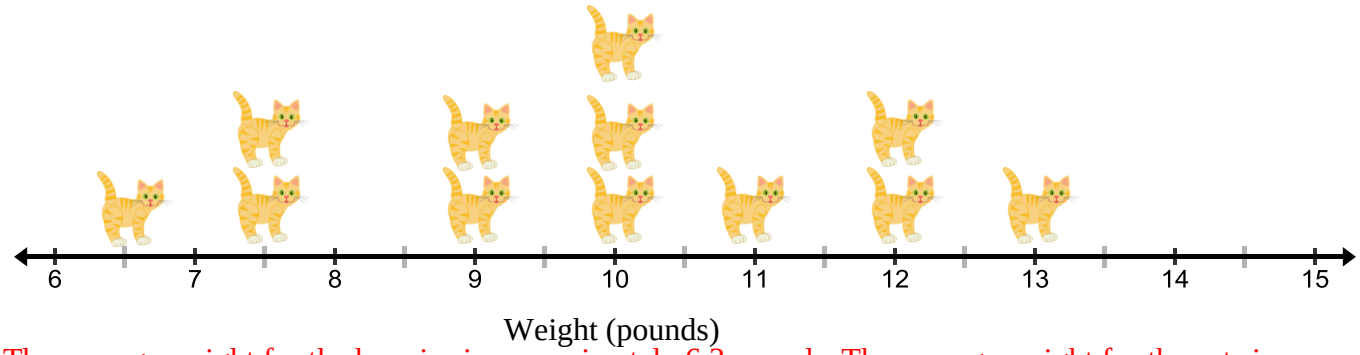
[1]

a. Use the bar graph to estimate the average time that each group of students spent studying for the test.
See student answers.

b. Calculate the average time each group spent studying for the test. Use the averages to determine which group of students spent more time studying for their math test.
6th graders spent an average of 0.9 hours studying for their test.
The college students spent an average of 3.55 hours studying for their test. This means that college students spent more than 3 times the amount of time studying for their test than the 6th graders.

8. The graphs given below show the weights of two different animals. Find the mean weight of each type of animal to determine which animal has a higher average weight.





The average weight for the bunnies is approximately 6.3 pounds. The average weight for the cats is approximately 9.8 pounds.

1 evatipp. "Time Spent Studying For Math Test" Plotly Technologies Inc., https://plot.ly/309/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

9. The table shows the low temperatures recorded for a week in early March.

Daily low temperatures in degrees Fahrenheit
30° 33° 33° 30° 15° 32° 34°

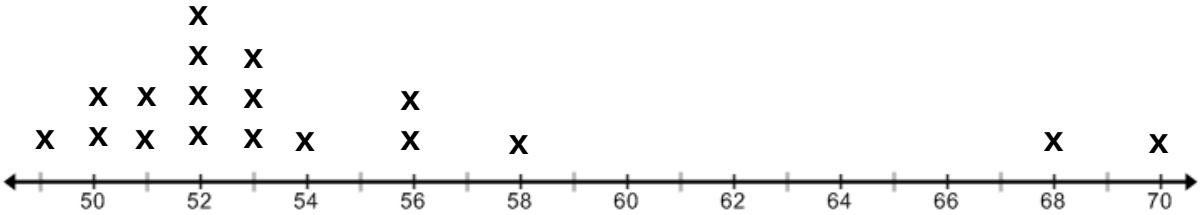
a. Identify any outliers in the data.
15° is an outlier.

b. Find the mean of the data with and without the outlier(s).
The mean with the outlier is ≈ 29.6°F
The mean without the outlier is ≈ 32°F

c. Describe how the outlier affects the mean.
With the outlier the mean is less than all but one of the temperatures. Without the outlier the mean is higher and better represents the data.

If students struggle to identify the outlier(s), have them make a dot plot.

10. The line plot shows the speed of several cars traveling down the highway.



Speed of Cars on Highway in MPH

a. Identify any outliers in the data.
The outliers are 68 mph and 70 mph.

b. Find the average speed of the cars with and without the outlier(s).
With the outliers, the average speed is 54.4 mph
Without the outliers, the average speed is 52.63 mph.

c. Describe how the outlier(s) affects the data.
With the outliers, the mean is higher than 13 of the 18 data values. Without the outliers the mean is lower and better represents the data.

11. Create two different sets of data that have 4 values each and a mean of 15.
Any four numbers that sum to 60 will have a mean of 15.

Students may approach this problem in a variety of ways. They could work backwards by reasoning that 15 was obtained by dividing the sum of the four values by 4. Multiplication is the inverse operation of division so 15 times 4 is 60. The sum of the values must be 60. It does not matter what the values are as long as they sum to 60. They might reason that since the mean is 15 all or most of the numbers must be around 15, they can then guess and check until they find 4 values with a mean of 15. This would be a great problem to return to after students have learned how to view the mean as a balance point.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

12. Owen has recorded four of his quiz scores for his history class below.

7, 10, 8, 8

a. He has one more quiz to take and his mom has told him that if he has an overall average of at least 8 on all his quizzes he can earn extra time playing video games. Determine the lowest possible score he can get on his final quiz to earn his reward.
Owen must get at least a 7 to get an overall mean of 8 for the 5 test scores.

b. Owen's history teacher wants to get an idea of how four of her students, including Owen, did on their 5 quizzes. The teacher's grade book is shown below. Assume that Owen gets a 7 on his last quiz. Show how Owen's teacher can find out how these 4 students did on their quizzes as a whole.

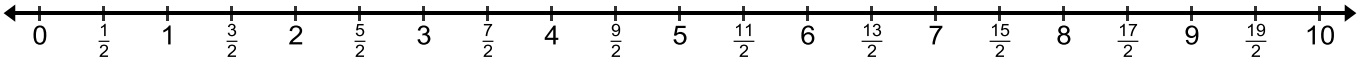
Student	Quiz #1	Quiz #2	Quiz #3	Quiz #4	Quiz #5
Leta	*6*	*5*	*8*	*6*	*7*
Owen	*7*	*10*	*8*	*8*	*7*
Seth	*8*	*7*	*8*	*8*	*7*
Somer	*10*	*10*	*9*	*9*	*10*

Leta's Average: 6.4
Owen's Average: 8
Seth's Average: 7.6
Somer's Average: 9.6
Group Average: 7.9

While it is very common to find the average of the averages, especially when taking measurements or performing an experiment, it is most reliable when the groups are the same size. Students should be cautious of this approach.

Spiral Review

1. Use a ruler to find the length of four items in your desk to the nearest $\frac{1}{4}$ of an inch. Then mark these measurements on the number line given below.
Answers will vary.



2. Solve the following problems without a calculator.
   a. 34×25
   850
   b. 542×33
   17,886
   c. 432×8332
   3,599,424

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2b Homework: Finding the Mean

Directions: Find the mean for each data set. Round to the nearest tenth and be sure to include units in your answer if applicable.

1.

Height of Players (cm)			
132	140	172	144
150	155	165	149
145	161	144	163

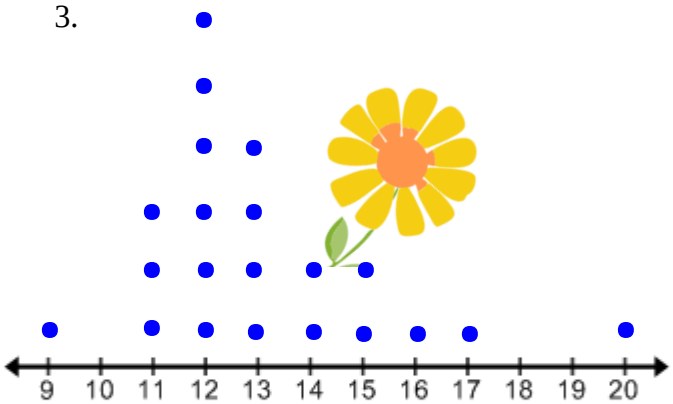
151.7 cm

2.

Number of Boxes Ordered	
Lonna	IIII
Gary	II
Hunter	~~IIII~~
Miriam	~~IIII~~ III
Jonas	III

4.4 boxed ordered

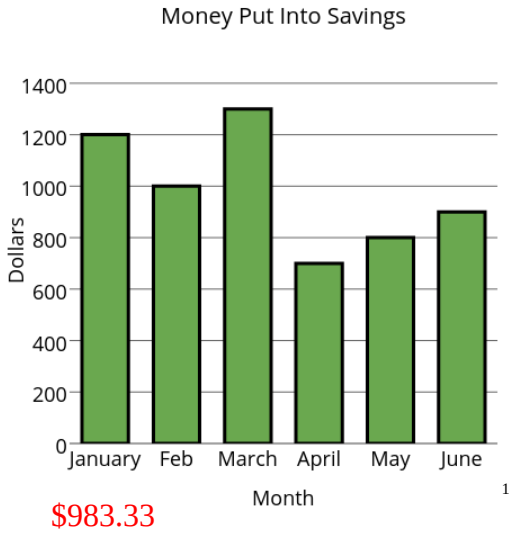
3.



Number of Yellow Flowers in Vases

13.2 yellow flowers per vase

4.



[1]

$983.33

5. Number of sit-ups: 30, 34, 20, 28, 31, 28, 27

28.3 sit-ups

6. Movies seen this summer: 4, 6, 3, 0 , 2, 4, 1, 3, 2, 5, 4, 6, 3, 4, 2, 3, 1

3.1 movies seen this summer

7. 71, 67, 73, 73, 66

70

8. 4.5, 5.1, 4.6, 4.7, 4.5, 5.0, 4.9

4.8

1 evatipp. "Money Put Into Savings" Plotly Technologies Inc., https://plot.ly/345/~evatipp (Last Accessed: March 2016)

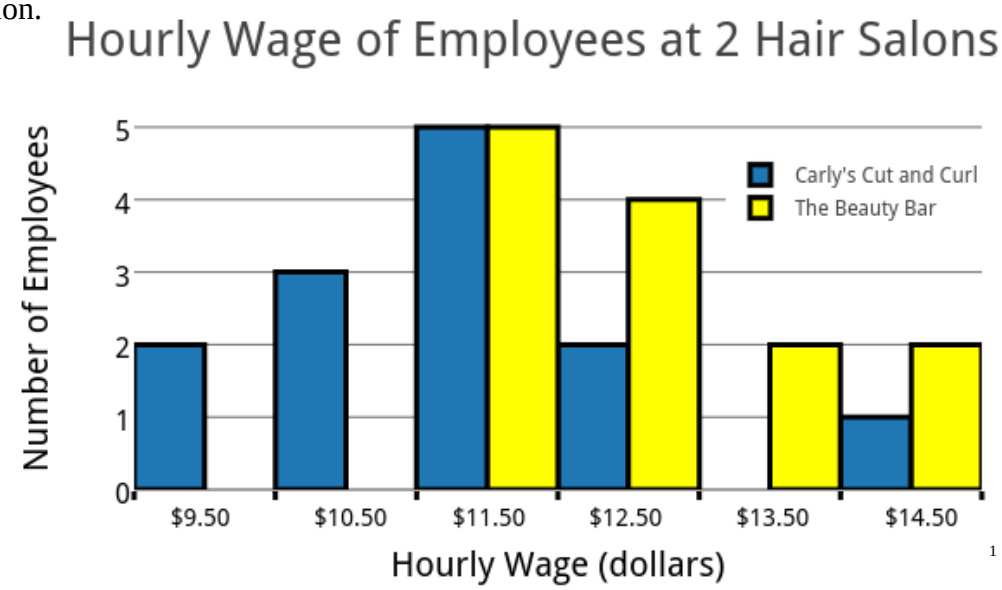
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

9. In order to qualify for the next round of a snowboarding competition, Tara must have an average score of at least 80 on her first 4 rounds of competition. Her scores for the first 3 rounds are below. What is the lowest possible score that she can get on her fourth round to move on in the competition?

Round 1	Round 2	Round 3	Round 4
90	75	72	

Tara must score at least an 83 on the 4th round.

10. The bar graph below shows the hourly wage of employees that work at two different hair salons. Determine which salon will most likely pay you more by finding the average hourly wage for employees at each salon.



[1]

Carly's Cut and Curl has an average hourly wage of $11.35. The Beauty Bar has an hourly wage of $12.58. Thus the Beauty Bar will most likely pay you more because it has a higher average hourly wage.

11. The data below shows the attendance of fans at each home game for the high school baseball team.

Fan attendance at home games: 34, 37, 30, 15, 34, 38, 41, 33

a. Which data value is an outlier? Explain.
15 is the outlier, it is much less than the other data values.

b. Calculate the average attendance with and without the outlier. Then describe how the outlier affects the mean.
Average Attendance with the outlier: 32.75

Average Attendance without the outlier: 35.29
With the outlier the average is less than 6 of the 8 data values. Without the outlier the average is a better representation of how the data is distributed.

c. Describe a situation that could have caused the outlier in this set of data.
Possible situations might include that it was raining on the day of this game, another important game was being played on that day that fans would rather attend, this game was played over spring break so there were not many fans in town.

1 evatipp. "Hourly Wage of Employees at 2 Hair Salons" Plotly Technologies Inc., https://plot.ly/345/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

12. The table below shows how much money Bethany spent on new school clothes.

Item	T-Shirt	Pants	Skirt	Shoes	Sweater	Shorts	Dress	Button – up Shirt
Cost	$15.00	$25.00	$18.50	$48.00	$22.00	$18.00	$25.00	$20.00

a. Find the average cost per item?
$23.94

b. Do any outliers affect the average cost? If so, eliminate the outlier and find the new mean.
The amount paid for shoes is an outlier; the mean without the outlier is $20.50.

c. Bethany insists that, on average, her friends spend more money per item on school clothes than she does. To be sure, she asks 6 of her friends to tell her the average amount of money per item that they spent on school clothes. Their data is given below.

Tami- $13.50, Dusty- $25.50, Wade- $13.00, Bobby- $28.30, Luis- $17.00, Giselle- $9.75

Find the average amount of money that Bethany's friends spend per item on school clothes and then compare it to her average cost per item.
The average cost per item for her friends is $17.84. This means than Bethany spends more per item than most of her friends on school clothes.

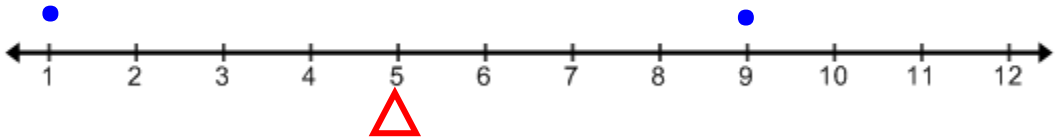
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2c Class Activity: Mean as a Balance Point

Understanding the mean as a balance point will help students better understand why and how mean absolute deviation is a measure of variability in **section 4.3**. This approach is another tool that students can use to better understand mean.

Carly wants to know how long her friends can do a handstand. She asks two of her friends to do a handstand. One friend can do a handstand for 9 seconds, her other friend can only do a handstand for 1 second. She records their times on the dot plot below.

Number of Seconds My
Friends Can Do A Handstand.



1. Carly's teacher says that the ***center*** of this data is the ***balance point***. Discuss with a neighbor what you think she means by this. Then draw a triangle under the dot plot where you think the balance point is. Ask students to justify where they placed the triangle. They may argue that 5 is the same distance away from 1 and 9 or that 5 is directly in the middle or center of 1 and 9.

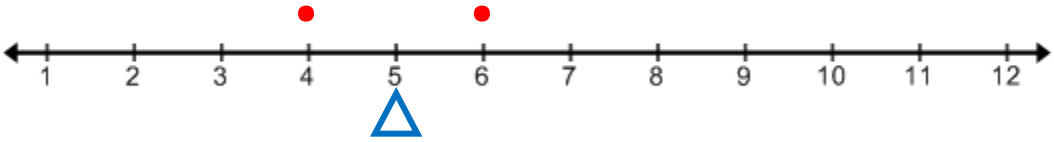
2. Carly notices that the balance point is also the mean of the two data values. Confirm this by calculating the mean below.

$$\frac{1+9}{2} = \frac{10}{2} = 5$$

Point out that this arithmetic shows that 1 and 9 are the same distance away from 5.

3. Can you place the values in a different location so that the plot still remains balanced at 5? Redraw the plot below with your new values. Sample values are given.

Number of Seconds My
Friends Can Do A Handstand.



Discuss student answers and have them justify their values. They may plot any two values that are the same distance away from 5. Ask them to justify how they know that the plot is still balanced. In the sample values given above they may argue that they moved both of the original values or dots 2 spaces towards 5 so it is still balanced. Or that 4 and 6 are both 1 space away from 5. Make sure that it is clear that 5 is the balancing point. Students might begin to surface ideas about variability as they discuss how the mean is still 5 regardless of how close the data values are to the balancing point as long as they are the same distance away from 5.

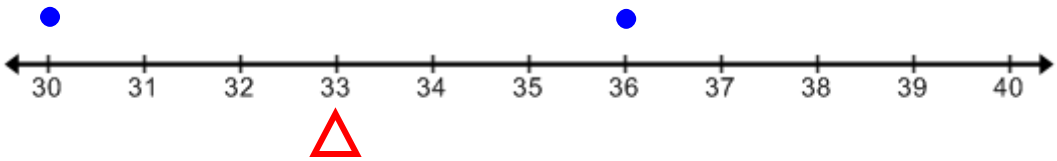
4. Check that the plot is balanced by calculating the mean of your two new data values. Calculations will vary depending on their values. Sample calculation is given.

$$\frac{4+6}{2} = \frac{10}{2} = 5$$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

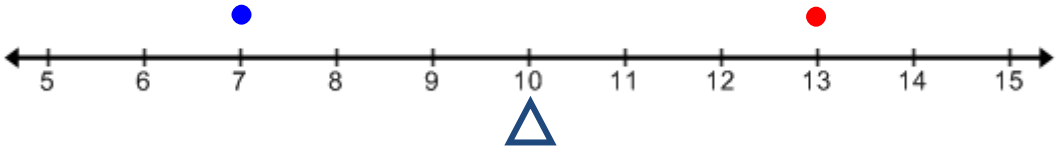
5. What do the **mean** and the **balancing point** of the data have in common?
   They are the same thing; they both describe the *center* of the data.

6. Find the balancing point of the two data values on the plot below. Draw a triangle to indicate the point of balance.



   Ask students how they know that 33 is the point of balance. Discuss that 33 is the balancing point because it is the same distance away from 30 and 36. Emphasize that the balancing point is the mean. Check their answer by finding the mean of the two data values. $\frac{30+36}{2} = \frac{66}{2} = 33$

7. One data value of 7 is shown on the plot below. Where should the other value be placed so that the plot is balanced at 10?



   Discuss that 13 is the other value in the data set because it is also 3 spaces away (and on the opposite side) from 10. Ask students to confirm their answer by calculating the mean of 7 and 13. $\frac{7+13}{2} = \frac{20}{2} = 10$

8. If the dot at 7 is moved to the left two spaces, where should the other value be moved to so that the plot remains balanced at 10? Justify your answer.
   The other value should be moved to 15. If the value of 7 is moved to 5 then it is now 5 spaces away from 10. This means that the value of 13 must also be moved 2 spaces away from 10 to 15 so that the plot remains balanced.

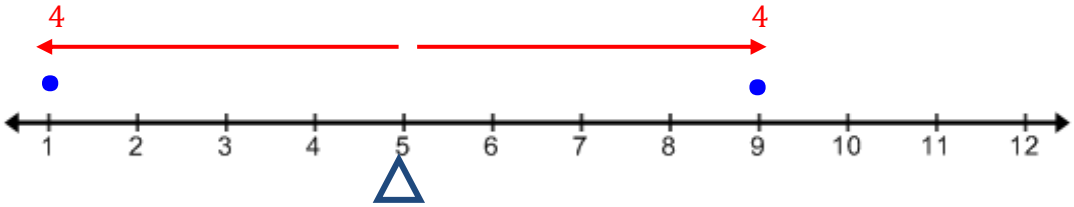
9. Confirm that the mean of the new data points found in #8 is 10.

$$\frac{5+15}{2} = \frac{20}{2} = 10$$

The original data for Carly's two friends have a balance point of 5. We justified that the balancing point was 5 because 1 and 9 are both 4 spaces away from 5. In mathematics we call these distances *absolute deviations* from the mean. The difference between the balancing point or mean and each data point is an ***absolute deviation.*** This is shown with Carly's handstand data below.

The distance from a data value to the mean is positive which is why it is called an absolute deviation. You can relate this to absolute value.

Number of Seconds My Friends Can Do A Handstand.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

10. What relationship does the absolute deviation on the right of the balance point have with the absolute deviation on the left of the balance point?
    The absolute deviations are the same.

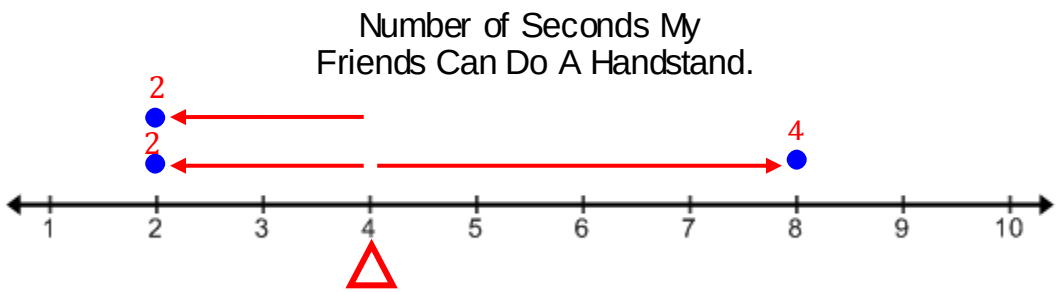
It is true that deviations to the left are negative and deviations to the right are positive. Later on in this lesson we will be finding sums to make comparisons. However, operations with integers do not occur until 7th grade. Since this is the case, we are only going to deal with absolute deviations, which are positive (or non-negative).

11. Analyze the absolute deviations for the values you used in problems 3, 6, and 7. What relationship do the absolute deviations to the right of the balance point have with the absolute deviations to the left of the balance point?
    The absolute deviations to the right of the balance point are the same as the absolute deviations to the left of the balance point.

The mean for a set of data is where the absolute deviations to the right of the balance point are the same as the absolute deviations to the left of the balance point.

The concept of deviations is very important in future statistical analysis. In the next section, students will analyze Mean Absolute Deviation as they look at variability. In future grades they will learn standard deviation.

What if Carly collected handstand data from more than 2 friends? The plot below shows how long three other friends can do a handstand.



Many students may struggle with understanding that you must sum the absolute deviations on each side of the balancing point, writing the value for each deviation above the corresponding dot is helpful.

12. What is the balancing point for this data? Draw a triangle to indicate a point of balance. Use the mean and the absolute deviations of the data values to justify your answer.
    The point of balance is 4.
    The absolute deviation from 2 to 4: 2
    The absolute deviation from the other 2 to 4: 2
    The absolute deviation from 8 to 4: 4

    Total of the absolute deviations on the left of the balance point is the sum of the absolute deviations: $2 + 2 = 4$.

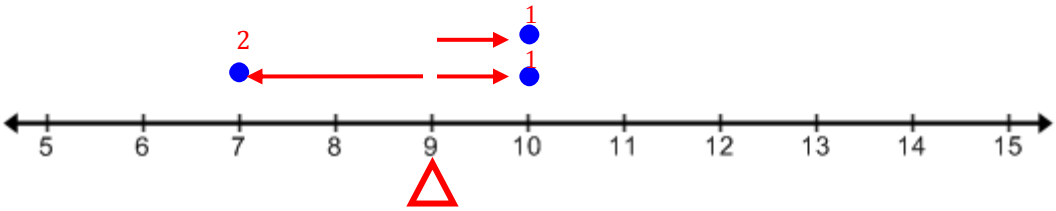
    The sum of the absolute deviations on the left are equal to the sum of the absolute deviations on the right $4 = 4$
    The mean of the data is $\frac{2+2+8}{3} = \frac{12}{3} = 4$

Some students may argue that 5 is the balancing point because it is both 3 spaces away from 2 and 8. Discuss with them that we must account for the "total" distance and since there are two values of 2 their total distance away from 5 is $3 + 3 = 6$.
Point out that when there is more than one data value to the left or right of the balance point be sure to find the total absolute deviations for each side. In this example there are two values of 2 on the left so that total of the absolute deviations for the left side is 4.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

13. Find a balancing point for each set of data. Draw a triangle to indicate a point of balance. Check that the sum of the absolute deviations to the left of the balance is equal to the sum of the absolute deviations to the right of the balance point. Also calculate the mean to check your answer.

a.



At this point students might notice that you can find the absolute deviation of the values by finding the difference between the data values and the mean or through subtraction.

Absolute deviations to the left: $9 - 7 = 2$

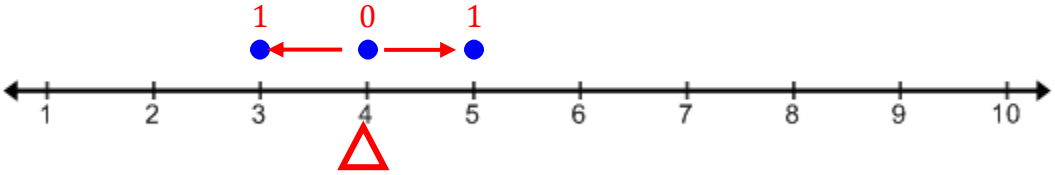
Absolute deviation to the right: $10 - 9 = 1$, $10 - 9 = 1$ $\Big\}$ $1 + 1 = 2$

Absolute Deviation Comparison→ $2 = 2$

Mean: $\frac{7+10+10}{3} = \frac{27}{3} = 9$

It is not nessacary for students to show all the work above, it is given as a guide. Encourage students to use whatever method (subtraction or counting distance on the number line) they are most compfortable with to find the absolute deviations.

b.



Absolute deviations to the left: $4 - 3 = 1$

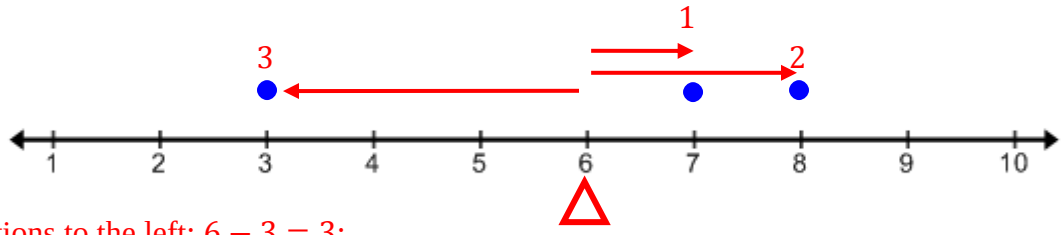
Absolute deviations to the right: $5 - 4 = 1$

Absolute deviation on the balance point: $4 - 4 = 0$

$\rightarrow 1 = 1$

Mean: $\frac{3+4+5}{3} = \frac{12}{3} = 4$

c.



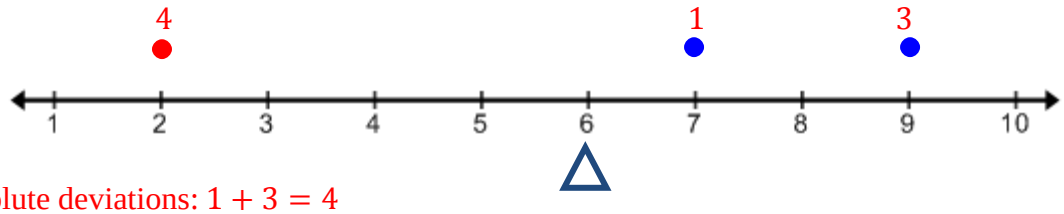
Absolute deviations to the left: $6 - 3 = 3$;

Absolute deviations to the right: $7 - 6 = 1$, $8 - 6 = 2$ $\Big\}$ $1 + 2 = 3$

$\rightarrow 3 = 3$

Mean: $\frac{3+7+8}{3} = \frac{18}{3} = 6$

14. Two values are shown in the plot below, where should a third value be placed so that the plot is balanced at 6? Check your answer by comparing the sum of the absolute deviations on each side of balance point and calculating the mean of your data points.



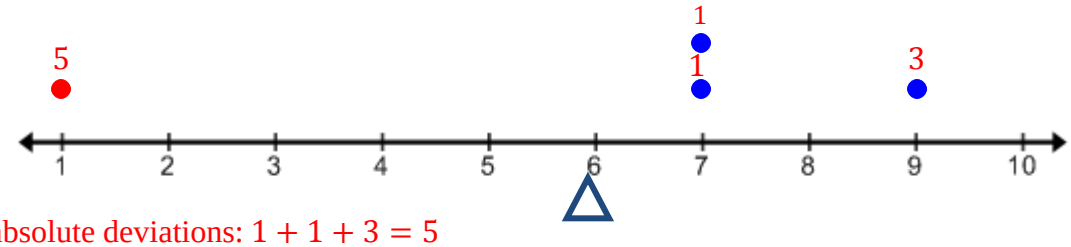
Sum of right absolute deviations: $1 + 3 = 4$

Left absolute deviation needs to be 4 also. $6 - 4 = 2$

Mean: $\frac{2+7+9}{3} = \frac{18}{3} = 6$

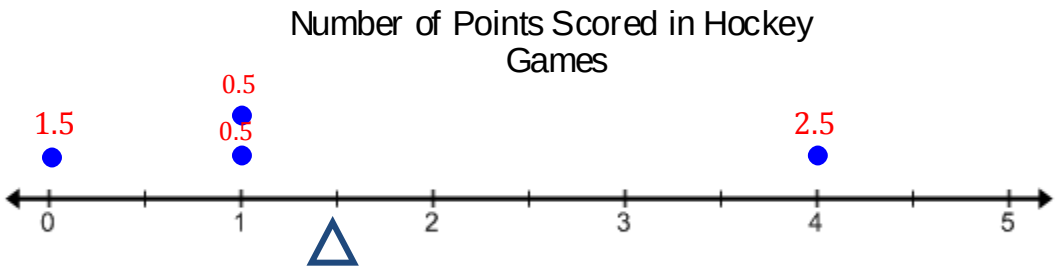
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

15. What if an additional value of 7 is added to the plot, where should the fourth value be placed so that the plot is still balanced at 6? Check your answer by comparing the sum of the absolute deviations on each side of the balance point and calculating the mean of your data points.



Sum of right absolute deviations: $1 + 1 + 3 = 5$
Left absolute deviation needs to be 5 to balance. 6 – 5 = 1
Mean: $\frac{1+7+7+9}{4} = \frac{24}{4} = 6$

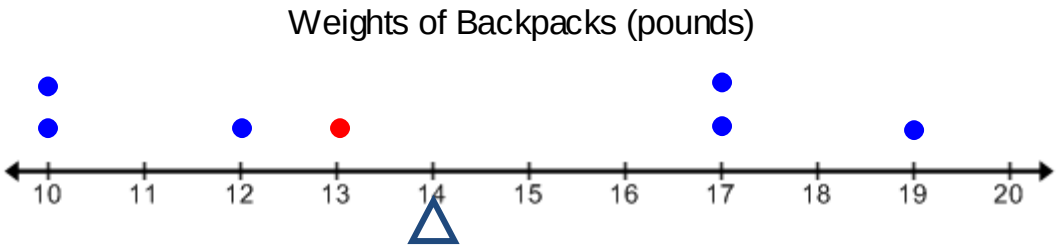
16. The dot plot shows the number of points that the Ice Rebels scored in 4 of their hockey games last season. Use the "balancing" process to explain why the mean number of goals for the team is 1.5.



Sum of left absolute deviations: $1.5 + 0.5 + 0.5 = 2.5$
Right absolute deviations: 2.5
The sum of the deviations on each side of the mean is the same.

17. The weights in pounds of six loaded backpacks are shown in the plot below. The weight for the seventh backpack is missing but the mean for the weights of all seven backpacks is 14 pounds. Use the "balancing" process to find the weight of the seventh backpack.

.



The sum of the left absolute deviations is $(\mathbf{14 - 10}) + (\mathbf{14 - 10}) + (\mathbf{14 - 12}) = \mathbf{4 + 4 + 2 = 10}$.
The sum of the right absolute deviations is $(\mathbf{17 - 14}) + (\mathbf{17 - 14}) + (\mathbf{19 - 14}) = \mathbf{3 + 3 + 5 = 11}$.
The seventh backpack must have an absolute deviation that causes the sum of the left and right absolute deviations to be the same. This means that we must add one more absolute deviation to the left side of the mean. 1 space to the left of 14 is 13. This means that the weight of the seventh backpack must be 13 pounds.

Students do not need to use subtraction to find the absolute deviations for each point. However, if students are given a data set with values that are very far away from the mean you will want them to know that they can use subtraction to find the absolute deviation rather than counting the distance on the number line.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Spiral Review

1. Write each fraction as a mixed number or whole number.

   a. $\frac{15}{4}$ b. $\frac{43}{10}$ c. $\frac{8}{3}$ d. $\frac{66}{2}$

2. Solve each problem without using a calculator. Use a tape diagram if needed.

   a. $8 \times \frac{1}{2}$ b. $30 \times \frac{1}{6}$ c. $48 \times \frac{1}{8}$ d. $100 \times \frac{1}{5}$

3. Estimate to the answers.

   a. 9×99 b. 15×9 c. 19×9 d. 20×49

4. Find the following without a calculator.

   a. One tenth more than five tenths.
   b. One tenth more than twenty-five hundredths.
   c. One hundredth more than thirty-five thousandths.
   d. One thousandth more than 45 hundredths.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2c Homework: Mean as a Balancing Point

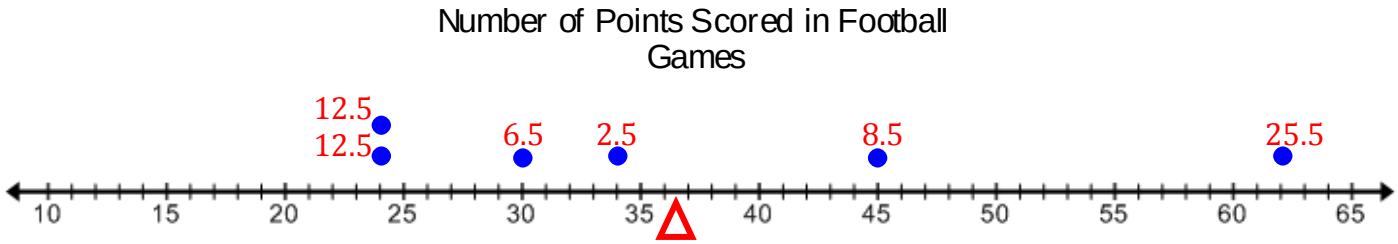
1. The number of Harry Potter books that 5 different friends have read are below.

$$0, 5, 7, 4, 4$$

   a. Find the mean number of Harry Potter books read for this group of friends.

   b. Make a dot plot of the data set and check your answer by comparing the sums of the absolute deviations on each side of the mean.

2. The number of points that the University of Utah football team scored in the first six games of their 2015 season are recorded in the plot below.



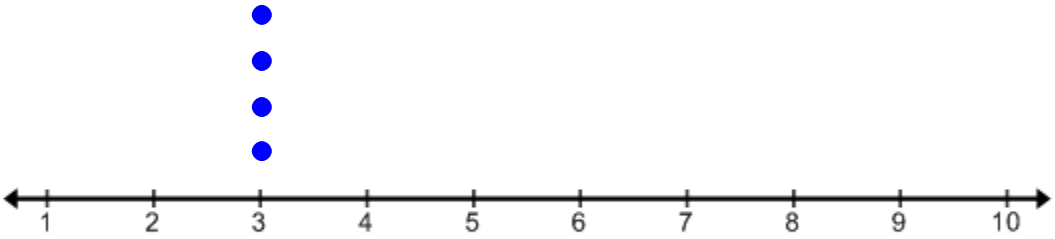
   a. Find the mean number of points scored for their first 6 games.

$$\frac{24 + 24 + 30 + 34 + 45 + 62}{6} = \frac{219}{6} = 36.5$$

   b. Verify your answer to part a by using the "balancing" process and comparing the sums of the absolute deviations on each side of the mean. Encourage students to use subtraction to find the absolute deviations of values that are far away from the mean.
   Sum of left absolute deviations: $12.5 + 12.5 + 6.5 + 2.5 = 34$
   Sum of right absolute deviations: $8.5 + 25.5 = 34$
   Sums of the deviations are the same on both sides of 36.5, thus it is the mean.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

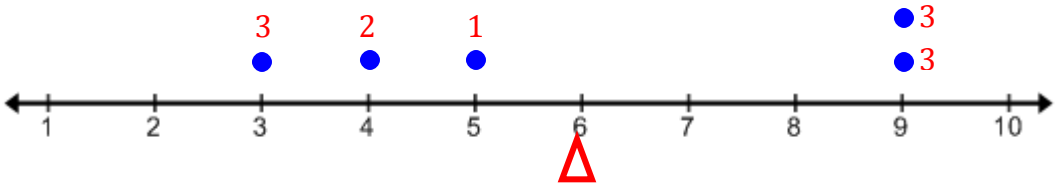
3. Use the "balancing" process to find the mean.



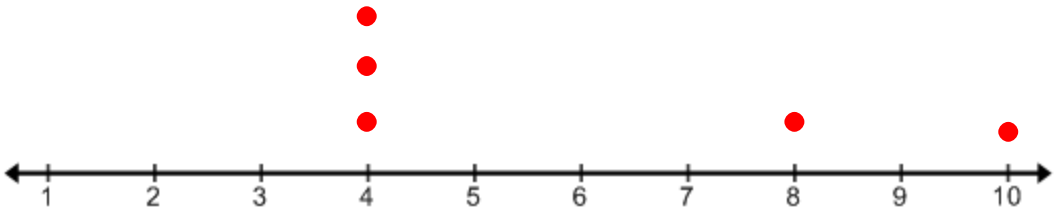
4. Use the data values given below to answer the questions that follow.

4, 5, 9, 3, 9

a. Plot these numbers on the dot plot given below and use the "balancing" process to find the mean. Draw a triangle to represent the mean.

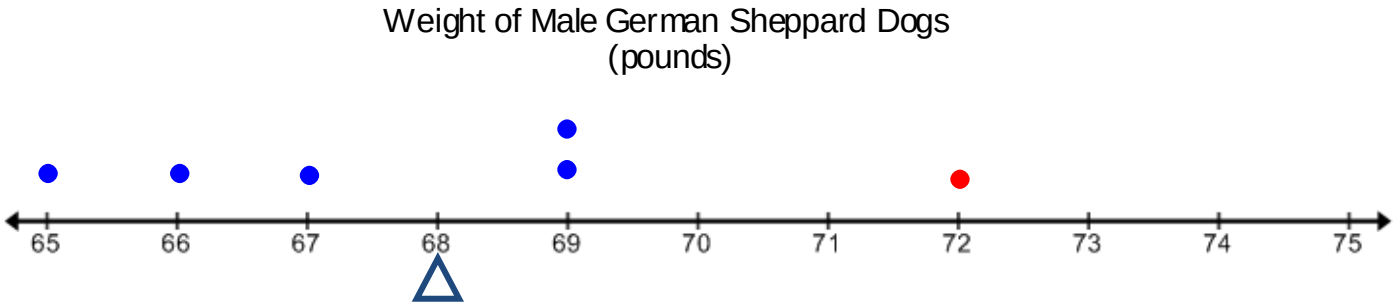


b. Rearrange the numbers from the plot above so that the mean or balance point remains the same. Explain your process.
Any arrangement will work where the sum of the deviations on both sides of 6 are the same. Sample answer is given.



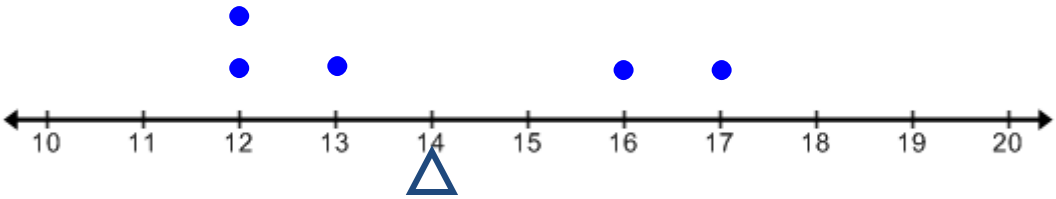
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. The weights in pounds of five male German Sheppard dogs are shown in the plot below. The weight for the sixth dog is missing but the mean for the weights of all six dogs is 68 pounds. Use the "balancing" process to find the weight of the sixth dog.

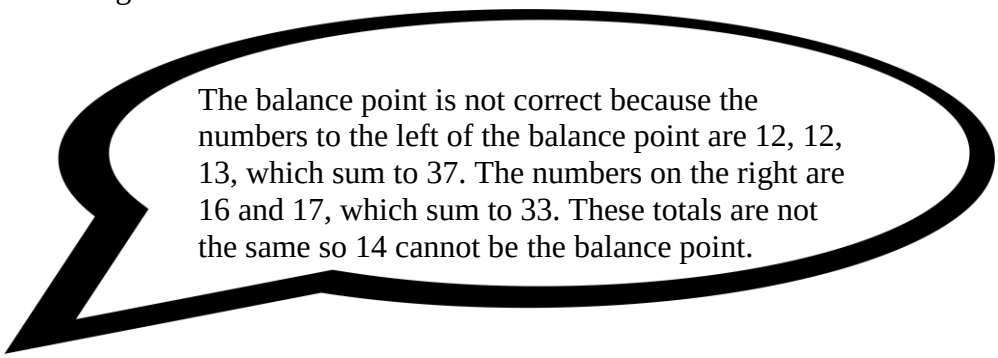


The sum of the left absolute deviations is $(\mathbf{68-65})+(\mathbf{68-66})+(\mathbf{68-67})=\mathbf{3+2+1=6}$.
The sum of the right absolute deviations is $(\mathbf{69-68})+(\mathbf{69-68})=\mathbf{1+1=2}$.
The sixth dog must have an absolute deviation that causes the sum of the left and right absolute deviations to be the same. This means that we must add 4 more absolute deviations to the right side of the mean. 4 spaces to the right of 68 is 72. This means that the weight of the sixth dog must be 72 pounds.

6. Find, Fix, and Justify
   Lila is examining the plot below. She is trying to decide if the balance point given is correct.



She makes the following statement.



Lila has made a mistake. Find her mistake and justify why 14 is the balance point of the data value.
She should have compared the sums of the absolute deviations on each side of the balance point.
The sum of the left absolute deviations is $(\mathbf{14-12})+(\mathbf{14-12})+(\mathbf{13-12})=\mathbf{2+2+1}=5$.
The sum of the right absolute deviations is $(\mathbf{16-14})+(\mathbf{17-14})=\mathbf{2+3=5}$.
Sums of the deviations are the same on both sides of 14, thus it is the mean.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2d Class Activity: Median and Mode

Recall that a *measure of center* describes a data set by summarizing all the values with a single number. There are other ways to measure the center of a data set in addition to finding the mean (the balancing point). These other measures of center are called the **median** and **mode**.

Directions: Once again grab a handful of linking cubes, count the number of cubes that you have, and write this number on a sticky note.

1. Explain how you found the **median** of the linking cube data below:

Option 1: Find the median of the data for the entire class. Ask students what they think the term median means, most likely someone will say it means "the middle". Explain to students that the median is found by counting to the middle number of an *ordered* data set. Have the students put their data in order either by placing their sticky notes in order on the board or by putting themselves in order. Have the class figure out how to find the middle number. They could do this by counting into the middle number or by having the students on both ends of the data sit down simultaneously one after another until the middle number is reached. If the number of students in your class is odd then the median will be the middle value. If the number of students in your class is even then have a discussion about how the median will be the average of the two middle values. For either scenario add an additional data value to the data (this could be your handful as a teacher) so that you can discuss both scenarios.

Option 2: Divide the class into groups and have them find the median for their group, this may be more manageable than option 1. Encourage student to link their cubes into rods and to find the median pictorially with their rods of cubes. They can order their data by placing the rods in order from shortest to tallest. They can then count to the middle of the data by pulling rods off the ends of the data simultaneously one after another until the middle number is reached. Or they can count into the middle number with their fingers. Have students share their methods with the class. Be sure to discuss scenarios of finding the median with an odd number of data values and an even number of data values.

Discussion questions might include: Does the median always have to be a value in the data set? Does the median always have to be a whole number? Explain how the median is a measure of center for a data set.

2. Explain how you found the **mode** of the linking cube data below:

Option 1: Find the mode by making a dot plot of the data values with the sticky notes on the board. Discuss how it is easy to find the mode on a dot plot by identifying the number that has the most marks (sticky notes) above it.

Option 2: Divide students into groups and have them use their rods of linking cubes to find the mode. They can do this by grouping rods that have the same number of cubes together. The group that has the most rods is the mode.
Some discussion questions might include: Can a set of data have more than one mode or no mode at all? Does the mode always have to be a value in the data set?

3. What is the mean of the data set?
   Answers will vary from class to class.

4. Compare and contrast the mean, median, and mode. Is there one measure of center that best represents your data?
   Discuss how sometimes these measures of center might be the same number. Also when you look at all three numbers together it gives you a better sense of the center of the data.

   Median: **The median is found by ordering the data from least to greatest and the finding the middle number. If there are two numbers in the middle then the median is the mean of those two middle numbers.**

   Mode: **The mode of a data set is the value or values that occur most often. There can be one mode, no mode, or more than one mode.**

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Examples

Find the median and mode of each data set

a. Data Set: 9, 10, 13, 10, 8, 11, 12

Median: Order the data from least to greatest. The median is the middle number.

$$8, 9, 10, 10, 11, 12, 13$$

The median is 10.

Mode: The mode is the value that occurs most often.

$$9, 10, 13, 10, 8, 11, 12$$

The mode is 10.

b. Data Set: 4, 7, 4, 8, 9, 5, 3, 7

Median: Order the data from least to greatest. If there is an even number of data values then the median is the mean of the two middle values.

$$3, 4, 4, 5, 7, 7, 8, 9$$

The median is 6. $\frac{5+7}{2} = \frac{12}{2} = 6$

Mode: The mode is the number that occurs most often.

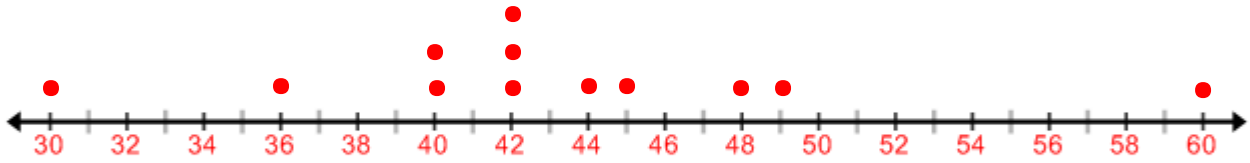
$$4, 7, 4, 8, 9, 5, 3, 7$$

The mode is both 4 and 7.

5. The list below shows the words per minute that a group of 12 students can type.

Words typed per minute: 40, 44, 49, 30, 40, 42, 42, 48, 60, 36, 42, 45

a. Make a dot plot of the data.



Words Typed per Minute

b. Find the mode of the data.
The mode is 42. Talk about how the dot plot makes finding the mode very easy.

c. Find the median of the data.
The median is 42. The dot plot also helps you to find the median because it puts the numbers in order.

d. Do you think that the median and mode will always be the same number?
No, the median could be a number that is not one of the original values and there could be no mode or more than one mode.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Directions: Find the median and mode for each set of data.

Making a dot plot is not necessary to find the median and mode. However, it may be a very helpful tool for students that struggle with organizing data. Encourage students to make a dot plot if they need to.

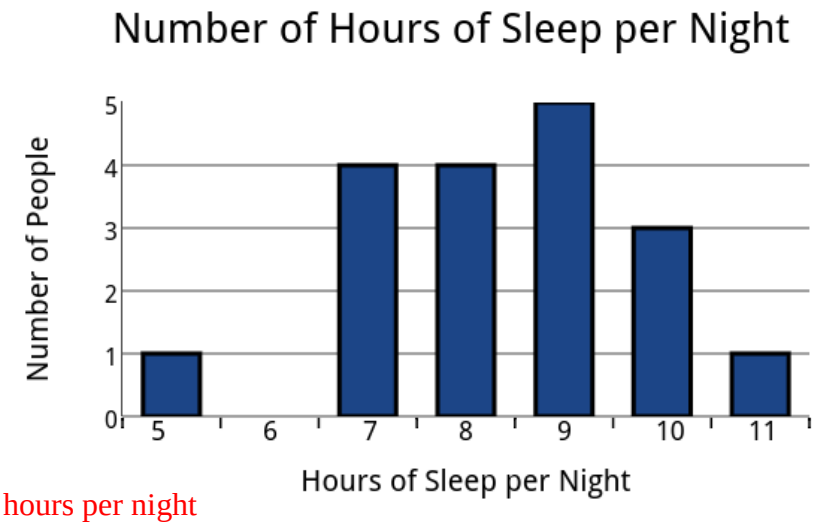
6. The table shows the number of students from each grade that are participating in the school Fun Run.

Grade	Kindergarten	1st	2nd	3rd	4th	5th	6th
Number of Students	75	60	55	63	40	58	65

Median- 60 students

Mode- None

7. The bar graph shows the average amount of sleep per night in hours for students in Ms. Kekoa's 6th grade class.



[1]

Median- 8.5 hours per night

Mode- 9 hours per night.

8. Amount of medicine (in milliliters) given to patients: 2.5, 7.5, 5, 5, 10, 5, 7.5, 5, 2.5, 10, 7.5, 7.5

Median- 6.25 ml

Mode-5 ml and 7.5 ml

9. Pounds of flour used per week at different bakeries: 300, 200, 250, 225, 250, 275

Median-250 pounds

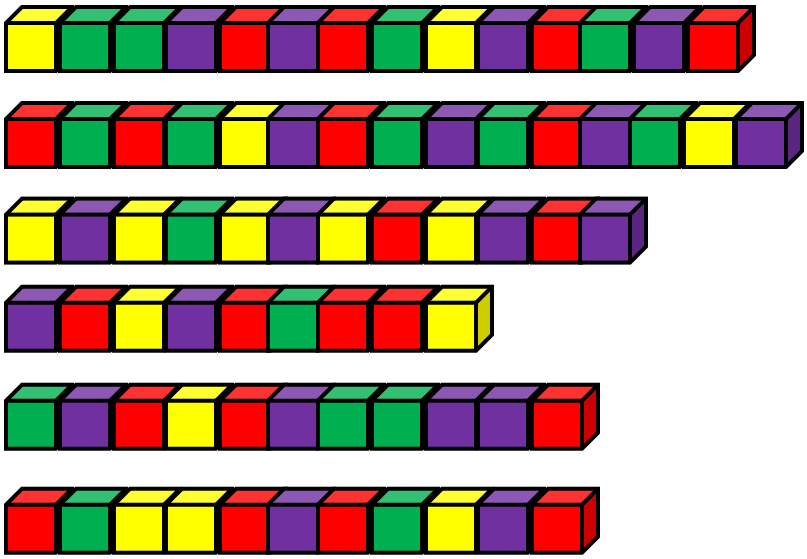
Mode-250 pounds

1 evatipp. "Number of Hours of Sleep per Night" Plotly Technologies Inc., https://plot.ly/351/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

10. Find, Fix, and Justify

A group of 6 students grabbed a handful of linking cubes. Each person attached their cubes into the rods shown below.

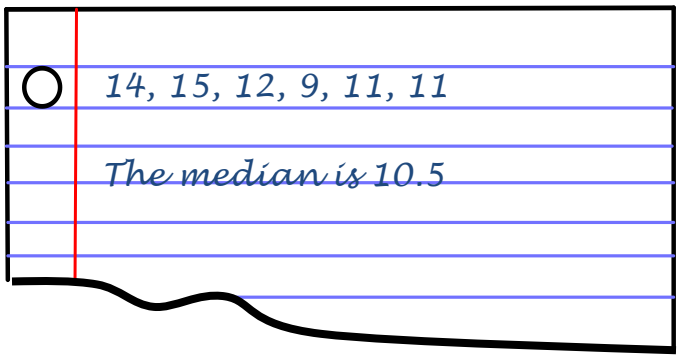


There is not a homework assignment for this lesson. You could assign this page as homework if needed.

As students critique the reasoning of others in the problem below, they must identify the mistakes of others. The mistakes represent common misconceptions that student's may have when finding the median.

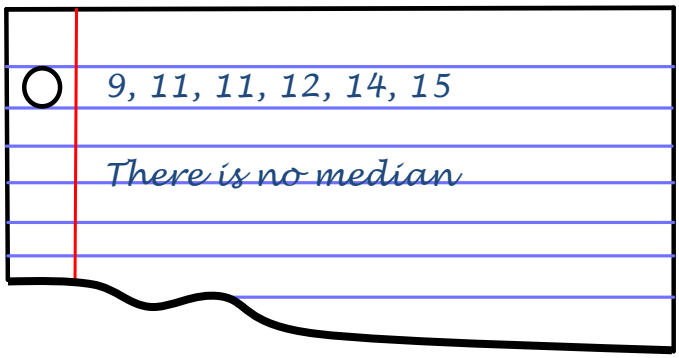
Monte and Ava have been asked to find the median number of cubes in a handful. Their answers are given below, but they have made a mistake in their reasoning. For each person, find their mistake and fix it.

a. Monte's Answer



Monte did not put the data values in order from least to greatest before finding the middle number. The median is 11.5.

b. Ava's Answer



Ava assumed that since there is not a single middle number there is not median. She needs to find the mean of the two middle numbers: median is 11.5.

c. What is the mode for the data set?
The mode is 11.

d. What is the mean number of cubes in a handful?
The mean is 12.

11. Find the mean, median, and mode for each set of data. Round your answer to the nearest hundredth.

a. 13.22, 11.05, 10.77, 15.04, 12.3, 12.89, 14.7, 16.3, 13.9
Mean- 13.35
Median- 13.22
Mode- None

b. $5\frac{1}{2}$, $3\frac{3}{8}$, $1\frac{1}{8}$, 2, $3\frac{5}{8}$, $2\frac{3}{4}$, $1\frac{1}{8}$
Mean- 2.79
Median- $2\frac{3}{4} = 2.75$
Mode- $1\frac{1}{8}$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2e Class Activity: Choosing the Best Measure of Center

1. Becky is the head of a small tech company. The salaries for all of her employees, including herself, are shown in the table below.

Title	Annual Salary	Number of People	Total
Chief Executive Officer	$110,000	1	$110,00
Marketing Manager	$80,000	1	$80,000
Sales Manager	$80,000	1	$80,000
Project Managers	$50,000	2	$100,000
Developers	$40,000	4	$160,000
Office Assistants	$30,000	3	$90,000
		Total	$620,000

a. How many people work at Becky's company?
12 people work at Becky's company.

b. Complete the table to find the total annual salary bill for the company?
$620,000

c. Find the mean (average) salary for employees at this company. Round your answer to the nearest dollar.
$51,667

d. Julia looks at the table and claims that the mode for this data is $80,000. What mistake has Julia made and what is the correct mode?
Julia looked at the Annual Salary's column and saw that $80,000 occurs more than the other salary amounts. However she should have taken into account the number of people that have each salary amount. The correct mode is $40,000 because 4 employees earn that much money.

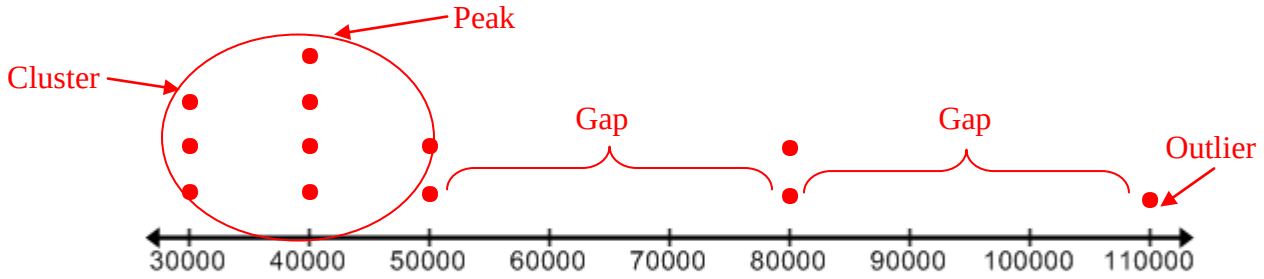
e. What is the median salary for Becky's company?
The median is $40,000.

f. List all the measures of center and decide which measure of center do you think best represents the data? Justify your answer.
Mean-$51,667
Median-$40,000
Mode-$40,000
This problem is meant to get students thinking about how to determine the most appropriate measure of center. Allow students' responses to be very informal and open here.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Throughout this lesson, students will be arguing which measure of center is best. Allow students the opportunity to justify their reasoning aloud and ask other student to critique their reasoning as they begin to learn about how the context and shape of the data can help them make this decision.

g. Make a dot plot of the salary data for Becky's company and use it to justify or change your answer from part f.
If you feel that making the dot plot takes too much time, have the finished dot plot ready to be displayed for a class discussion.



Annual Salaries (dollars)

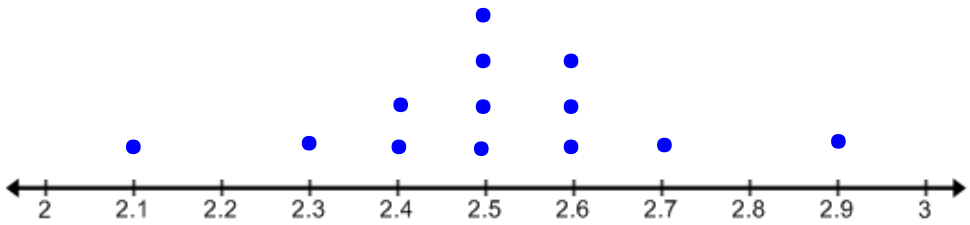
At this point, encourage students to use the "shape" of the data to justify their answers. Possible discussion questions are given below.
What is the shape of this data? –This data is skewed right
Point out any peaks, clusters, gaps, or outliers. How do these features affect the measures of center?-The outlier affects the mean. The salary of $110,000 will "pull up" the mean. If students have a hard time understanding this have them visualize each salary as a rod of linking cubes. If we were going to "fair share" all the cubes the stack of 110,000 would make all the other shares get bigger as it is redistributed. Alternatively, have the students compare the mean without the outlier and compare the two results.
Which measure of center is closest to where the data is clustered?-The median and mode fall right within the cluster.
Ask students to summarize their arguments-The shape of this data is skewed right with most of the data clustered around $40,000. There is also an outlier at $110,000 which pulls the mean up. This suggests that the median or mode of $40,000 is a better measure of center. This means that most employees make closer to $40,000 than $51,667.

h. If Becky wants to attract future employees with a desirable average pay, which measure of center should she use and why?
She should use the mean because $51,667 is over $10,000 more than the median, which is $40,000.

i. Becky's company does not do very well during the next year of business. She decides not to pay herself. Which measure(s) of center will not change?
The mode and median will not change.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. Bethany works as a certified nursing assistant (CNA). Her job is to record the amount of milk-based formula that each newborn baby at a hospital consumes every 2 hours. The dot plot below shows this data.



Amount of formula consumed by newborns every 2 hours (ounces)

Before jumping into the questions provided for this set of data, you may want to ask students the following questions. How many babies are at this hospital? What attribute is being measured? What unit of measurement is being used?

a. Find the mean, median, and mode for the data set. Round to the nearest tenth.
   Mean- 2.5 oz
   Median- 2.5 oz
   Mode- 2.5 oz

Be sure to discuss these measures of center in terms of the context. The *mean* means that on average each baby consumed 2.5 oz of formula every 2 hours. You could connect this to what they know about ratio and ask them "how many ounces is this per hour?". Another way to interpret the mean is as a balancing point. In this case, a mean of 2.5 suggests that for every ounce of formula consumed that is less than 2.5 there is a corresponding ounce of formula that is consumed that is greater than 2.5. If students struggle to understand this, ask them to think of ounces of formula as absolute deviations from the mean. These absolute deviations balance at the mean of 2.5. The median of 2.5 means that half of the babies consumed less than 2.5 ounces every 2 hours and half of the babies consumed more than 2.5 ounces every 2 hours. The mode means that the greatest number of babies consumed 2.5 ounces of formula every 2 hours.

b. Use the shape of the data to justify which measure of center best represents the data.
   In this case the mean, median, and mode are all the same indicating that 2.5 ounces is a good measure of center. The shape of this data is symmetric and there does not appear to be any outliers. This means that the *mean* is a good summary of the amount of formula that most newborns consume every 2 hours.

c. Why might this information be useful?
   Doctors and parents may want to know if their baby is consuming too much or too little formula for proper growth.

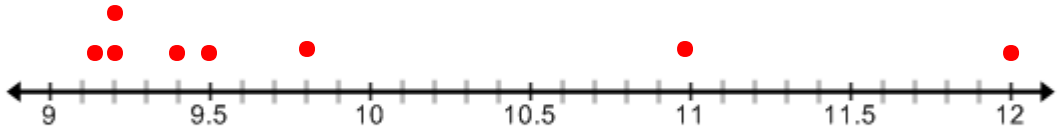
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. Raul has been training for a road bike race that takes a cyclist from the town of Logan, Utah to Jackson Hole, Wyoming. Raul's times for this exact same race the previous 8 years are shown in the table below.

Raul's Bike Race Times (Hours)							
9.2	9.15	9.5	12.08	9.2	9.4	10.9	9.8

Again, you may want students to report on the number of observations, describe the attribute being measured and the units of measurement.

a. Make a dot plot of this data. Discuss the shape of the data and use it to make a prediction about which measure of center will best summarize the data.



Raul's Bike Race Times (hours)

The shape of the data is skewed right, and most data values are clustered between 9 and 9.5 hours. The time of 12.08 hours is an outlier. Ask students what might have caused this outlier time to occur? (An injury, weather conditions, bike breakdown, or improper training) How will this outlier affect the mean? This outlier will bring the mean up.

b. What is Raul's *mean* for these previous races? What does this *mean* tell you in terms of the context? Round to the nearest hundredth.

9.9 hours. For every hour that Raul completed a race that is less than 9.9 hours there is a corresponding hour that Raul completed a race that is greater than 9.9 hours.

c. What is Raul's *median* time for these previous races? What does this *median* tell you in terms of the context?

9.45 hours; Raul completed half of his races in less than 9.45 hours and half of his races in more than 9.45 hours.

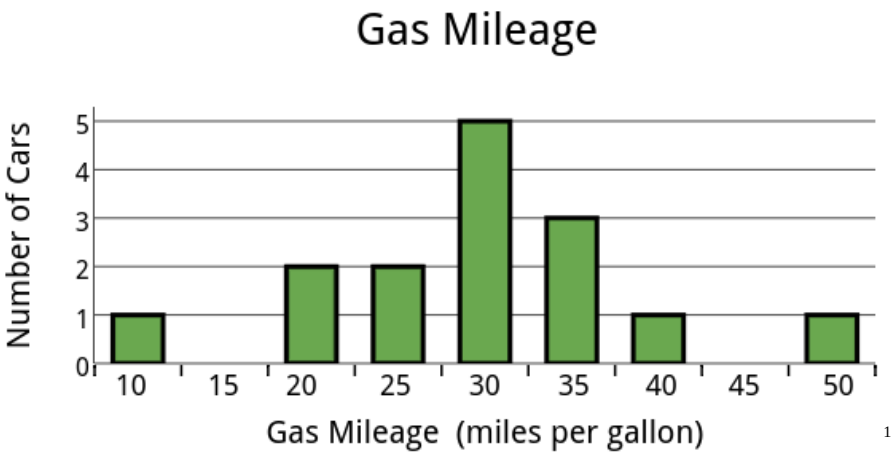
d. Which is the better measure of center for this data?

Discuss the following. Based on the shape of the data distribution, most of the data values are clustered around 9.5 hours. There are also two outliers in the data that bring the mean up. This would suggest that the median is a better measure of center because the mean does not inform us about how well Raul performs when he is not affected by extraneous conditions. The median gives us a better indication of how well Raul will do on a "good day". However, the mean is still important because it gives us an indication of how likely Raul is to complete the race in general.

A box plot might be an ideal representation of this data because you could see the outlier (a race that Raul was affected by extraneous conditions) and the times that he was able to get when he was not affected by the outlier conditions (interquartile range). You could return to this problem once students have learned how to interpret a box plot in sec. 4.3.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. The bar graph shows gas mileage of several cars. Use the shape of the data to determine which measure of center best summarizes the data. Be sure to account for clusters, peaks, gaps, and outliers. Then find this measure of center and describe what it means in terms of the context.



[1]

Discuss that if this graph were a histogram it would not be possible to find the exact measures of center. However, you can still use the shape of the data to determine which measure of center best represents the data and you can also find an interval in which the measure of center would fall. Also discuss that just because a set of data has an outlier, sometimes that outlier may not be significant enough to greatly affect the mean.

The shape of this data is fairly symmetrical. One may argue that the values of 50 and 10 mpg are outliers. The value of 10 could be a very old, large automobile and the 50 a new hybrid automobile. Regardless, the outliers "cancel" each other out because they are on opposite ends of the plot. The values peak at, and are clustered around, 30 mpg. This indicates that the mean would be a good measure of center. The mean for this data is 29.7 mpg, this means that you can expect that the average car will get around 29.7 mpg.

Note: Some students may feel overwhelmed with articulating their reasoning through writing. Encourage them to exhibit their justifications by marking or highlighting key features about the shape of the data directly on the graph or summarizing their thinking verbally to a partner or to you.

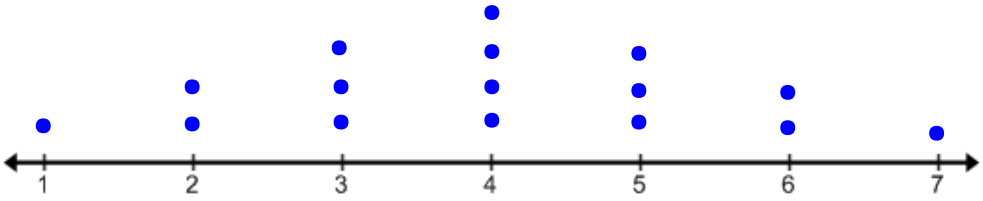
5. What does the shape of the data tell you about which measure of center to use to summarize the data?

At this point students should begin to understand through repeated reasoning that generally, if the shape of the data is fairly symmetrical, then the mean is a good measure of center. If the shape of the data is skewed or there are significant outliers then a good measure of center is the median.

1 evatipp. "Gas Mileage" Plotly Technologies Inc., https://plot.ly/368/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

6. Use the dot plot below to answer the following questions.



a. Find the mean, median, and mode.
The mean, median, and mode are all the same, they equal 4.

b. Move the least number of points so that the mean, median, and mode still remain the same value.
Students can move any two points (except points with a value of 4) that have the same absolute deviation from the mean, the same number of units away from the mean. This will not change the mean, median, or mode. For example, they might move the points at 1 and 7 to 2 and 6 respectively.

c. Move one point so that the mean is greater than the median.
Students can move any point to the right to increase the mean; the median will remain at 4.

d. Move one point so that the mean is less than the median.
Students can move any point to the left to decrease the mean; the median will remain at 4.

Spiral Review

1. For each written phrase, write a numerical expression and then evaluate your expression.
   a. Two thirds of the sum of forty-eight and twelve.
   $\frac{2}{3}(48 + 12); 40$
   b. Four times the difference of three fourths and one half.
   $4\left(\frac{3}{4} - \frac{1}{2}\right); 1$
   c. 5 less than one third of 21.
   $\frac{1}{3} \cdot 21 - 5; 2$
2. Lucy and Rachel are frosting cupcakes. Lucy can frost 5 cupcakes in 1 minute and Rachel can frost 8 cupcakes in 1 minute.
   a. Fill in the chart using the information given above.

Minutes	Lucy	Rachel
1	5	8
2	10	16
3	15	24
4	20	32

   b. Identify a pattern or rule that describes how long it takes each girl to frost cupcakes.
   Students may describe their pattern or rule in a variety of ways. Lucy frosts 5 cupcakes per minute; for every additional minute add 5 more cupcakes. Rachel frosts 8 cupcakes per minute; for every additional minute add 8 more cupcakes.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2e Homework: Choosing the Best Measure of Center

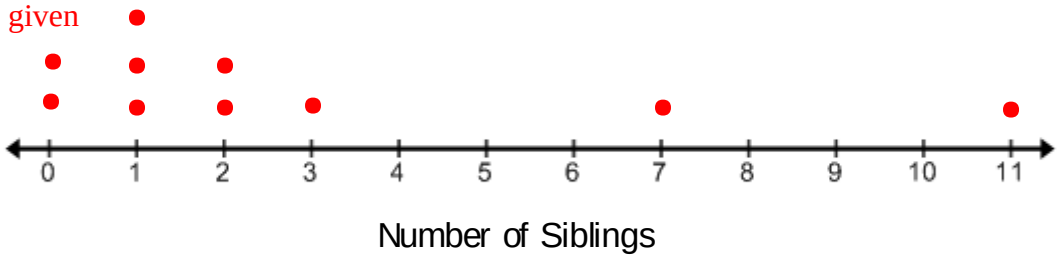
Directions: For each set of data, find the mean, median, and mode. If needed, plot the data using a dot plot, bar graph, or histogram. Then justify which measure of center you think best summarizes the data. Be sure to use the shape of the data, accounting for clusters, peaks, gaps, and outliers if needed.
Interpreting which measure of center is best can be subjective in some cases; accept arguments for either measure of center if they are accompanied with viable arguments.

1. The number of siblings for each of 10 students in a 6th grade class is shown below.

   2, 3, 7, 0, 1, 1, 11, 0, 1, 2

   Mean 2.8 siblings Median 1.5 siblings Mode 1 sibling

   Which measure of center best describes the data?
   Sample plot is given



   There are two outliers 7 and 11. These outliers pull the mean up. The *mean* communicates that the average family for these 10 students has about 2.8 siblings, however it does not account for the fact that most of these siblings are only in two families. The dot plot of the data shows that it is skewed right and that most of the data is clustered around 1. These shape features indicates that the median is the best measure of center for the data. The median of 1.5 tells us that half of the students have less than 1.5 siblings and half have more than 1.5 siblings.

2. The graph shows the price of the same running shoe at several different stores.

[1]

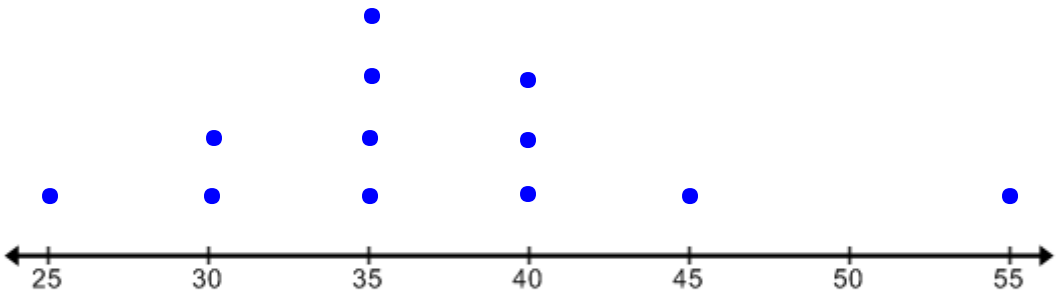
   Mean Not possible to calculate Median Not possible to calculate Mode Not possible to calculate

   Which measure of center best describes the data?
   The median will most likely best describe this data set because it is skewed left with 2 outliers between $40 and $60. Most of the data is clustered around $80 to $90. The median is between $80 and $90.

1 evatipp. "Price of Running Shoes" Plotly Technologies Inc., https://plot.ly/375/~evatipp (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. The dot plot shows the length of lunch breaks at several different middle schools.



Length of Lunch Breaks (minutes)

Mean 37.08 min Median 35 min Mode 35 min

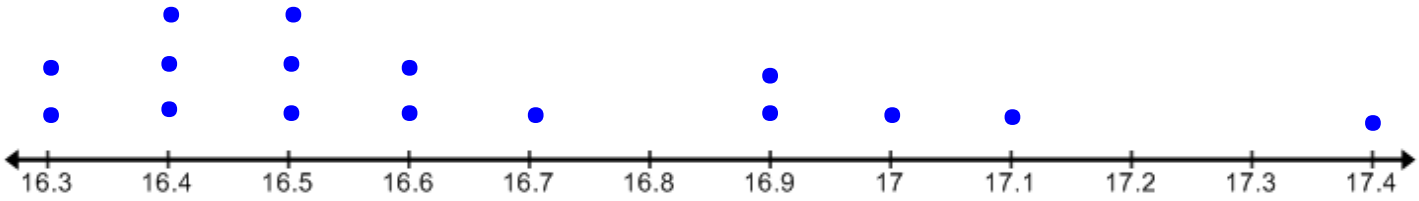
Which measure of center best describes the data?
The mean or the median can be used to summarize this data. The shape of the data is fairly symmetrical and there are no major outliers. Then both the *mean* and median suggest that the typical lunch break for these middle schools is approximately 35 to 37 minutes long.

4. In your homework assignment for **Lesson 4.1b** you collected data for a statistical question that you created and asked to several people in your class. You made a dot plot of the data and analyzed the shape of the data. Return to this data and find its mean, median, and mode. Then use the shape of the data to determine which measure of center best summarizes the data. Be sure to justify your answer.

Mean answers will vary Median answers will vary Mode answers will vary

Which measure of center best describes the data?
Answers will vary

5. Addison has just completed a 100-Yard Dash with a time of 16.6 seconds. Circle the statements below that are true.



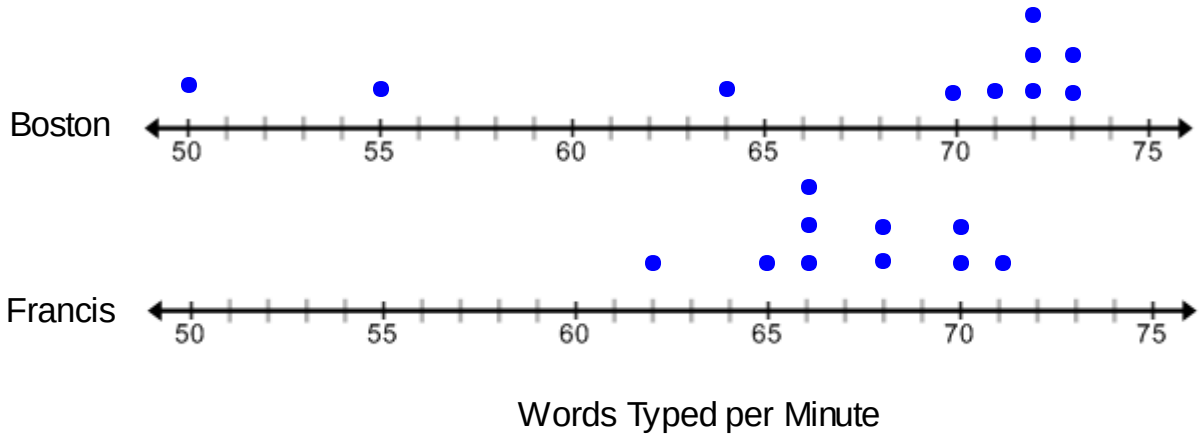
100 Yard Dash Times (seconds)

A. Addison's time is greater than the mode of the data.
B. Addison's time is less than the median.
C. Addison's time is greater than all the measures of center.
D. Addison's time is within one-tenth of the mean.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

The problems on this page may be more challenging for your students. You may want to make them bonus problems.

6. Boston and Francis are practicing their keyboarding skills. They have each taken a practice test 10 times. Their scores are recorded in the dot plots below. They both have a **mean** typing score of 67.2 words per minutes but different median scores.



a. Find the median words per minute typed for each person.
Boston's median is 71.5 words per minute; Francis' median is 67 words per minute.

b. To qualify to move on to the next class, you must have a final test score with a typing speed of at least 69 words per minute. You only get to take the final test once. Which person is more likely to pass the test if they take it right now? Justify your answer. Again interpretation of the best measure of center can be subjective, accept all answers that are thoroughly justified.
Suggested answer: Boston is more likely to pass the test; even though she has a mean typing score of 67.2 words per minute her median score is 71.5 words per minute. This means that half of her scores were greater than 71.5 words per minute. Francis has a median of 67, this is not close enough to the minimum requirement of 69 words per minute.

7. Use the data given to answers the questions that follow.

8, 2, 2, 5, 3, 5, 3,

a. Find the mean and median for this set of data
Mean- 4
Median- 3

b. Suppose each data value was increased by 10. How would this change the mean, median, and mode?
Each measure of center would increase by 10. The mean would 14 and the median would be 13.

c. Suppose each data value was decreased by 2. How would this change the mean, median, and mode?
Each measure of center would decrease by 2. The mean would be 2 and median would be 1.

d. Find one data value that if added to this set of data will keep the median the same.
Any value that is less than or equal to 3

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.2f Self Assessment: Section 3.2

Consider the following skills/concepts. Rate your comfort level with each skill/concept by checking the box that best describes your progress in mastering each skill/concept. Corresponding sample problems, referenced in brackets, can be found on the following page.

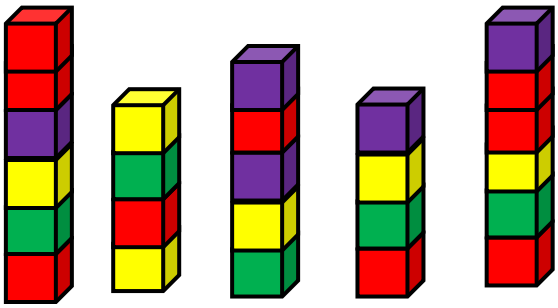
Skill/Concept	Minimal Understanding 1	Partial Understanding 2	Sufficient Mastery 3	Substantial Mastery 4
1. Understand the mean as a "leveling out" of data or a fair share. Also understand how the mean can be interpreted as the "balancing point" for the data values.	I can kind of explain how the mean describes a fair share of the data. I do not know how it can be interpreted as the "balancing point".	I can explain how the mean describes a fair share of the data. I can find the absolute deviations of the values in the data set but I do not know how they relate to the mean.	I can explain how the mean describes a fair share of the data. I struggle to relate this to the formula for the mean. I can show how the mean is the "balancing point" of the data by comparing the sum of the absolute deviations of the values to the right and left of the mean.	I can explain how the mean describes a fair share of the data and relate it to the formula for finding the mean. I can show how the mean is the "balancing point" of the data by comparing the sum of the absolute deviations of the values to the right and left of the mean.
2. Find the mean from a set of data and interpret its meaning for a given context.	I can find the mean from a set of data but don't know what it means if the *mean* is not a whole number.	I can accurately find the mean from a set of data that is presented as a list.	I can accurately find the mean from a set of data that is represented in a variety ways. I can interpret what the *mean* means for a given context.	I can accurately find the mean from a set of data that is represented in a variety ways. I can interpret what the *mean* means for a given context and give suggestions for how the values were collected.
3. Find the median and mode from a set of data and interpret their meaning for a given context.	I know what the median and mode are but do not know how to find them.	I know how to find the median if there is an odd number of data values. I do not know how to find the median if there is an even number of data values. I can find the mode if there is only one.	I can find the median and mode from a set of data that is presented in a list. I understand that there can be more than one mode or no mode at all. I can interpret what the median and mode mean for a given context.	I can find the median and mode for any set of data that is represented in a variety of ways. I understand that there can be more than one mode or no mode at all. I can interpret what the median and mode mean for a given context and given suggestions for how the values were collected.
4. Determine the best measure of center for a given data set and justify why it is the best measure of center.	I know that there are different measures of center but don't really know how to choose one that best represents the data.	I can sometimes guess which measure of center is best but don't really know how to determine which measures it best for any given set of data.	Given a set of data, I can choose the measure of center that I think best represents the data. I struggle to construct good arguments and give justification as to why I think the measure of center best represents the data.	Given a set of data, I can choose the measure of center that I think best represents the data. I can give justification and construct good arguments as to why I think the measure of center best represents the data.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Sample Problems for Section 4.2

Square brackets indicate which concept/skill the problem aligns to.

1. A group of five students have each grabbed a handful and linking cubes. They each stick their cubes together into a rod. These rods are shown below. [1]



a. Show or explain how you can equally redistribute the blocks so that everyone gets the same number of blocks.

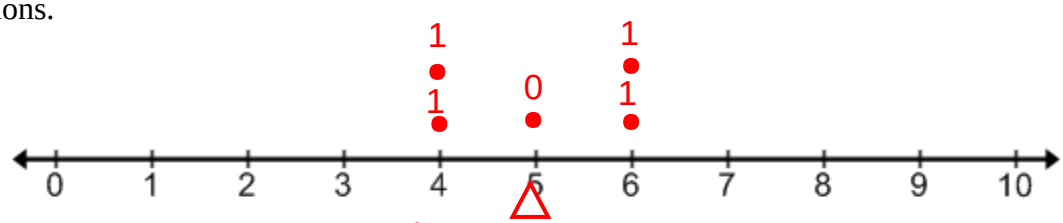
Students might suggest that you can move one block from each rod of 6 to each rod of 4. This makes it so each rod has 5 blocks. They could also combine all of the rods together making a total of 25 blocks and then redistribute the blocks into 5 equal groups of 5 blocks. Accept answers that are written arguments as well as pictorial or graphical arguments.

b. Find the *mean* number of blocks there are per handful using the mean formula and explain how this relates to the fair share or unit rate.

$$6 + 4 + 5 + 4 + 6 = 25 \rightarrow \frac{25}{5} = 5$$

This shows combining all the blocks together to get a sum of 25 and then redistributing or dividing into 5 equal groups.

c. Verify your *mean* by plotting each number of blocks in a handful on a number line and showing that the *mean* is the balancing point. Be sure to discuss how you know it is balanced with absolute deviations.



The absolute deviations on each side of the mean both sum to 2. We know that the plot is balanced at 5 because the sum of the absolute deviations on each side of 5 are equal.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. Find and interpret the mean, median, and mode for each set of data below. Then determine which measure of center best represents the data. Be sure to justify your answer. [2][3][4]

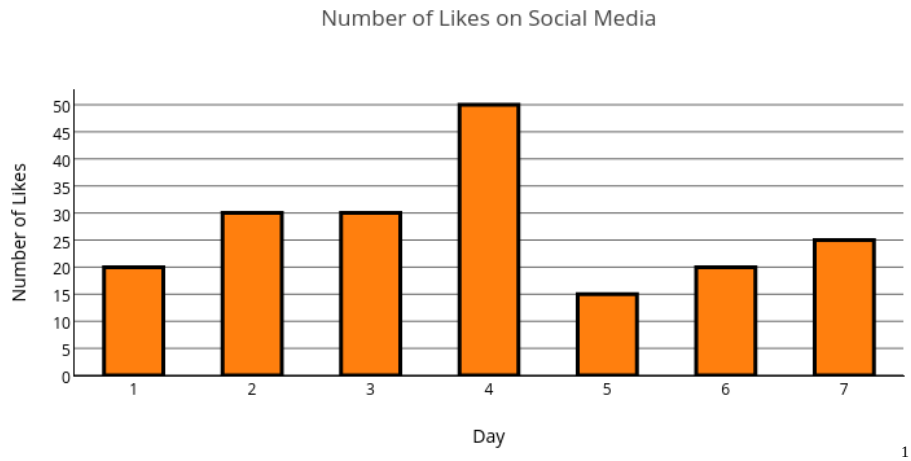
a.

Number of sit ups			
78	86	86	95
90	71	110	102
92	80	106	100

$Mean = 91.3$; The average number of sit ups is 91. $Median = 91$; Half of the values in the data set are above 91 and half are below 91. $Mode = 86$; 86 sit ups occur the most often. This data was likely collected by counting the number of sit ups that several different people can do during a given time interval.

Students may argue that either the mean or median is the best measure of center since these values are almost the same value. Also the data is fairly symmetrical and there are not any major outliers that affect the data.

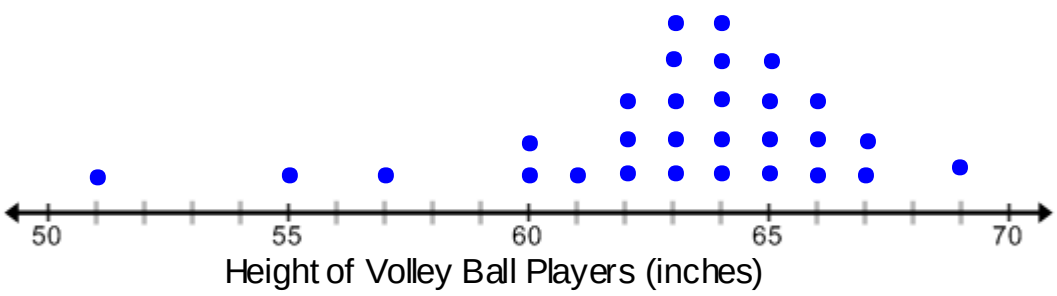
b.



[1]

$Mean = 27.1$; The average number of likes per day is 27.1. $Median = 25$; Half of the number of likes is less than 25 and half of the time the number of likes is more than 25. $Mode = 20$ and 30 likes occur the most often. This data was most likely collected by counting the number of likes posted on a social media website per day. For this data set, the median is probably the best measure of center because the value of 50 is an outlier that will pull the mean up.

c.



$Mean = 62.97$; The average height of a volleyball player is about 63 inches. $Median = 64$; Half of the players are less than 64 inches tall and half of the players are more than 634 inches tall. $Mode = 63\ and\ 64$; The greatest number of players are 63 and 64 inches tall. This data was most likely collected by measuring the height of several players on a volleyball team. Students might argue that the median is the best measure of center because the data is skewed left, which pulls the mean down. They might also argue that the mode is the best measure of center because it includes values of 63 and 64 inches.

1 evatipp. "Number of Likes on Social Media" Plotly Technologies Inc., https://plot.ly/375/~evatipp/# (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Section 4.3: Meaures of Variability

Section Overview:

At this point in the chapter, student's ability to interpret and analyze data should be growing. In addition to using shape and center to describe a data set, they can use variability. In this section, students will learn how variability is an important component to fully understanding a data distribution. Finding the range is one way that students can discuss the variability of a data set. However, it does not tell us what is happening with the values within the data set itself. For distributions where the mean is a better measure of center, variability is commonly measured in terms of how far the values deviate from the mean. Students will use the work they did in the previous section with absolute deviations to calculate the mean absolute deviation. This will set the stage for work that they will do in later years with standard deviation. In cases where the median is a better measure of center, students will use the range and interquartile range (IQR) to describe the measure of variability. The IQR is a numerical measure that gives the spread of the middle 50% of the data points. Finally, students will learn how to construct and analyze a box plot which gives a snap shot of the 5-number summary and can help students to visualize the shape of the data. The numerical measures of spread and box plots aid in the comparison of data as students will be able to quickly see how two or more data sets relate to one another.

Primary Concepts and Skills to Master in this Section:

By the end of this section, students should be able to:

1. Find and interpret the range for a data set.
2. Find and interpret the mean absolute deviation (MAD) for a data set.
3. Find and interpret values of the 5-number summary for a data set.
4. Find the interquartile range (IQR) for a data set and interpret its meaning.
5. Display numerical data in a box plot, use key terms to describe its shape, and interpret the distribution of data.
6. Draw conclusions about two or more data sets by comparing their box plots.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3a Class Activity: Variability and Mean Absolute Deviation

This lesson will take two days.

The problems in this lesson help students learn that they must look at other factors, in this case variability, which can affect the interpretation of data. This helps them to understand that they must look at all angles when solving a problem and take them into account when creating solutions.

1. Olivia has been offered a job in both Salt Lake City and San Francisco. She is trying to decide which city she would rather live in and climate is very important to her. She has heard friends that live in Utah talk about the great skiing and friends that live in California talk about the nice warm beaches. She decides to look at the average monthly temperature in degrees Fahrenheit for each state.

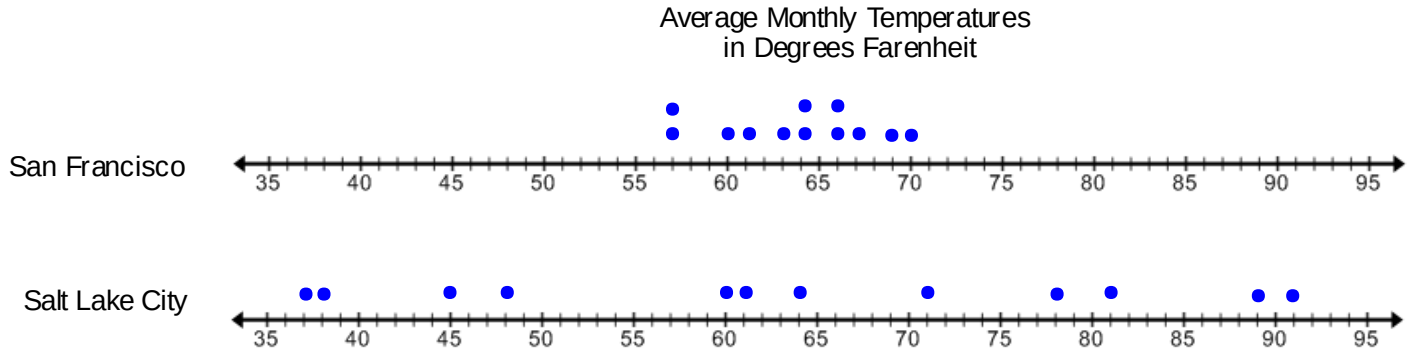
City	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Salt Lake City	37°	45°	60°	61°	71°	82°	91°	89°	78°	64°	48°	38°
San Francisco	57°	60°	61°	63°	64°	66°	66°	67°	70°	69°	64°	57°

a. How would you describe the temperatures for each city?
The monthly temperatures for Salt Lake City are more spread out; they change a lot throughout the year. The monthly temperatures for San Francisco are less spread out; they do not change very much throughout the year.

b. Find the mean temperature for each city. Round your answer to the nearest degree.
The mean temperature for Salt Lake City is 64° and the mean temperature for San Francisco is also 64 °.

c. Olivia is surprised that the mean monthly temperature for both cities is the same. Using the mean temperature as the only factor, where would you tell Olivia to move to.
Based on the mean temperature alone, it does not matter where Olivia moves to because they are the same.

d. Are there any other factors relating to the monthly temperatures that can help Olivia make her decision? Use the dot plots below to answer this question.



Analyze the dot plots at a class. Ask students about the shape of the data; is it symmetrical or skewed? The data for each plot is fairly symmetrical. Are there any outliers that would affect the mean? There are not any outliers. The answers to these questions suggest that the mean is a good measure of center for this data. However, it may not be the best/only numerical measure to consider when describing the data distribution.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Students will most likely point out that the data for Salt Lake City is more spread out and the data for San Francisco is more clustered together. Remind students what the range is for a data set and find the range for each city.

The *range* of a data set is the difference between the greatest value (maximum) and the smallest value (minimum).

$$\boldsymbol{Range = Max - Min}$$

e. Find the range for each city.

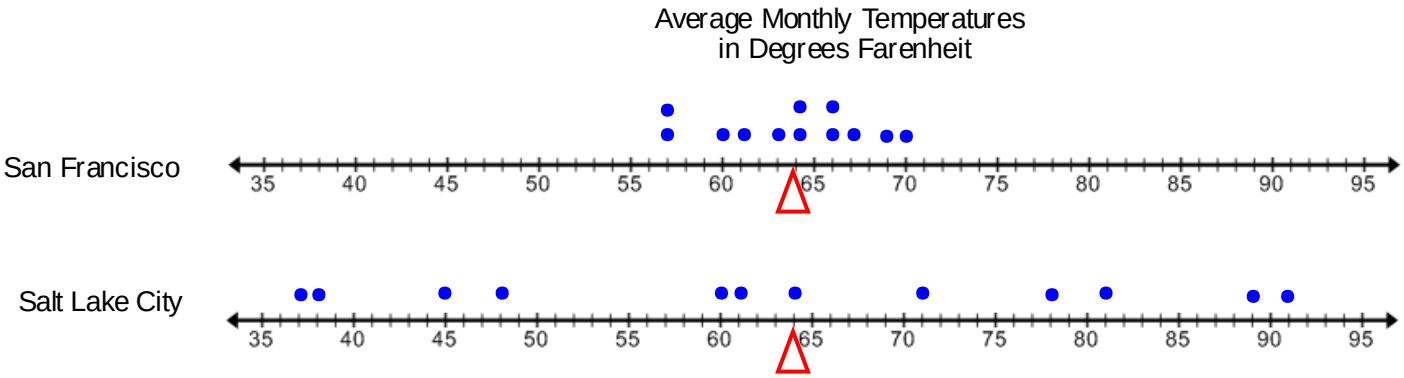
$\boldsymbol{SLC\ Range = 91 - 37 = 54}$

$\boldsymbol{SF\ Range = 70 - 57 = 13}$

Discuss that the range, just like mean and median, is a specific number that tells us something about the data distribution. It is not a measure of center though; it is a measure of ***variability*** or ***spread***. The range tell us the temperatures for SLC vary by 54 degrees and the range for SF tells us that its temperatures only vary by 13 degrees.

Range is not the only measure of variability. ***Variability*** is used in statistics to tell us how spread out data is from some focal point. We can also look at how spread out the data is from a measure of center, in this example the mean.

f. Mark the mean on both of the dot plots with a balance triangle. Describe the variability of the monthly temperatures to their mean for each city; determine which city has more variability in its average monthly temperatures.



The data values for SF are clustered closer to the mean. The data values for SLC are more spread out from the mean. SLC has more variability than SF in terms of temperature.

g. How does determining the level of variability in the data help Olivia make her moving decision?

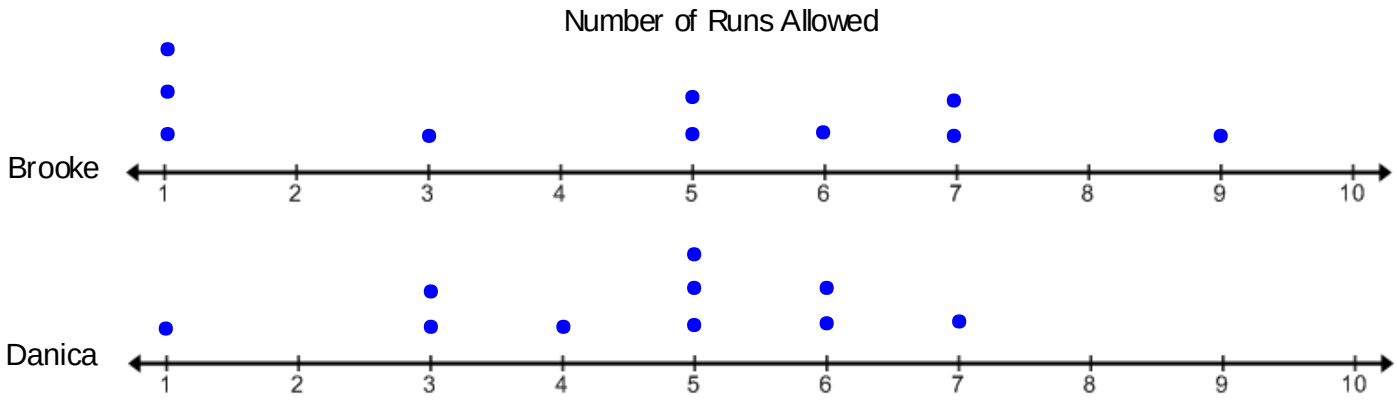
If Olivia were to only look at the mean she would conclude that there is no difference in the monthly temperatures for each city. By looking at the variability, she can see that throughout the year SF temperatures do not vary much away from 64 degrees. SLC on the other hand has temperatures that vary a lot from the mean. When she makes her decision, she must consider whether she wants variability in temperatures or not.

h. For which data set is the mean a better indicator of a typical value and why?

The mean is a better indicator for the SF data set because it has less variability.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. A coach is trying to decide which softball pitcher he should play for their next game. He is looking at the numbers of runs that each of his two pitchers have allowed in their last 10 games. The data for each pitcher is shown on the dot plot below.



a. Find the mean number of allowed runs for each pitcher.
Brooke-4.5 runs allowed
Danica-4.5 rules allowed

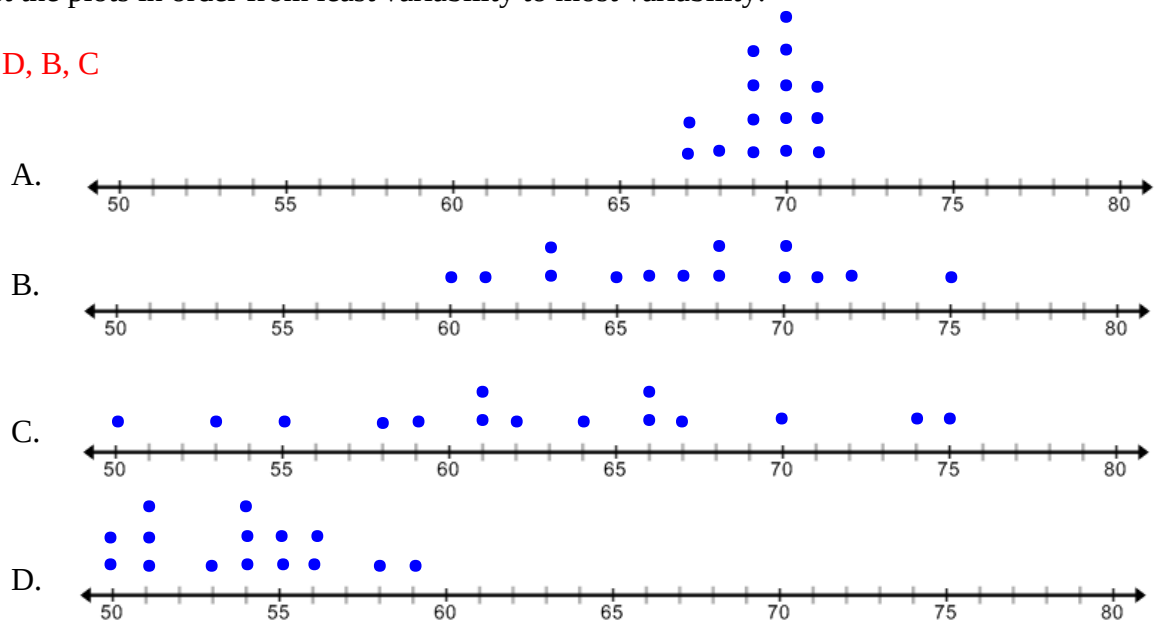
b. For which distribution does the mean give a better indicator of a typical value? Explain your answer.
Danica's data give a better indicator for the mean because there is less variability.

Discuss with students that the mean is only a precise indicator of an attribute if the variability in the data is low.

c. The softball coach states that both pitchers have the same mean number of runs allowed for their previous ten games so it does not matter which girl he should play. Is he correct in his thinking? Explain.
The coach is not correct. Although both pitchers have the same mean, Danica has less variability in her data. This means that she is more consistent in the number of runs she allows. This information should be used in the coach's decision.

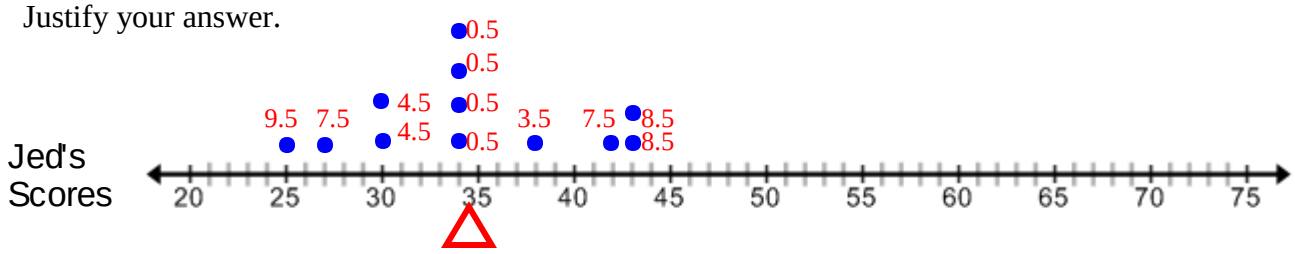
3. List the plots in order from least variability to most variability.

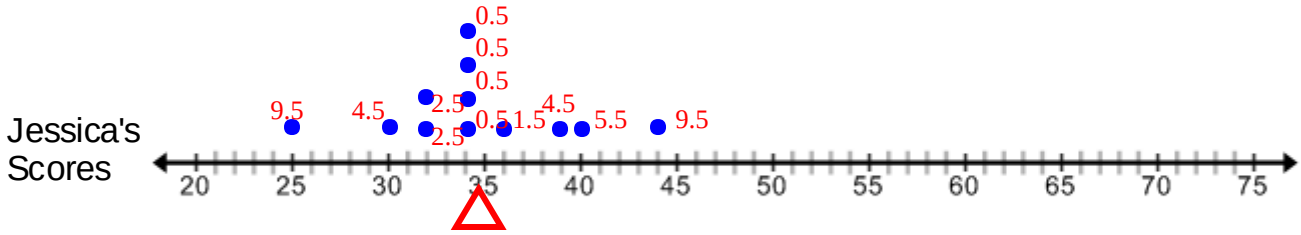
A, D, B, C



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. The plots below show the scores for two skateboarders from their last 12 skate boarding competitions. The mean score for both athletes is **34.5**. Which athlete appears to have less variability in their scores? Justify your answer.





Students might argue that Jed has less variability because the range for his scores is 18 and the range for Jessica's scores is 19. On the other hand, Jessica does have a lot of scores clustered around the mean of 34.5. Students should arrive at the conclusion that it is hard to compare the amount of variability between these two plots because they are quite similar. This presents the need for a better way of analyzing variability besides just looking at the shape of the data.

In the problem above both plots shows similar variability. Recall that variability shows us how spread out the data is from some focal point, in this case the mean. In previous lessons, we used **absolute deviations** to analyze how far away each data point is from the mean.

5. Use what you know about **absolute deviations** to try and find a better way to measure the variability for Jed's data. Review with students that an absolute deviation is the distance between the balancing point or mean and each data point. Talk about how absolute deviations relate to spread or variability.

Develop a method for finding the mean absolute deviation (MAD) by discussing the problems below.

- Ask them to analyze the absolute deviations, how can we summarize the absolute deviations for the data set? (Find the average absolute deviation).
- How do we find the average absolute deviation for the data set?
  a. Find the absolute deviations for each data value from the mean for Jed's data. Record them above each value on the dot plot.
     See plot above. If students struggle finding the absolute deviations on the plot encourage them to organize the data in a table that shows how you can find the absolute deviations with subtraction.
  b. Find the sum of the absolute deviations and divide by the total number of observations.
     $9.5 + 7.5 + 4.5 + 4.5 + 0.5 + 0.5 + 0.5 + 0.5 + 3.5 + 7.5 + 8.5 + 8.5 = 56$
     $\frac{56}{12} \approx 4.7$

We now have one number for Jed's scores that represents the average deviation of the data points from the mean. This is called the **mean absolute deviation (MAD).**

  c. Find the mean absolute deviation for Jessica's data.
     $9.5 + 4.5 + 2.5 + 2.5 + 0.5 + 0.5 + 0.5 + 0.5 + 1.5 + 4.5 + 5.5 + 9.5 = 42$
     $\frac{42}{12} = 3.5$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

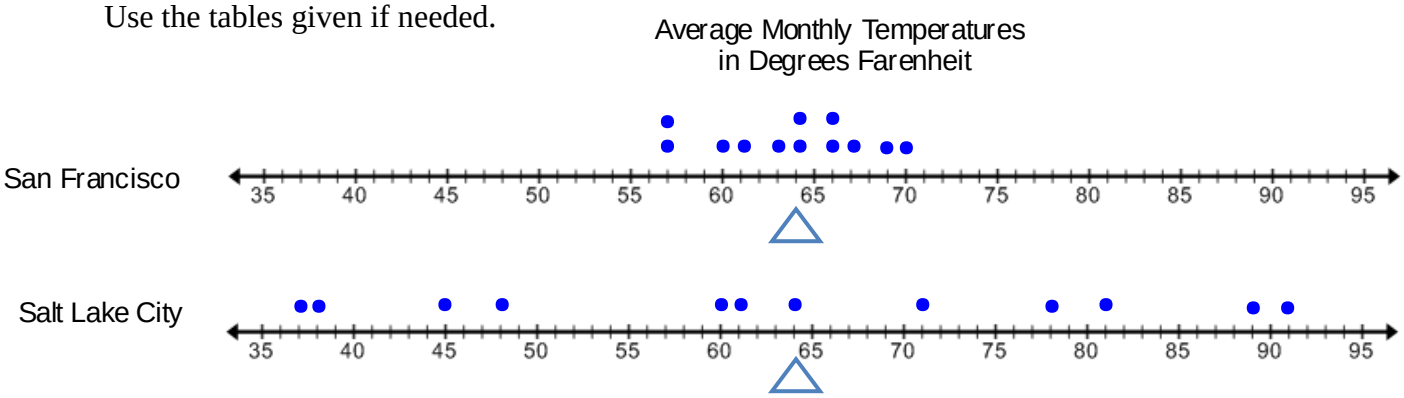
d. What does the mean absolute deviation (MAD) for each skate boarder mean?
Jed's MAD is 4.7, meaning that Jed's scores are on average 4.7 points away from his mean score of 34.5.
Jessica's MAD is 3.5, meaning that Jessica's scores are on average 3.5 points away from her mean score of 34.5

e. Which athlete has a higher MAD? What does this mean?
Jed has a higher MAD. This means that he has slightly more variability in his data than Jessica.

Mean Absolute Deviation, or _MAD_, is a measure of variation. It is computed by finding the _mean_ of the _absolute deviations_ in the distribution.

You may want students to write down that an absolute deviation is the positive distance from each data value to the mean.

6. Find the Mean Absolute Deviation for the Salt Lake City and San Francisco Temperature Data given earlier. The dot plots for both cities are below, recall that the mean temperature for both cities is 64°. Use the tables given if needed.



Encourage students to use the tool that best suits them to find the MAD. They might prefer to find and mark the absolute deviations on the dot plot or they can organize the data in a table.

San Francisco Average Monthly Temperatures

Temperature	Expression that shows counting distance	Absolute deviations
57	64 – 57	7
57	64 – 57	7
60	64 – 60	4
61	64 – 61	3
63	64 – 63	1
64	64 – 64	0
64	64 – 64	0
66	66 – 64	2
66	66 – 64	2
67	67 – 64	3
69	69 – 64	5
70	70 – 64	6

Sum of the Absolute Deviations: 40

MAD: $\frac{40}{12} \approx 3.3$

Salt Lake City Average Monthly Temperatures

Temperature	Expression that shows counting distance	Absolute deviations
37	64 – 37	27
38	64 – 38	26
45	64 – 45	19
48	64 – 48	16
60	64 – 60	4
61	64 – 61	3
64	64 – 64	0
71	71 – 64	7
78	78 – 64	14
81	81 – 64	17
89	89 – 64	25
91	91 – 64	27

Sum of the Absolute Deviations: 185

MAD: $\frac{185}{12} \approx 15.4$

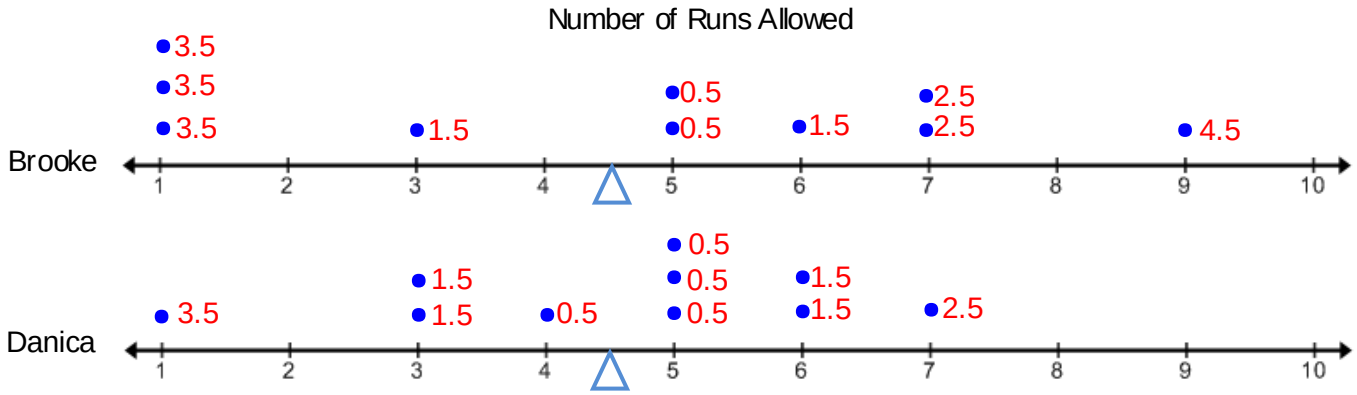
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

7. What does the mean absolute deviation for each set of data tell us?
The MAD for SF is 3.3. This means that the monthly temperatures for SF have an average deviation from the *mean* of about 3.3 degrees. The MAD for SLC is 15.7. This means that the monthly temperatures for SLC have an average deviation from the *mean* of about 15.4 degrees. This confirms that the monthly temperatures for SLC have much more variability than SF.

8. Fill in the blanks in the statement below with the words "greater" or "smaller".

The mean absolute deviation describes how each data value varies from the mean. The higher the MAD the __greater_____ the variability, the lower the MAD the _____smaller_____ the variability.

9. Find the mean absolute deviation (MAD) for Brooke and Danica's softball data. Recall the *mean* number of runs allowed for both pitchers is 4.5.



Some students might point out that you can just find the sum of the absolute deviations on one side of the mean and double it to find the total sum of all the absolute deviations. This is possible because we know that the sums of the absolute deviations on each side of the mean are equal.

Brooke's MAD: $\frac{3.5+3.5+3.5+1.5+0.5+0.5+1.5+2.5+2.5+4.5}{10} = \frac{24}{10} = 2.4$

Danica's MAD: $\frac{3.5+1.5+1.5+0.5+0.5+0.5+0.5+1.5+1.5+2.5}{10} = \frac{14}{10} = 1.4$

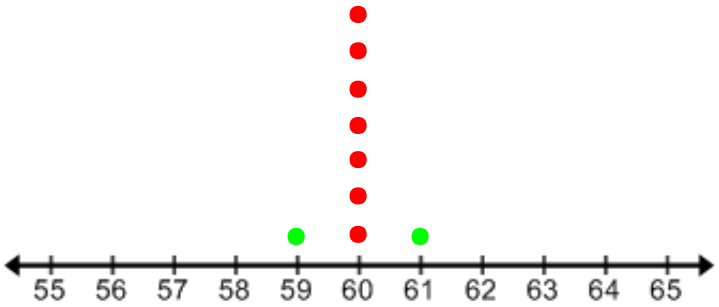
Take time to discuss that the MAD for each pitcher tells you the average deviation of data values from the *mean*. Danica's MAD is smaller so she has less variability in her distribution. The *mean* is a better indicator of a typical value for Danica's distribution than Brooke's distribution.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

10. Use the data given below to answer the following questions.

60, 60, 60, 60, 60, 60, 60

a. Plot the data on a dot plot



b. Find the mean and the MAD of the data. Interpret what the MAD tells us about the amount of variability in the data.
The mean is 60 and the MAD is zero. This means that there is no variability in the data.

c. Suppose an additional data value of 61 is added to this data set. Add this value to the dot plot. Without doing any calculations determine how this will affect the mean? How will this affect the MAD?
It will make the mean go up slightly. The MAD will also go up because now there is variability in the data.

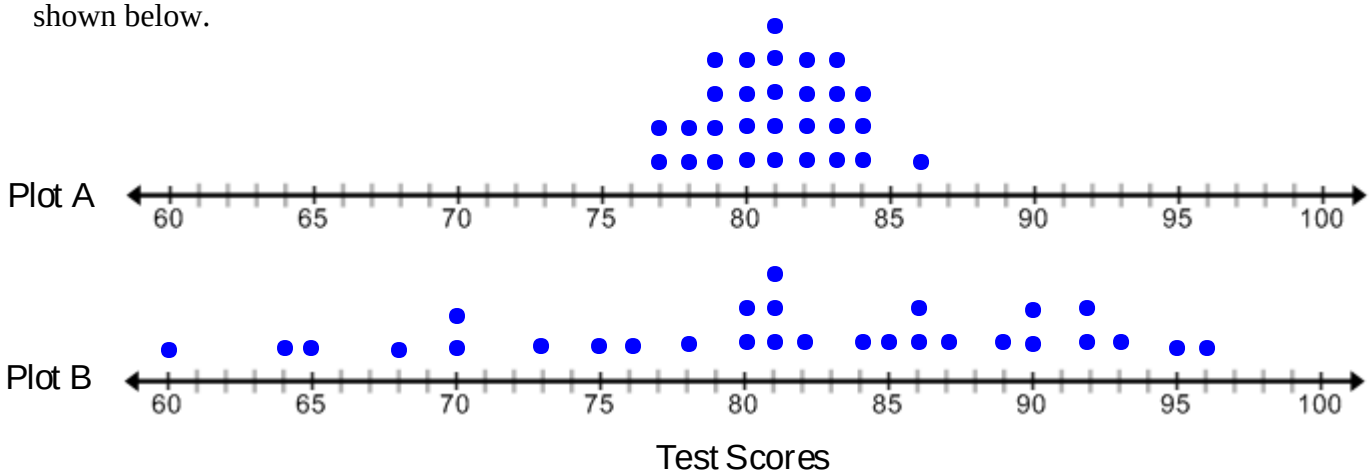
d. Now suppose that in addition to adding 61 to the data set a value of 59 is added as well. Without doing any calculations determine how this will affect the mean? How will it affect the MAD?
This will make the mean go back to 60. The MAD will go up even more because adding an additional point that is not 60 will add even more variability.

e. Suppose an additional value was added at 65. What single value must also be added to the plot to keep the mean at 60? Explain your answer.
You must add a value on the left side of the mean that has the same absolute deviation from the mean as 65. Since the absolute deviation for 65 is 5, you would need to add a value of 55.

f. Suppose an additional value was added at 64. What two additional values must be added to the plot to keep the mean at 60? Explain your answer.
You must add values on the left side of the mean that have absolute deviations that sum to the same value of the absolute deviation for 64. 64 has an absolute deviation of 4, so any two values to the left of the mean that have absolute deviations that sum to 4. Answers may include values at 59 and 57 or 58 and 58.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

11. Raphael has just received his test scores for English and Science. The bell has just rung to go home and he is in too much of a hurry to write down his scores. His teacher gives him a paper on the way out the door that shows a dot plot of the class scores (out of 100) for each test. The dot plots are not labeled but Raphael's teacher told the class that both tests had the same average score of 81. The dot plots are shown below.



a. Find the range of data for each plot.

Plot A Range: $86 - 77 = 9$

Plot B Range: $96 - 60 = 36$

b. Without doing any calculations, which dot plot represents a class with a higher MAD? Explain.
Dot plot B represent class scores with a higher MAD because is clearly has more variability than dot plot A.

If needed, encourage students to put a triangle on the plot where the mean is.

c. On the way home from school Raphael is thinking about his tests scores and can't remember them. He does remember that his teacher told him that he received one of the high scores in the class for his English test. Which dot plot would Raphael want to be the distribution of scores for his English test if he had one of the high scores? Justify your reasoning.
Raphael would most likely want dot Plot B to represent the English scores because the high scores for this plot are in the 90s. The highest score for plot A is only 86.

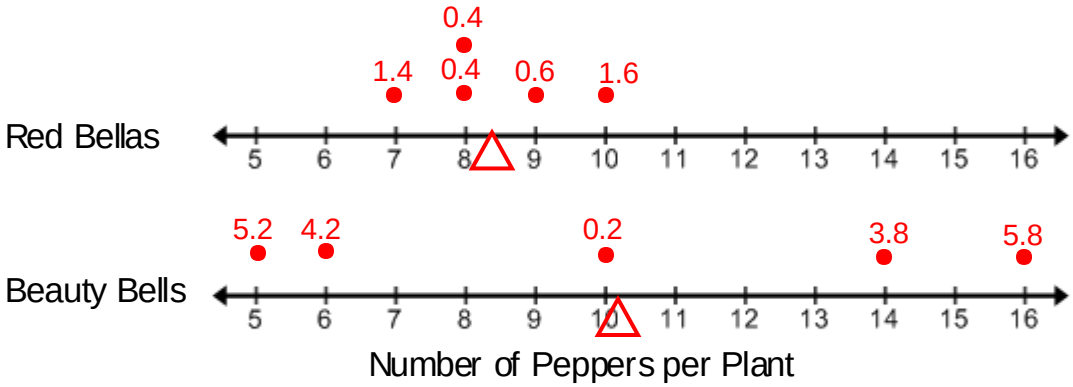
d. He also remembers that his score on his science test was below average. Which dot plot would Raphael want to be the distribution of scores for his science class? Justify your reasoning.
Raphael would most likely want dot Plot A to represent the Science scores because the scores below the average are all within a range of $77 - 80$. The scores below the average for plot B range from 60 to 80.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

12. Claudia loves fresh peppers from the garden. She has decided to plant two varieties of bell peppers, Red Bellas and Beauty Bells. For each variety, she plants 5 plants and grows them in the same conditions. The table below shows the number of peppers that each of pepper plants produced.

Plant	Red Bellas	Absolute Deviations	Beauty Bells	Absolute Deviations
1	7	1.4	14	3.8
2	8	0.4	10	0.2
3	10	1.6	5	5.2
4	8	0.4	6	4.2
5	9	0.6	16	5.8

a. Draw a dot plot to represent the number of peppers produced for each plant.



b. Find the *mean* number of peppers produced for each variety of pepper.

Red Bellas *mean*: ≈ 8.4

Beauty Bells *mean*: ≈ 10.2

c. Find the MAD for each variety of pepper. Use the table or the dot plot to track the deviations.

Red Bellas MAD: $\frac{1.4+0.4+0.4+0.6+1.6}{5} = 0.88$

Beauty Bells MAD: $\frac{5.2+4.2+0.2+3.8+5.8}{5} = 3.84$

d. Claudia is only going to plant one kind of pepper plant next season and she only cares about the number of peppers a plant will produce. What pepper plant should she choose next season? Use the *mean* and MAD for each variety of pepper to justify your answer.
Accept good arguments for either variety. If Claudia wants consistency in the number of peppers produced, she should choose Red Bellas. Even though it has a smaller mean than Beauty Bells, there is less variability in the number of peppers produced. If she is a risk taker she should choose Beauty Bells. This kind of pepper has a higher average, but has more variability. She might end up with a "bumper" crop or get as little as 5 or 6 peppers.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Spiral Review

Multiplying Decimals

1. 5×3.21
   16.05
2. 8×0.001
   0.008
3. 3.012×3
   9.036
4. 3.2375×5
   16.1875

Add and Subtract Decimals

5. $0.8 + 0.007 - 0.004$
   0.803
6. $5 + 0.682 - .03$
   5.652
7. Four plus seventy-three thousandths minus four hundredths
   4.033
8. Four and three tenths minus five hundredths.
   4.25

Measurement Conversions

9. $1\ ft =$ __12__ in
10. $2.5\ ft =$ ___30____ in
11. $54\ in =$ __4.5____ ft
12. $36\ in =$ __3__ yd

Finding Volume

13. Find the volume of a rectangular prism that has a length of 7 inches, a width of 5 inches, and a height of 9 inches.
    $V = 315\ in^3$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3a Homework: Variability and Mean Absolute Deviation

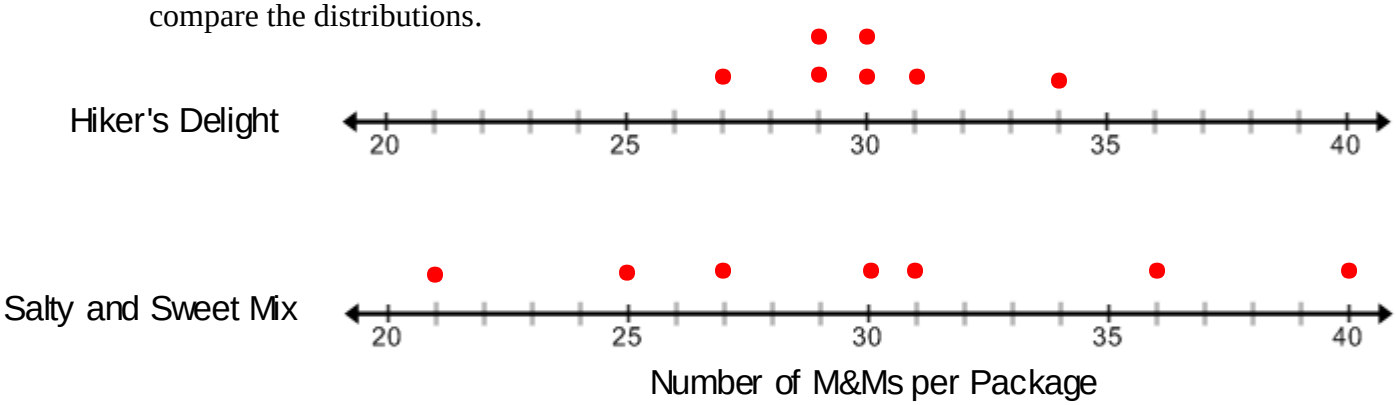
Note: There are a variety of homework problems provided for this lesson. Choose the ones that best meet the needs of your students.

1. You are trying to decide what brand of trail mix to buy. You think that the best trail mix has a lot of M&Ms in it. The table below shows the amount of M&Ms in 7 randomly selected bags of each brand of trail mix.

Trail Mix Bag	1	2	3	4	5	6	7
Hiker's Delight	30	27	34	29	30	31	29
Salty and Sweet Mix	36	40	21	27	31	25	30

a. What is the range of the number of M&M's for each brand of trail mix?
Hiker's Delight range: $34 - 27 = 7$
Salty and Sweet Mix range: $40 - 21 = 19$

b. Make a dot plot for each set of data. Be sure to use the same scale for each plot so that you can compare the distributions.



c. Find the *mean* number of M&M's for each brand of trail mix.
Hiker's Delight *mean*: 30
Salty and Sweet Mix *mean*: 30

d. Use the table below to find the MAD for each set of data. Explain what the MAD means for this context.

Hiker's Delight	30	27	34	29	30	31	29
Absolute deviation	0	3	4	1	0	1	1
Salty and Sweet Mix	36	40	21	27	31	25	30
Absolute deviation	6	10	9	3	1	5	0

Hiker's Delight MAD: $\frac{0+3+4+1+0+1+1}{7} = \frac{10}{7} \approx 1.43$ This means the average absolute deviation of data values from the *mean* is approximately 1.43. This data has less variability than the data for the Salty and Sweet Mix.

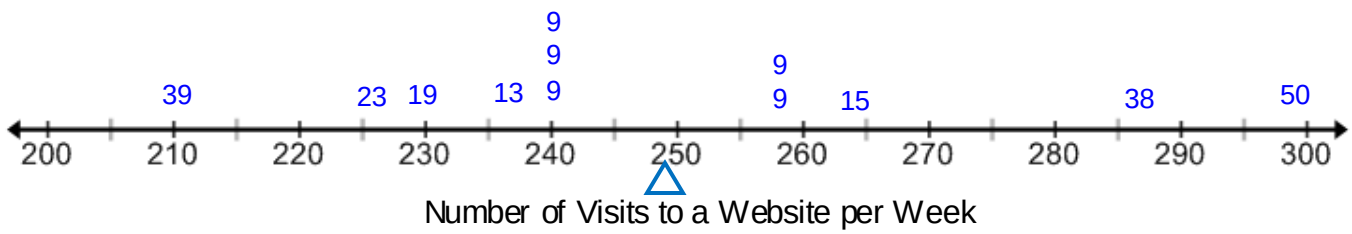
Salty and Sweet Mix MAD: $\frac{6+10+9+3+1+5+0}{7} = \frac{34}{7} \approx 4.86$ This means the average absolute deviation of data value from the *mean* is approximately 4.86. This data has higher variability than the data for Hiker's Delight.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

e. For which distribution is the mean a better indicator of a typical value? Explain your answer.
The MAD for Hiker's Delight is less than the MAD for Salt and Sweet Mix. This means that the mean for Hiker's Delight is a better indicator of a typical value of the data.

f. Considering variability, which brand of trail mix do you prefer? Explain your answer.
Students that prefer Hiker's Delight might argue that they are likely to get 29 or 30 M&M's and at the very least 27 M&Ms. They might even get as many as 36 M&Ms sometimes.
Students that prefer Salt and Sweet Mix may argue that while there is more variability in this data set they would tolerate a chance of getting only 21 M&M's at the chance at getting 36 or 40 M&Ms.

2. In the dot plot given below, each value is shown as its absolute deviation from the *mean* of **249**.



a. Find the MAD of the data.

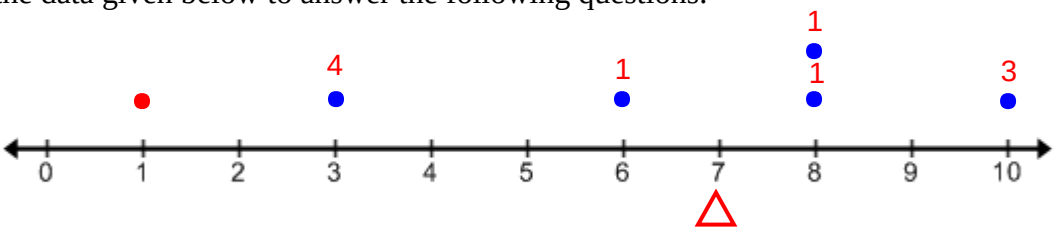
$$\frac{39+23+19+13+9+9+9+9+9+15+38+50}{12}=\frac{242}{12}\approx 20.2$$

b. Which of the following is the correct interpretation of the MAD for this context?

i. The number of visits to the website per week is approximately 20.2 visits away from the mean.
ii. The number of visits to the website per week is approximately 20.2.

iii. The number of visits to the website per week is, on average, approximately 20.2 visits from the *mean* of 249.

3. Use the data given below to answer the following questions.



a. Find the *mean* of the data and mark it with a balance triangle on the plot.
The *mean* is 7

b. Mark the absolute deviations from the *mean* for each data value on the dot plot and use them to find the MAD.

$$\frac{4+1+1+1+3}{5}=\frac{10}{5}=2$$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

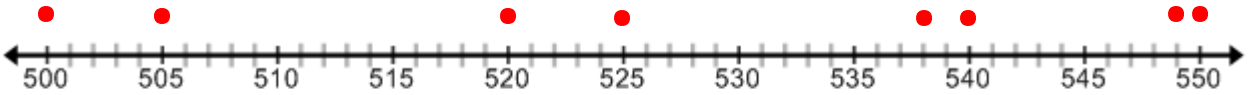
c. What value could you add to this data set that will not change the *mean*? Explain your answer.
A value of 7 will not change the mean because it has an absolute deviation of 0.

d. Suppose a value of 1 is added to the data set. Add this point to the dot plot. Without doing any calculations determine how this will affect the mean and MAD.
Adding a 1 to the data set will make the *mean* go down because in order to "balance" the data you would have to move your balancing point down. The MAD will go up because adding a value of one makes the data set have more variability.

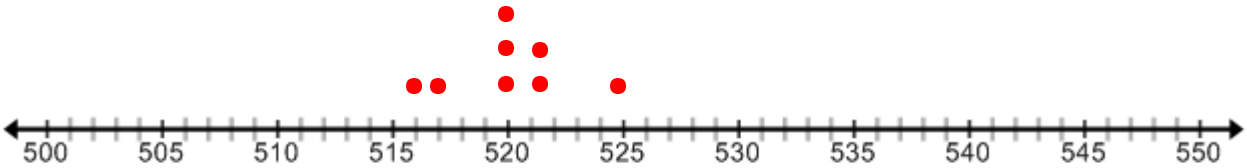
e. What value or values would you add to the data set in addition to the value of 1 to make the *mean* return to 7?
Since 1 has an absolute deviation to the left of the *mean* of 6, you would have to add values that have absolute deviations to the right of the mean that sum to 6. For example you could add two values of 10 or one value of 13.

4. For the number lines below, make a dot plot with at least 6 values that matches the following variability requirements. Sample answers are given, answers will vary.

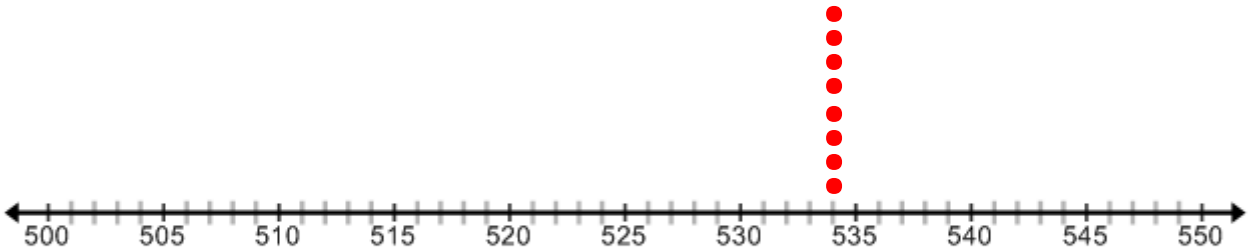
a. A plot with a lot of variability.



b. A plot with less variability than the plot in part a.

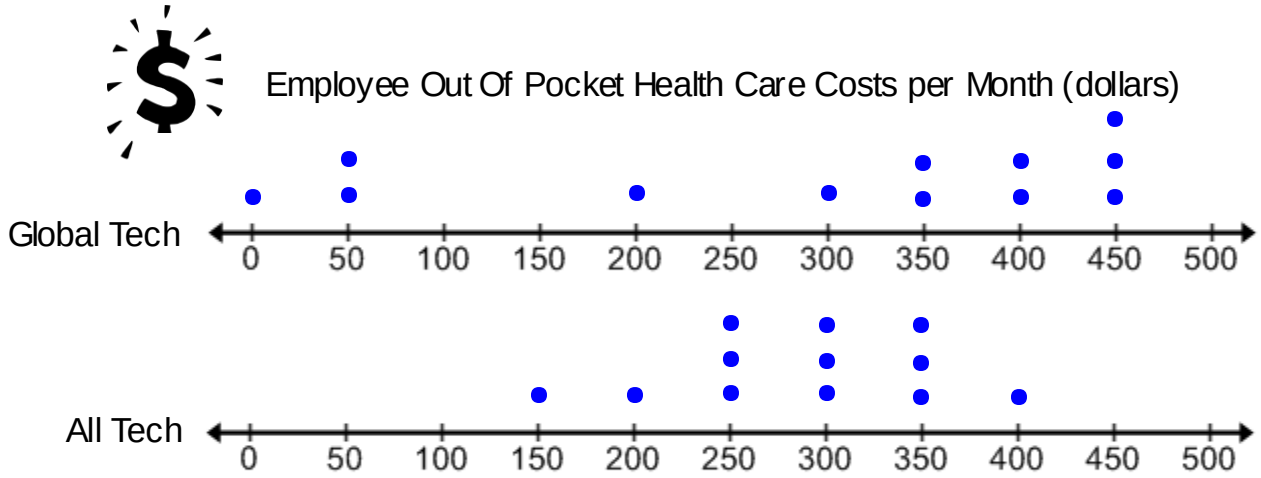


c. A plot with no variability.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. The dot plots below shows the out of pockets cost of health care per month for several employees at two different technology companies. The *mean* cost of health care for both companies is approximately $287.50.



a. Describe the shape of the data for each company.
The shape for Global Tech appears to be skewed left and the shape for All Tech is fairly symmetrical.

b. The MAD for each company is listed below. Interpret what the MAD is telling us for these data sets.

Global Tech MAD: ≈ 141.7

All Tech MAD:≈ 56.25

The MAD for Global Tech indicates that the out of pocket cost of health care for employees has an average deviation of $141.70 from the *mean* cost.
The MAD for All Tech indicates that the out of pocket cost of health care for employees has an average deviation of $56.25 from the mean cost.

Note that the context must be accounted for when analyzing the size of the MAD. In this case, the MAD for All Tech is $56.25 which might seem really big compared to some of the other MADs we have discussed. However, comparing it to the MAD for Global Tech and the context, it is relatively small.

c. For which company is the mean a better indicator of a typical value of the data distribution? Use the shape and MAD to justify your answer.
The *mean* is a better indicator of a typical value of the distribution for All Tech because it has a smaller MAD.

d. What might be a more appropriate measure of center for the data set with the larger MAD?
The data for Global Tech is skewed left and it has a rather high MAD, this indicates that the median would probably be a better measure of center for this data.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

6. Adila is going to purchase her first cell phone and is trying to decide between two different brands of phones. She is going to base her final decision on how long the phone can last without having to be recharged. She decides to survey her friends that own these two phones on the battery life of their phones. The results of her survey are shown in the table below.

Phone Brand A

Battery Life (hours)	Absolute Deviation
10	0.75
9	0.25
8	1.25
9	0.25
10	0.75
7	2.25
11	1.75
10	0.75
Total:	8

Phone Brand B

Battery Life (hours)	Absolute Deviation
7	2.25
14	4.75
13	3.75
8	1.25
6	3.25
8	1.25
9	0.25
9	0.25
Total:	17

a. Find the mean battery life for each phone brand.
Brand A *mean*: 9.25 hours

Brand B *mean*: 9.25 hours

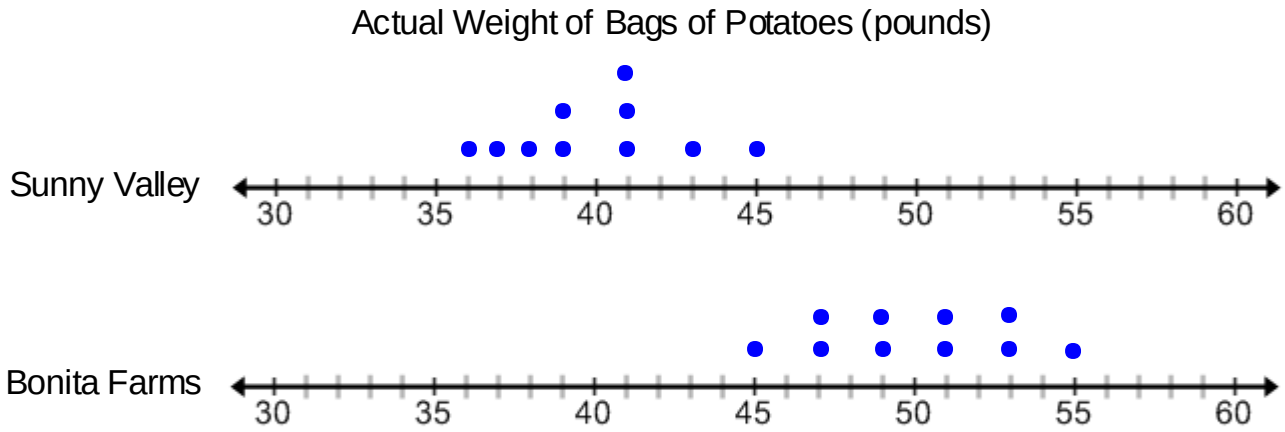
b. Fill in the missing information in the table to find the MAD for each set of data. Explain what the MAD means for each data set as well.
Brand A MAD: 1; The average deviation from the mean is 1 hour of battery life

Brand B MAD: 2.125; The average deviation from the mean is 2.125 hours of battery life.

c. Make an argument for choosing each brand of phone, be sure to discuss the data's variability in your argument.
Students that argue that Brand A is better may suggest that less variability in battery life is important. If you choose this brand, it likely that you will get a battery that lasts around the mean time of 9.25 hours.
Students that argue that Brand B is better might prefer to take a chance on getting a battery that lasts as long as 13 or 14 hours.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

7. The manager of a cafeteria is ordering potatoes. He can choose between two different brands of potatoes that are of the same quality and have the same cost per pound. Sunny Valley offers 40 lb bags and Bonita Farms offers 50 lb bags. Each manufacturer guarantees that their bags of potatoes will weigh close to the labeled weight. The manager at the cafeteria has weighed the last ten bags of potatoes that he received from each brand. His results are shown below. None of the bags are the exact weight.



The manager at the cafeteria is trying to determine which brand of potatoes offers bags that are closer to their labeled weight so he can decide which brand to buy in the future. Circle the statements below that are true.

- The MAD for Sunny Valley is 40 lbs.
- The MAD for Bonita Farms is 2.6 lbs.
- The MAD for Sunny Valley is 2.2 lbs.
- The MAD for Bonita Farms is 50 lbs.
- The manager cannot determine which bags are closer to their labeled weight because the average for each data set is different.
- Sunny Valley offers bags that are closer to their labeled weight because the MAD for its data is smaller than the MAD for Bonita Farms.
- The manger should go with the 50 lbs bags over the 40 lb bags regardless of their exact weights because he is always going to get more potatoes for the same price.
- The average deviation of weights from the *mean* weight of 40 pounds for Sunny Valley is smaller than the average deviation of weights from the mean weight of 50 lbs for Bonita Farms.

8. Two classes took the same test. A summary of the distribution of scores is provided below.

Class	Mean	MAD
1st period	76	4
2nd period	76	10

a. Suppose you had the highest score in the class on this test. Would your score most likely have been higher if you were in 1st period or 2nd period? Explain your answer.
Since the *means* are the same your score would have most likely been higher if you were in 2nd period because it has a higher MAD and therefore more variability.

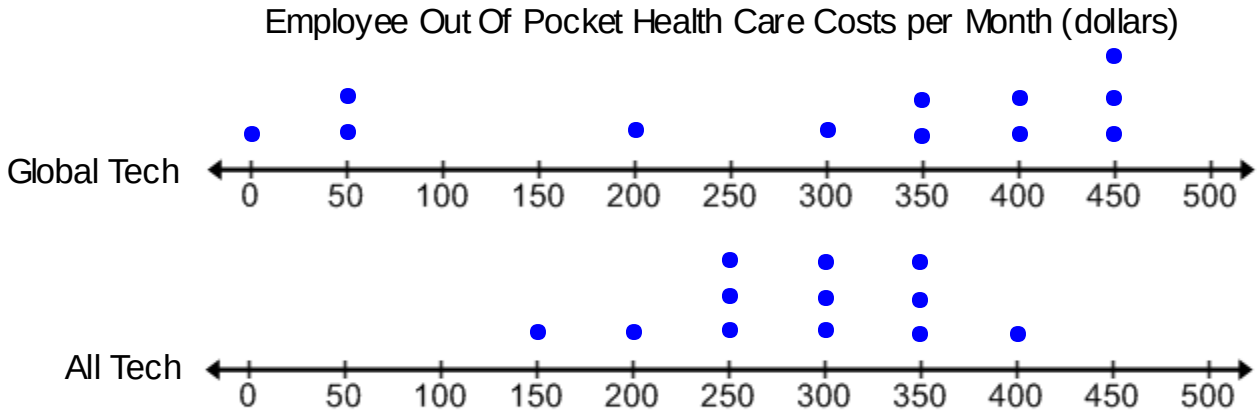
b. Suppose you had the lowest score in the class on this test. Would your score have most likely have been higher if you were in 1st period or 2nd period? Explain your answer.
Since the *means* are the same your score would have most likely been higher if you were in 1st period. Since this class has a lower MAD there is less variability so your score would be closer to the *mean*.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3b Class Activity: The Interquartile Range

In the previous homework assignment, you were given the out of pocket cost of health care for two different companies. This data is shown below.

Not all measures of variability are the same. You must account for which measure of center is best when using a measure of variability to analyze data. Students must adhere to precision as they choose which numerical measures to use when interpreting data.



1. Discuss which measure of center best describes each set of data? Justify your answer using the questions and answers from your homework assignment.
   Review with students that the mean was a good measure of a typical value for All Tech. This is because the MAD for this data is approximately $56.25 while the MAD for Global Tech is $141.60. Also the shape of All Tech appears to be fairly symmetrical; this also suggests the mean as a good measure of center. The shape of Global Techs data is skewed left; this might suggests that the median would be a better measure of center.

2. Find the median from the Global Tech data and compare it to the mean.
   The median for the Global Tech data is $350. This is approximately $62.50 more than the mean. Not only does the median describe the center of the data but is also divides the data in half. The same number of values will be above the median as below it.

3. Discuss how to analyze the variability of a data set when the median is the best measure of center?
   In order to analyze the variability of Global Tech's data we would not use the MAD because it is based off of the deviations from the mean. Rather the range would be a better indicator. Review how to find the range of a data set.

4. Find and interpret the range of the Global Tech data.
   $\mathbf{450 - 0 = 450}$. The data for Global Tech ranges from $450 a month to $0 a month.
   It might be important to note that the median does not split the range in half but rather the number of data values in half. You can point this out by discussing the median for this data is 350 while the range cut in half is 225.

This is a perfect example where you can discuss how someone may choose a certain measure of center to report because it plays to their advantage. While both companies have the same mean cost for health care, the medians have a $50 difference. Discuss with students why one company might choose one measure of center over another to report the typical cost for health care and how variability plays into this as well.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

The range for this data set gives us an idea of how spread out the data is from min to max. However, it does not really let us know how the numbers are organized within the range. **Quartiles** can help us to better interpret the variability within the data set. Just like the median, quartiles split the data into equal parts.

Possible discussion questions are given below.

- How many parts do you think the quartiles split the data into?
  Quartiles split the data into 4 parts. Talk about how the pre-fix *quart* means 4.

  - ***Quartiles* are the values of the points that split the data into four equal parts.**
  - **The *1st quartile* (or *lower quartile*) separates the lower half of the data into two equal parts. It is the median of the lower half of the data.**
  - **The *3rd quartile* (or *upper quartile*) separates the upper half of the data into two equal parts. It is the median of the upper half of the data.**

- What quartile is the median?
  The median is the 2nd quartile.

There are different methods used to find the quartiles by statisticians and statistical software. The Common Core Standards use the method which excludes the median to create two halves when the number of data points is odd. See footnote in 6 – 8 Statistics and Probability Progressions Document page 4.

5. Find and label the median, 1st quartile, 3rd quartile, max, and min for the Global Tech data below.

$$IQR = 425 - 125 = 300$$

0, 50, 50, 200, 300, 350, 350, 400, 400, 450, 450, 450

Min — Q1=125 — Median=350 — Q3=425 — Max

   a. How much of the data is below the median?
      Half of the data or 50%

   b. How much of the data is below the 1st quartile?
      1/4th of the data or 25%

   c. How much of the data is between the median and 3rd quartile?
      1/4th of the data or 25%

   d. How much of the data is between the 1st quartile and 3rd quartile?
      Half of the data or 50%

   e. How much of the data is below the 3rd quartile?
      3/4th of the data or 75%

The **interquartile range (IQR)** can be found by finding the difference between the 3rd quartile and 1st quartile.

8. Find the interquartile range (IQR) of the Global Tech data. Then mark it on the data above.
   The IQR is $\mathbf{425 - 125 = 300}$.

9. How much of the data falls within the interquartile range (IQR)?
   50 %, be sure to discuss that since quartiles split the data into 4 groups, and the IQR is the difference between Q3 and Q1, then for every data set, 50% of the data will fall within the IQR.
10. What does the IQR tell you about the data?
    In general, the IQR is an indicator of the variability of the middle 50% of the data. That is the data surrounding the median. The IQR of this data indicates that the spread of the data around the median is within $300 of each other or that 50% of the data values vary by no more than $300.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

The interquartile range or *IQR* is the difference of the 3rd quartile and 1st quartile.

$$IQR = Q3 - Q1$$

The *IQR* is a single number that describes that variability of a data set.

11. Find the interquartile range of the data given below. As you work through the data write down your steps for finding the IQR. Encourage students to work with a partner or group on this problem. If students need help, tell them to refer to the definitions on the previous page.

Top Speeds of the World's Fastest Motorcycles in Miles Per Hour

186, 174, 175, 176, 176, 176, 175, 227, 190, 230, 248, 275, 169, 325, 190

Step 1-Order the data from least to greatest;
169, 174, 175, 175, 176, 176, 176, 186, 190, 190, 227, 230, 248, 275, 325
Step 2-Find the median of the data set; median is 186
Step 3-To find Q1 find the median of the lower half of the data; Q1 is 175
Step 4-To find Q3 find the median of the upper half of the data; Q3 is 230
Step 5-To find the IQR find the difference between Q3 and Q1; IQR is 55

Review these steps with the class together. Consider asking the following questions in a class discussion. What is the range of the data? How much of the data is below the median? What percent of the data falls within the IQR? How much of the data is below 230 mph?

12. What does the IQR tell us about this data?
The middle 50% of the tops speeds are within 55 mph of each other.

13. The data sets given below show the number of words typed per minute for Boston and Francis.

Boston: 70, 71, 50, 73, 73, 72, 55, 72, 64, 72

Francis: 71, 70, 65, 66, 68, 70, 62, 66, 68, 66

a. Find the range for the words typed per minute for each person.
Boston's range = 23
Francis's range = 8
b. Find the median words typed per minute for each person. Show your work above.
Boston's median = 71.5 words per minute
Francis's median = 67 words per minute
c. Find Q1 and Q3 for each set of data.
Boston $\boldsymbol{Q1 = 64, Q3 = 72}$
Francis $\boldsymbol{Q1 = 66, Q3 = 70}$
d. Find the IQR for each set of data.
Boston $\boldsymbol{IQR = 72 - 64 = 8}$
Francis $\boldsymbol{IQR = 70 - 66 = 4}$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

e. Use the IQR to compare the variability of each data set.
The IQR is smaller for Francis. This suggests that the middle half of the counts of data is closer to the median than the middle half of the counts of data for Boston. Francis has less variability in the middle 50% of his scores than Boston does. This means that Francis' typing scores were generally closer to his median score than Boston's scores were to her median score.

f. Use the information above to make an argument for who will type more words per minute in the future.
Boston's median number of words typed per minute is higher than Francis so you could argue that she will be able to type more words in the future. However, you could also argue that Francis data has less variability so the likelihood of him typing close to his median number of words typed per minute is greater and since there is not a huge difference between the two medians Francis may type more words per minute than Boston.

14. The tables below represent the ages of people at two different city skate parks on a given afternoon.

Ages of People at City Skate Parks	
South Jordan City	Salt Lake City
4	3
10	8
10	10
11	11
14	11
15	19
15	22
15	23
16	28
20	29
32	32
40	39

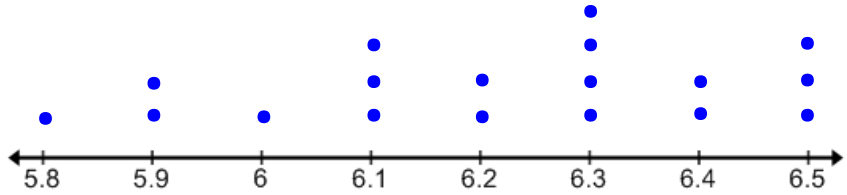
a. Make a prediction about how the sizes of IQR for the ages of people at each skate park compare.
The IQR for South Jordan City's Skate Park is probably smaller because the ages of people in the middle of the data set (when the values are placed in order) appear to vary less than the ages of people in the middle of the data set at Salt Lake City's Skate Park.

Some students may argue that the ranges are the same so the IQR should be the same for each data set. Talk about how even if you have two data sets with the same range we still do not know how the data is grouped or arranged within the range of values.

b. Find and interpret the IQR for each city and compare your answers to your prediction in part a.
Encourage students that is acceptable to mark the median and quartiles on the table itself since the numbers are already in order rather than re-writing the values horizontally.
South Jordan City $IQR = 7.5$; The middle 50% of the people at this park have ages that vary by 7.5 years.
Salt Lake City $IQR = 18$; The middle 50% of the people at this park have ages that vary by 18 years.
The IQR for Salt Lake City is more than double the IQR for South Jordan City. This means that the ages of the people at the Salt Lake City skate park have more variation than the ages of the people at the South Jordan City Skate Park.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

15. Find the range and interquartile range of the data given in the dot plot below.



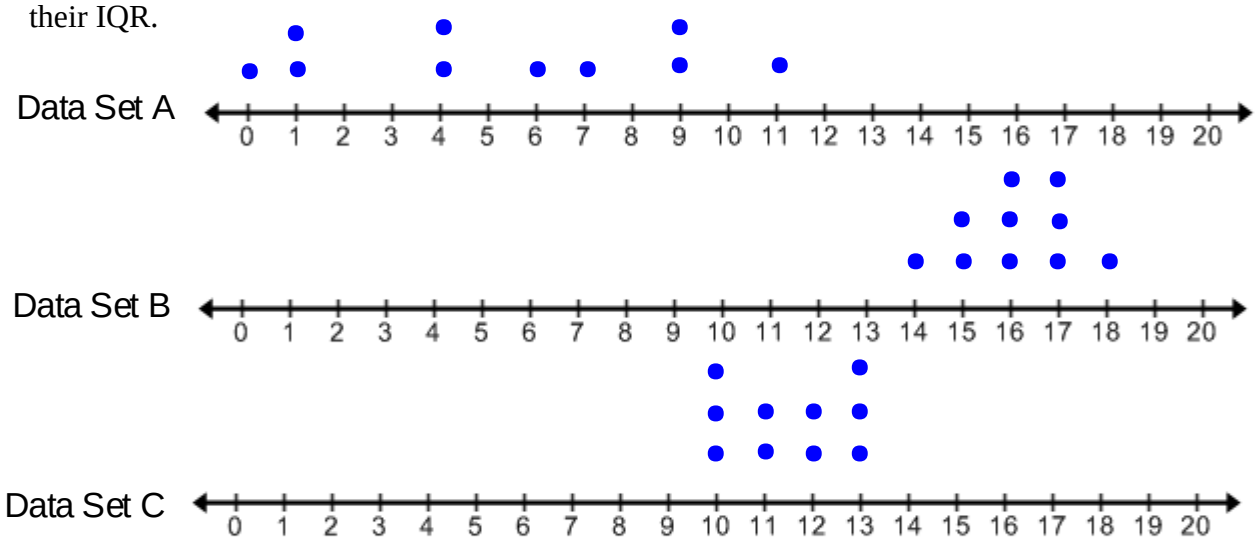
$Range = 0.7$

$Q1 = 6.1, Q3 = 6.4, IQR = 0.3$

Students may struggle finding the median and quartiles from a dot plot. Encourage them to write the numbers in a list from least to greatest if needed. Another option is to count the number of values in the data set, for this problem there are 18. This means that the median is between the 9th and 10th point on the graph if you start with the minimum point. In addition, since there are also 9 values to the right of the median and to the left of the median the 1st quartile will be the middle value of the lower 9 points, which will be the 5th value (starting with the minimum). Likewise the 3rd quartile will be the 5th value from the median. This is a good strategy to use for finding the quartiles and median if there are a lot of values in a data set.

Discuss with students that just like the MAD, the numerical value of the IQR must be considered along with the context in which it is given. For example, the IQR of 0.3 may seem very small, but if this data were representing some sort of dosage of medicine or precision instrument of some kind, then it might be a significant measurement.

16. Without doing any calculations order the data sets below from least to greatest according to the value of their IQR.



Data Set B has the smallest IQR, then Data Set C, and then Data Set A.

Be sure to discuss possible misconceptions such as confusing data sets that have large quartile values as having a larger IQR. Also just because a data set has a larger range does not mean that its IQR is also larger, as is the case with Data set B and C. The range for Data Set B is larger than the range of Data Set C but the IQR for Data Set B is the smaller than the IQR for Data Set C.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

17. Create a data set with at least 8 data values that match each set of given conditions.
Sample answers are given.

a. A data set with an IQR that is very close to the range.
1, 1, 2, 5, 5, 9, 10, 10
$Range = 10 - 1 = 9$
$IQR = 9.5 - 1.5 = 8$

b. A data set with a small IQR and a large range.
1, 50, 50, 50, 50, 50, 50, 100
$Range = 100 - 1 = 99$
$IQR = 50 - 50 = 0$

c. A data set where the median and the 3rd quartile are the same number.
1, 10, 15, 50, 50, 50, 50, 100
$Median = 50$
$Q3 = 50$

Spiral Review

Fill in each blank with $<, >, or =$.

1. 14.87_ > __ 14.85
2. 329.4 _ > _3.294
3. 0.68 __ = __ $\frac{68}{100}$
4. 0.4003_ < __4.03

Round each number to the nearest tenth.

5. 25.43
25.4
6. 5.48
5.5
7. 2.72
2.7
8. 0.446
0.4

Fill in each blank with the correct measurement conversion.

9. $5\ m =$ __500_cm
10. $30\ m =$ _0.03_km
11. $2000\ mm =$ _200_cm
12. $40\ km =$ _40,000_m

Find each product.

13. $\frac{1}{2} \times \frac{1}{3}$
$\frac{1}{6}$
14. $\frac{1}{2} \times \frac{1}{4}$
$\frac{1}{8}$
15. $\frac{3}{4} \times \frac{1}{3}$
$\frac{1}{4}$
16. $\frac{2}{5} \times \frac{2}{3}$
$\frac{4}{15}$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3b Homework: The Interquartile Range

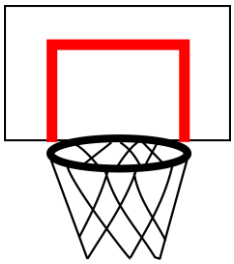
1. Find the range for each situation given.
   a. The youngest person in a family is 3 years old. The oldest person in the same family is 38 years old.
      $Range = 35\ years$
   b. The longest snake in a reptile exhibit is 10.5 ft long. The shortest snake in the exhibit is 0.5 ft long.
      $Range\ 10\ ft$

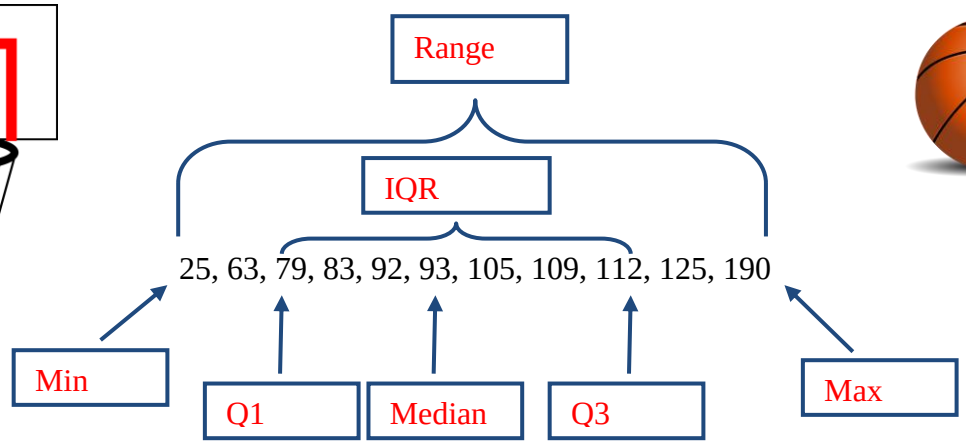
2. The data below represents the number of points scored by a team in their first season of basketball games.

   a. Place the words below in the correct blank in the data set.

   Max Min Median Q1 Q3 Range IQR

Number of Points Scored in Basketball Games





   a. Calculate the Range for the data above.
      $Range = 165$

   b. In how many games did the team score more than 93 points?
      The team scored more than 93 points in half of the games because 93 is the median.

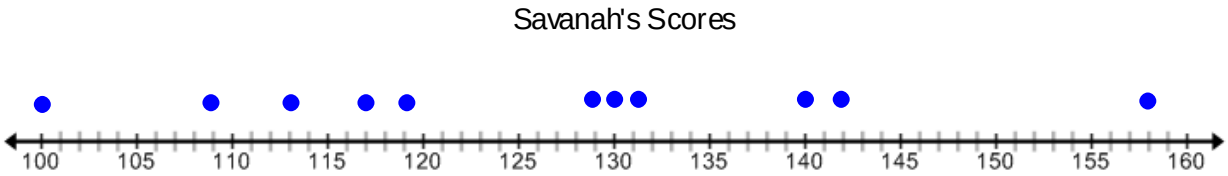
   c. In 75% of the games, the team scored less than how many points?
      112 points

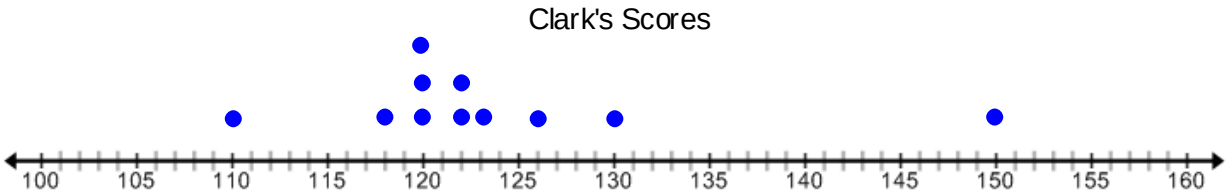
   d. Calculate and interpret the IQR for the data above.
      $IQR = 33$; This means that the points scored were within 33 points of each other for the middle 50% of the games.

   e. Which measure of variability, the range or the IQR, better reflects this distribution of data? Explain.
      The IQR is a better measure of variability for this data because the range of the points is 165; the scores of 25 and 190 are outliers and make the range big.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. Clark and Savannah love to go bowling. They have recorded their bowling scores for the last 11 games they have played. These scores are shown in the dot plots below.





a. Find the range for each persons set of scores.
   Savannah's range = 58
   Clark's range = 40
b. Find the median score for each person. Show your work above.
   Savannah's median= 129
   Clark's median = 122
c. Find Q1 and Q3 for each set of data.
   Savannah's $Q1 = 113, Q3 = 140$
   Clark's $Q1 = 120, Q3 = 126$
d. Find the IQR for each set of data.
   Savannah's $IQR = 140 - 113 = 27$
   Clark's $IQR = 126 - 120 = 6$
e. Use the IQR to compare the variability of each data set.
   The IQR is smaller for Clark, this suggests that the middle half of the counts of data is closer to the median than the middle half of the counts of data for Savannah. Clark has less variability in his scores than Savannah does. This means that Clark's scores for the middle 50% of his data were closer to his median score than Savannah's scores for the middle 50% of her data were to her median score.
f. Use the information above to make an argument for who will get a higher score in the future.
   Savannah's median score is higher than Clark's so you could argue that she will have a better score in the future. However, you could also argue that Clark's data has less variability so the likelihood of him scoring close to his median score is greater than the likelihood of Savanah scoring close to her median score.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. The lunch menu at Anthony's school advertises that each lunch comes with a serving of 12 tater tots. Anthony is convinced that the lunch server with the red hair at his school cafeteria hands out about 12 tater tots each day. His friend thinks the lunch server with the glasses hands out about 12 tater tots each day. They decide to track the number of tater tots they get on their lunch trays for two weeks. Each day Anthony goes to the lunch server with the red hair and his friend goes to the lunch server with the glasses. At the end of the two weeks they find the median number of tater tots given out by each server. Their findings are in the table below.

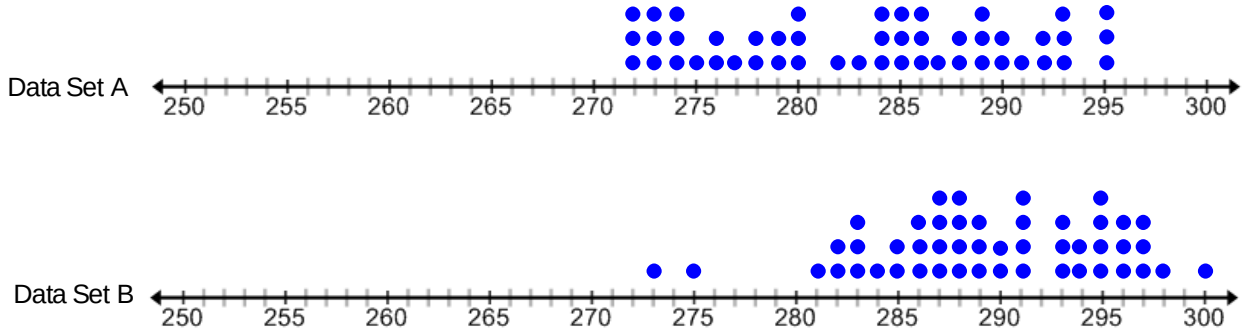
Number of Tater Tots										
Lunch Server	Mon	Tue	Wed	Thur	Fri	Mon	Tue	Wed	Thur	Fri
Red Hair	13	9	12	11	14	15	11	12	10	14
Glasses	15	13	11	10	13	5	20	9	17	10

a. Find the median for each lunch server.
$Red\ Hair\ median\ = 12$
$Glasses\ median = 12$

b. Find the IQR for each lunch server.
$Red\ Hair\ IQR =\ 3$
$Glasses\ IQR = 5$

c. Use median and IQR for each lunch server to determine who is most likely to give out the advertised 12 tater tots.
For the lunch server with the red hair the middle 50% of the data varies around its median by 3 tater tots. For the lunch server with the glasses the middle 50 % of the data varies around its median by 5 tater tots. Thus the lunch server with the red hair has a higher likelihood of handing out 12 tater tots.

5. Jonas was asked to decide which of the two data sets given below has a greater IQR without making any calculations. His answer and reasoning are given below. He has made a mistake in his thinking. Explain the error in his thinking and correct his answer.



"Data Set B has the larger IQR because its 1st and 3rd quartiles and median are bigger than the quartiles and medians for Data Set A. Also Data Set B has a larger range."

Jonas has made the mistake of confusing higher values for the quartiles with a higher IQR. IQR is measuring the difference between the first and third quartiles. Also he assumed that the data set with the larger range would have a higher IQR. Upon inspection, the Data Set A has a higher IQR because its data is more spread out around its median than Data Set B.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

6. Create a data set with at least 8 data values that match each set of given conditions.
   Sample answers are given.
   a. A data set where the IQR is equal to the range.
      10,10,10,10,10,10,10,10
      $Range = 0$
      $IQR = 0$
   b. Two different data sets that have the same range but a different IQR.
      $1, 2, 3, 4, 5, 6, 7, 8;\ IQR = 4$
      $1, 1, 1, 1, 1, 1, 1, 8;\ IQR = 0$
      $Range\ = 7$
   c. A data set with a range of 10 and an IQR of 3.
      $20, 24, 24, 25, 26, 27, 30$
      $Range = 10$
      $IQR = 3$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3c Class Activity: Box Plots

In this section, students are introduced to another way to model a data distribution. In addition to knowing how to make and interpret a box plot, students must recognize when it is appropriate to use this tool for data analysis.

In the previous section we discussed how to analyze the variability of a data set where the best measure of center is the median. To do so we looked at the *maximum value, minimum value, median value, 1st quartile, and 3rd quartile*. These values also help us to better understand the distribution of a data set. They are often called **the 5-number summary**. Review what each value in the 5-number summary describes. Emphasize the vocabulary below.

#1 Maximum Value The smallest value in a data set.

#2 Minimum Value The largest value in a data set.

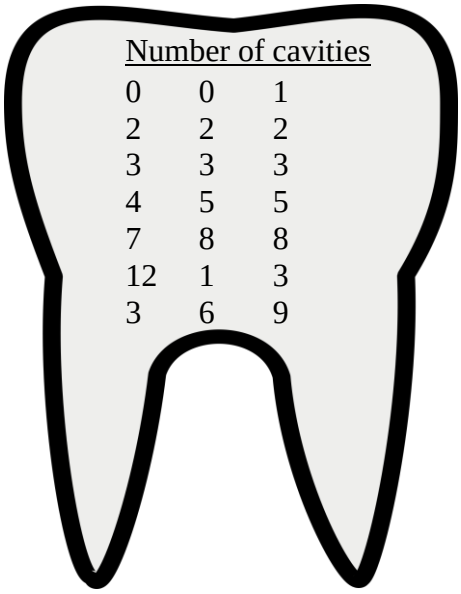
#3 Median The middle value of a data set, it splits the data in half, and is a measure of center.

#4 First Quartile (lower quartile) The median of the upper half of the data. 75% or 3/4 of the data values are below it, 25% or 1/4th of the data value are above it.

#5 Third Quartile (upper quartile) The median of the lower half of the data. 25% or ¼ of the data values are below it, 75% or 3/4th of the data values are above it.

These values can best be reviewed through a class discussion. It may not be necessary for students to write the definitions above.

1. The number of cavities that a class of 6th graders have had are shown.



Number of cavities

0	0	1
2	2	2
3	3	3
4	5	5
7	8	8
12	1	3
3	6	9

a. Find the 5-number summary for this data set.

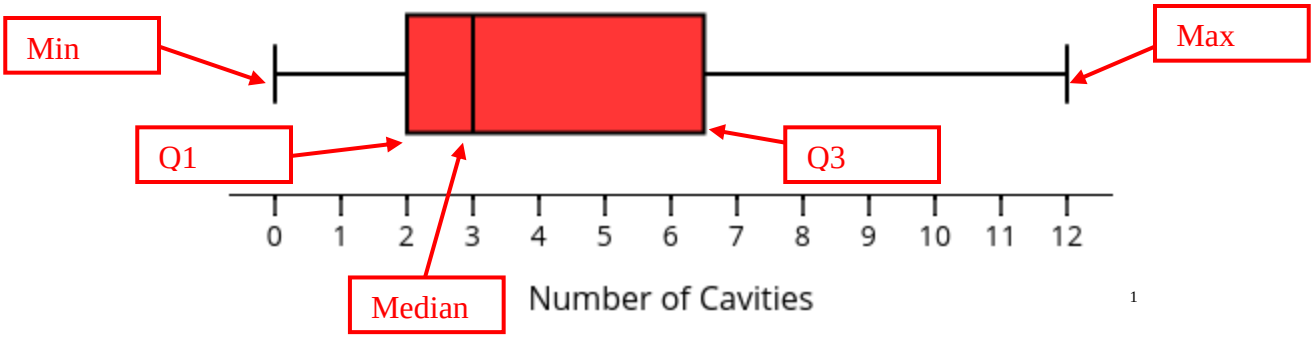
$Max = 12$
$Min = 0$
$Median = 3$
$Q1 = 2$
$Q3 = 6.5$

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

A **box plot** is a data display that divides a set of data into four parts using the median and quartiles. It can be used to better help us to understand a data distribution. It is a picture that shows how the 5-number summary values are related to each other on a number line.

b. Make a Box Plot of the cavity data on the previous page using the values in its 5-number summary.

Show students how to make a box plot of the data. Be sure to label the 5-number summary values on the plot. Have students list the steps for making a box plot so that they can refer to it later.



[1]

1. Graph the median, min, max, lower, and upper quartiles with points above the number line.
2. Draw a box around the two middle quartiles.
3. Draw a line segment through the median.
4. Draw a line segment (whiskers) that extend from the quartiles to the max and min values.

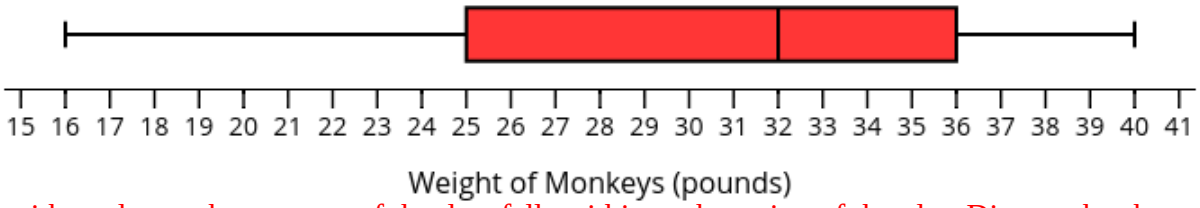
Be sure to label the plot with a title. Discuss what the plot tells us about how the data is distributed. What was the median number of cavities in this class? (3) What was the least amount of cavities? (0) What was the greatest amount of cavities? (12) What is the range of this data? (12) How many of the students had between 2 and 6.5 cavities? (50%) What fraction or percent of the class are in each of these sections? (Refer to quartile sections) (1/4 or 25%) How can you interpret the shape of the data from a box plot? (Look at where the box lies within the range of data, and where is the median line located inside the box. In this case the data is slightly skewed right, as shown by the long segment extending to the max and the median is to the left of the center of the box.) In what sections are the data values close together, in what section are they spread out? (Look at the length of the line segments and the area of the boxes within the IQR box.) Are there any outliers?

2. The data below shows the weight of monkeys at a zoo.

Weight of Monkeys in pounds: 17, 22, 16, 40, 32, 38, 32, 32, 29, 28, 35, 37

a. Work with a partner to make a box plot of the data. (Hint: begin by finding The 5-Number Summary values of the data.)

$Max = 40, Min = 16, Median = 32, Q1 = 25, Q3 = 36$



[2]

Review with students what percent of the data falls within each section of the plot. Discuss the shape of the data and include any outliers.

b. Use the box plot to find the IQR of the data. Use the median and IQR to interpret the variability of the data.

$Q3 - Q1 = 36 - 25 = 11$; The median is 32 pounds, meaning the middle 50% of the monkey weights that surround the median vary by 11 lbs.

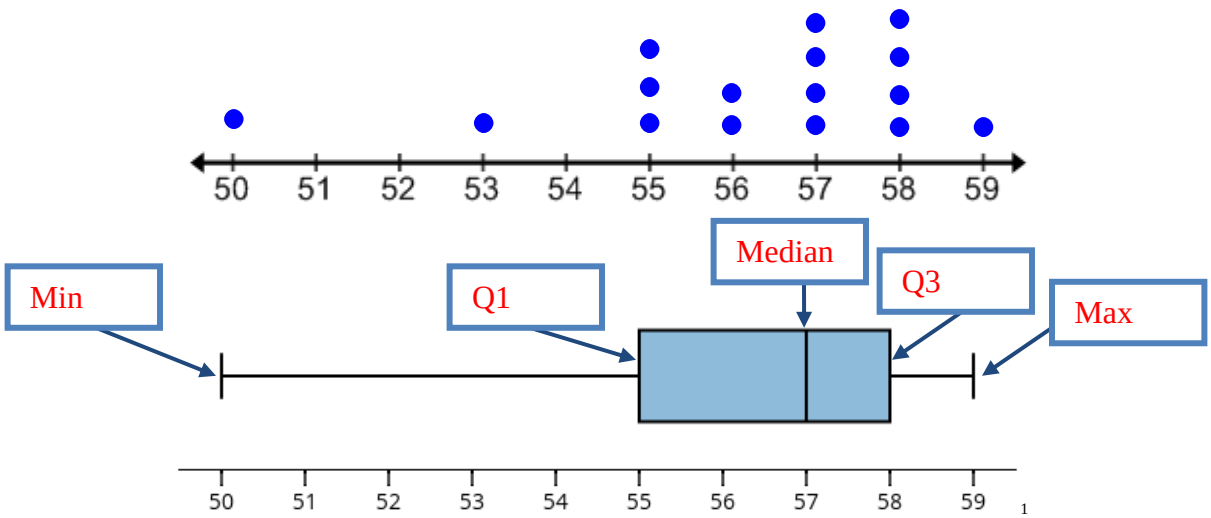
c. What might explain the variability in weight for the monkeys at the zoo?

Diet, health of the monkey, age of the monkey, living conditions, type of monkey, etc

1 evatipp. "Cavities" Plotly Technologies Inc., https://plot.ly/426/~evatipp/ (Last Accessed: April 2016)
2 evatipp. "Weight of Monkeys" Plotly Technologies Inc., https://plot.ly/398/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. The dot plot and the box plot below represent the same set of data.
   a. Fill in the missing labels with the 5-number summary values on the box plot.



[1]

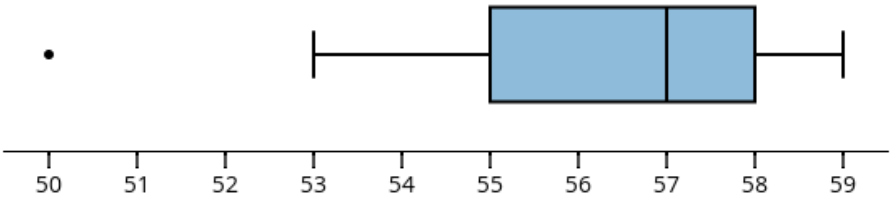
   b. In which display is it easier to identify the median and the quartiles of the data?
   The box plot because the vertical line segments identify the median and quartiles.

   c. In which display is it easier to identify exactly how many data points have a value of 58.
   The dot plot. Similar to a histogram, in a box plot you cannot determine the number of data points that have a specific value.

Help students to see the connection between the small size of the box between the median and the 3rd quartile and its correlation to the 25% of the data values being between 57 and 58. Recall that each quartile contains 25% of the data, in this case 4 data values (1/4 of 16 = 4). You can think of these 4 values being "crammed" into this box indicating less variability. You can see that the shape of the data is skewed left in the box plot by examining the long "whisker" that extends to the minimum value and the location of the median line in the box.

   d. Use the plots to describe the distribution of this data, be sure to discuss the *center*, *shape* and *spread/variability* of the data.
   The median of the data is 57; half of the data is above 57. The data appears to be skewed left. The IQR for the data is 3, this indicates that the middle 50% of the data surrounding the median fall within 3 units of each other.

   e. The plot below is another box plot of this data. Compare this plot to the one above, how is it the same? How is it different? Which plot is better?



[2]

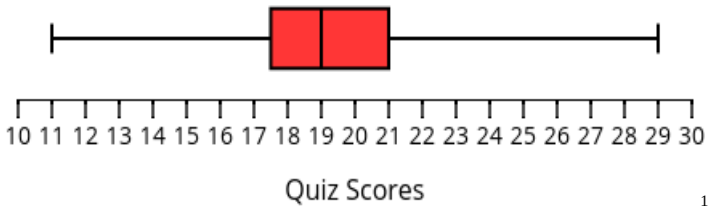
   This plot distinctively shows the outlier at 50. It shows the minimum (non-outlier) value as 53 and not 50. This plot is better because that you can see outliers in the data. Talk about how a long line segment extending to a max or min value does not necessarily indicate that these values are outliers; however, it is common for both types of plots to be used.

1 evatipp. "Compare Box and Dot" Plotly Technologies Inc., https://plot.ly/407/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Compare Box to Dot (1)" Plotly Technologies Inc., https://plot.ly/403/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

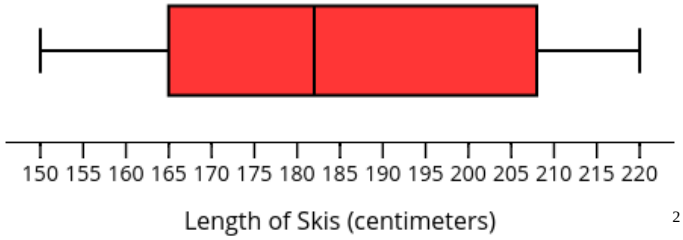
4. Make a box and whisker plot of the following data sets.

   a. Quiz Scores: 18,22, 29, 20, 22, 18, 15, 19, 17, 20 ,18, 19, 11



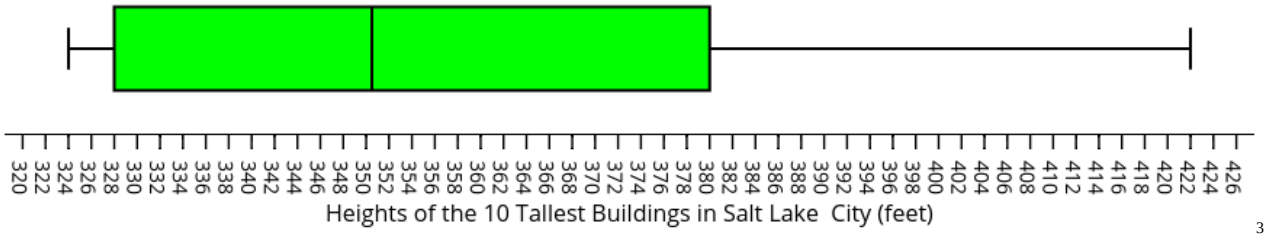
[1]

   b. Ski Lengths (centimeters): 182, 174, 208, 150, 220, 165, 200, 215, 170, 165, 205

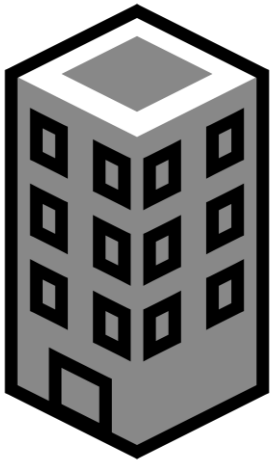


[2]

5. The box plot shows the height of the tallest buildings in Salt Lake City. Identify if the following statements are true or false. Justify your answer.



[3]



   i. The tallest building in Salt Lake City is 422 feet.
   True, the tallest building is the value at the end of the line segment to the right.

   ii. At least half of the buildings are less than about 350.5 feet tall.
   True, the median is about 350.5 so half of the buildings are less than this height.

   iii. Half of the buildings are taller than 380 feet tall
   False, only 25% of the buildings are taller than 380 feet.

   iv. Most of the buildings are greater than 328 feet
   True, 75% of the data is greater than 328 feet.

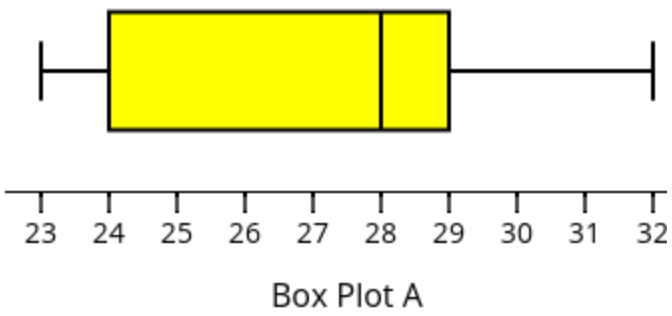
   v. Half of the buildings heights are within 52 feet.
   True, the IQR is 52, and 50% of the data is within the IQR.

   vi. We know that the value of 422 feet is an outlier because the line segment connecting it is so long.
   False, a long line segment extending to a max value does not always indicate that the max is an outlier.

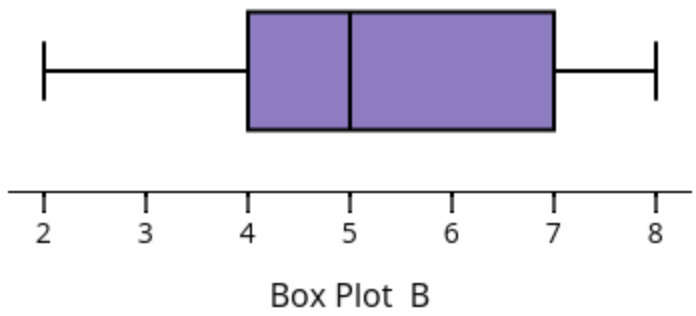
1 evatipp. "Quiz Scores (2)" Plotly Technologies Inc., https://plot.ly/429/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Length of Skis" Plotly Technologies Inc., https://plot.ly/432/~evatipp/ (Last Accessed: March 2016)
3 evatipp. "Tallest Buildings in SLC" Plotly Technologies Inc., https://plot.ly/411/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

6. Find the IQR of the two plots given; use the IQRs to determine which data set has less variability around the median.



[1]



[2]

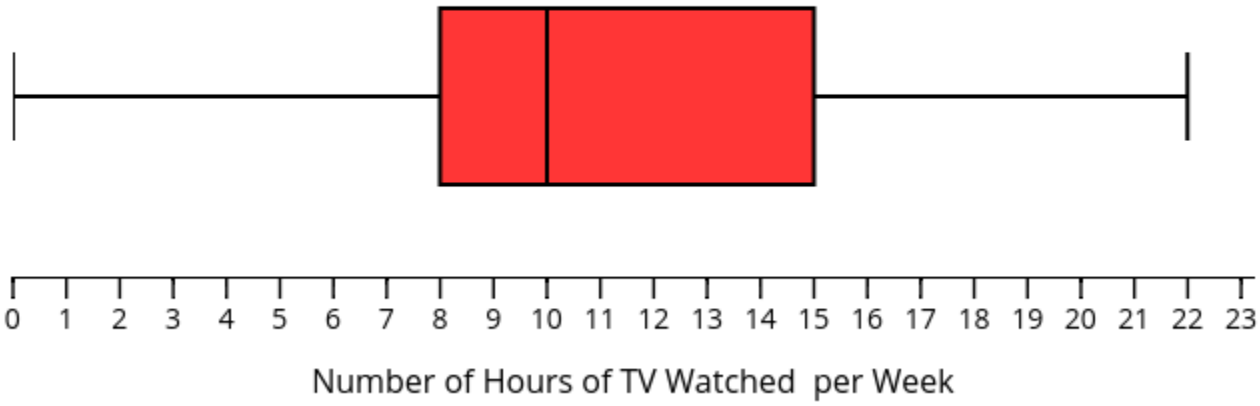
$Box\ Plot\ A\ IQR = 29 - 24 = 5$
$Box\ Blot\ B\ QR = 7 - 4 = 3$

The IQR for Plot B is smaller than Plot A, thus Plot B has less variability around its median.

Talk to students about being sure to pay attention to the scale when they compare box plots. While a box plot gives us a "snap shot" of the data, we must be careful when comparing them if they do not have the same scale.

7. Create a box plot for the information given below. Then find the IQR and describe the distribution of the data.

A group of 6th graders are interested in the number of hours of TV they watch each week. The most amount of TV watched per week in the group is 22 hours. The least amount of TV watched per week is 0. Half of the group watched less than 10 hours of TV per week. 25% of the group watched more than 15 hours of TV per week and the 1st quartile of the data is 8 hours of TV watched per week.



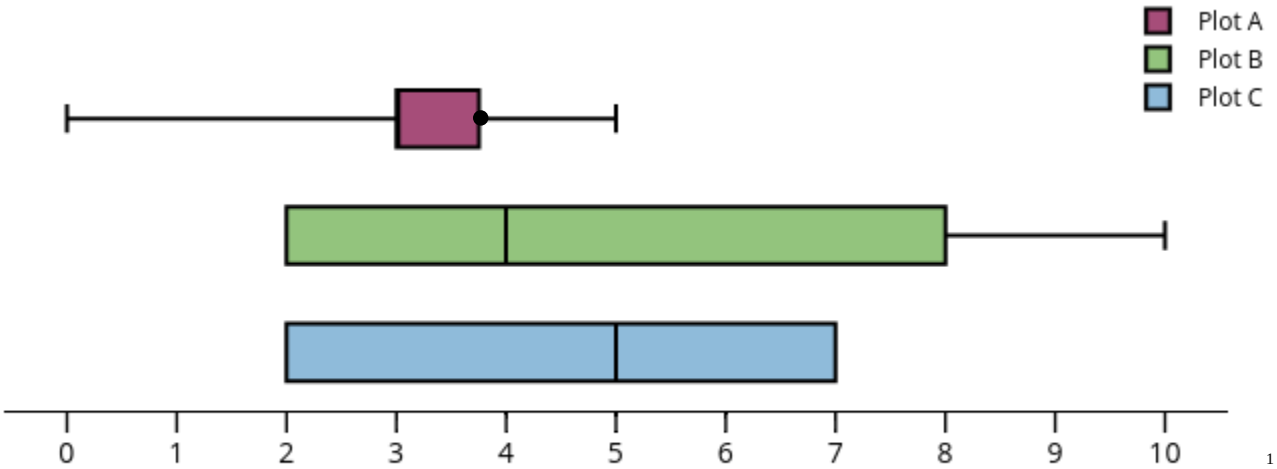
$IQR = 15 - 8 = 7$; The median of the data is 10 hours of TV watched per week. The IQR is 7, which means that the 50% of the data that surrounds the median varies by 7 hours. The range of the data is 22 hours. 1/4th of the students watch less than 8 hours of TV per week while 1/4th of them watch more than 15 hours of TV per week.

Ask students to consider the variability within the context. Do they think that a median of 10 and an IQR of 7 means that there is a lot of variability in the amount of TV these 6th graders watch?

1 evatipp. "Box Plot A" Plotly Technologies Inc., https://plot.ly/435/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Box Plot B (1)" Plotly Technologies Inc., https://plot.ly/438/~evatipp/ (Last Accessed: March 2016)
3 evatipp. "Hours of TV per Week (1)" Plotly Technologies Inc., https://plot.ly/441/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

8. Discuss the three box plots below. What makes them different from the plots previously discussed?



[1]

Plot A does not appear to have a median value. In this plot the median is the same value as the 3rd quartile. This is indicated with the point on the plot.
Plot B does not have a line segment extending to a minimum value. In this case, the 1st quartile and the minimum have the same value of 2.
Plot C does not have line segments extending to the max and minimum values. In this case, the 1st quartile and the minimum have the same value of 2 and the 3rd quartile and the max have the same value of 7.

Spiral Review

1. How does the value of the digit of 2 in 2,356 compare to the value of the digit of 2 in 3,256?

   a. It is 100 times the value.
   b. It is $\frac{1}{100}$ of the value.
   c. It is 10 times the value.
   d. It is $\frac{1}{10}$ of the value.

2. How does the value of the digit of 5 in 34.58 compare to the value of the digit of 5 in 85.43?

   a. It is 100 times the value.
   b. It is $\frac{1}{100}$ of the value.
   c. It is 10 times the value.
   d. It is $\frac{1}{10}$ of the value.

3. What decimal number is represented by $8 \times \frac{1}{10} + 3 \times \frac{1}{100} + 2 \times \frac{1}{1000}$?
   0.832

4. Write 34.06 in expanded notation.
   $3 \times 10 + 4 + 6 \times \frac{1}{100}$

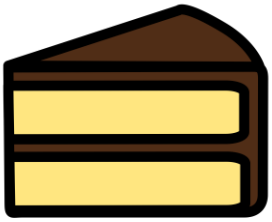
1 evatipp. "Test Scores Box Plot (2)" Plotly Technologies Inc., https://plot.ly/447/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3c Homework: Box Plots

1. The amount of calories per slice in several different types of cake is shown below.

Calories in a Slice of Cake			
Type of Cake	# of Calories	Type of Cake	# of Calories
Carrot	500	Angel Food Cake	70
Chocolate Fudge	450	Lite Coconut Crème	95
Red Velvet	300	White Chocolate Raspberry	340
Lemon Crème	325	Dulce De Leche	350
Devil's Food Cake	350	Hazelnut N Crème	400
German Chocolate	470	Cherry Chocolate Chip	425



a. Find the 5-Number Summary values for this set of data.
$Min = 70$
$Max = 500$
$Median = 350$
$Q1 = 312.5$
$Q3 = 437.5$

b. Make a box plot of the data.

[1]

It is also acceptable for students to create the box plot with the line segment on the left extending all the way to the minimum value. The plot above is indicating that 70 and 95 are outliers.

c. What kind of cake has the most calories per slice?
Carrot cake has the most calories at 500 calories per slice.

d. What percentage of the cakes have at least 350 calories?
50%

e. Why is the median line not in the center of the box?
We know that 25% of the data falls within each quartile. The median line that it not directly centered suggests a small range of numbers for this quartile. 25% of the data values are between 312.5 calories and 350 calories compared to the 25% of the data values that are between 350 and 437.5 calories.

1 evatipp. "Calories in Cake" Plotly Technologies Inc., https://plot.ly/525/~evatipp/ (Last Accessed: March 2016)

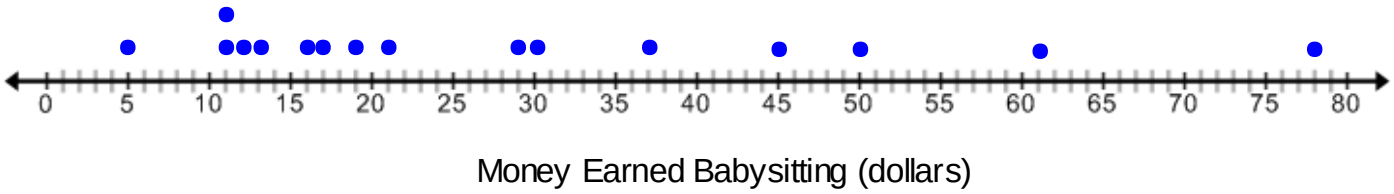
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

f. If you were on a diet what kind of cake would you choose?
Either angel food cake or lite coconut crème.

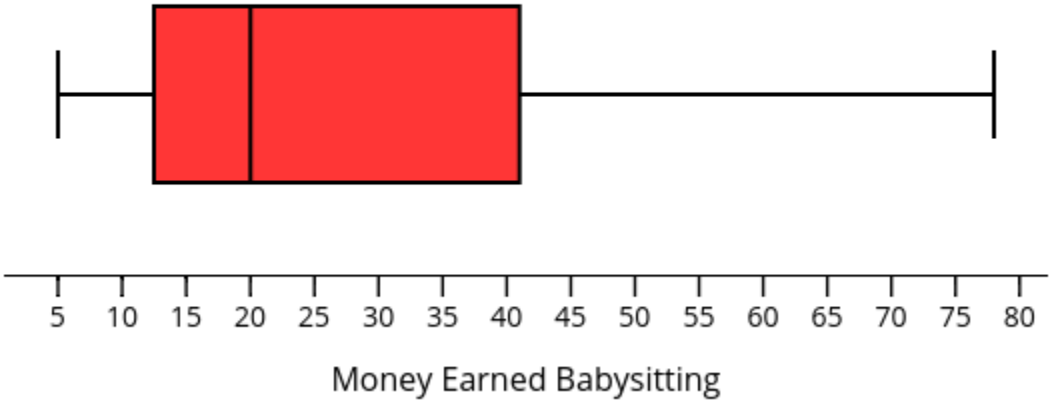
g. Find and interpret the IQR from the box plot.
$IQR = 437.5 - 312.5 = 125$; The 50% of the data that is clustered around the median varies by 125 calories.

h. Write a few sentences that summarize the distribution of this data. Be sure to discuss the shape, center, and spread.
The greatest number of calories per slice is 500 and the smallest number is 70 calories. The median number of calories per slice is 350. The data is slightly skewed left with 50% of cakes ranging between 312 and 437 calories. Only 25% of the cakes have less than 312.5 calories.

2. The dot plot shows the amount of money that Tommy has earned babysitting this year.



a. Make a box plot of the data.



[1]

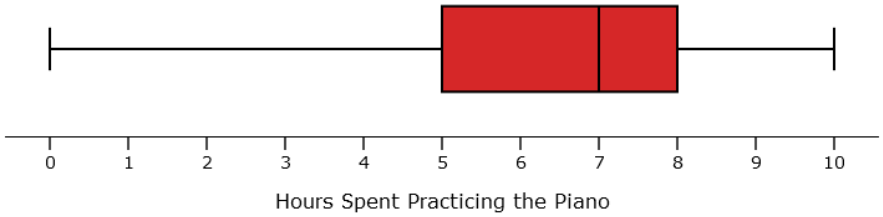
b. List one way a dot plot is a better display of data than a box plot. Then list one way a box plot is a better display of data than a dot plot.
In a dot plot you can see every exact value. You cannot see this in a box plot
In a box plot you can immediately see the value in the 5-number summary. You can also calculate the IQR fairly easily from a box plot.

c. Write a few sentences describing the distribution of this data. Be sure to discuss the shape, center, and spread.
The median amount of money that Tommy earned babysitting is $20. The range of the data is $73 with an outlier at $78. The data appears to be slightly skewed right with 50% of the data between between $12.50 and $41.00.

1 evatipp. "Babysitting" Plotly Technologies Inc., https://plot.ly/459/~evatipp/ (Last Accessed: March 2016)

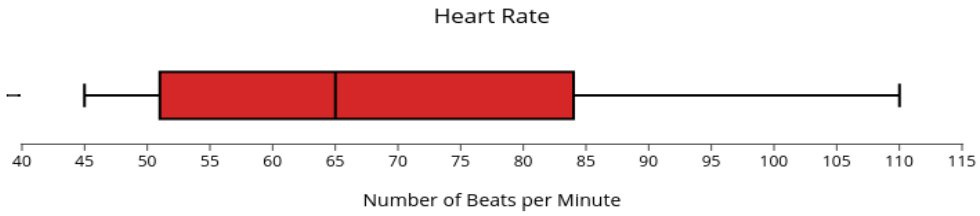
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. Make a box plot for each set of data below.
   a. Hours spent practicing the piano: 5, 8, 5, 10, 5, 8, 7, 9, 10, 3, 0, 8, 7, 6



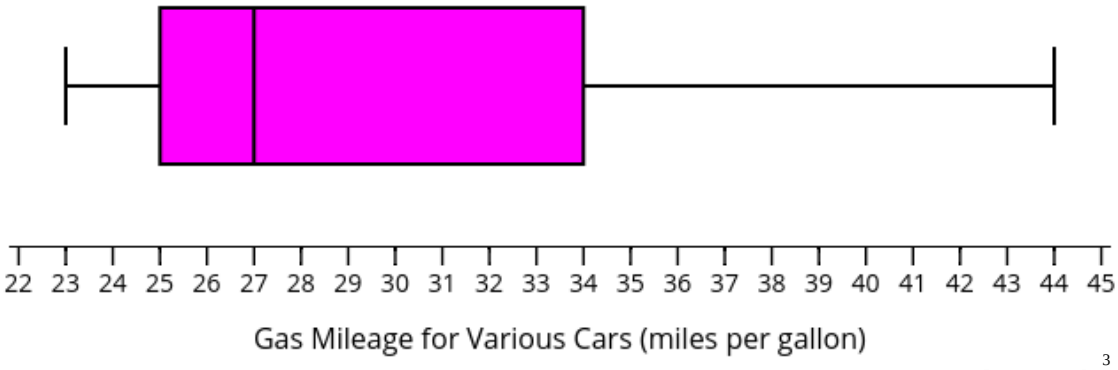
[1]

   b. Heart rate in beats per minute: 72, 65, 45, 110, 45, 51, 92, 57, 63, 70, 84



[2]

4. The box plot shows the gas mileage for various cars. Identify if the following statements are true or false. Justify your answer.



[3]

   i. The car with the best gas mileage gets 44 miles per gallon (mpg).
      True, the maximum value on the plot is 44.

   ii. It can be determined with certainty from this plot that at least one car gets 27 mpg.
      False, while the median value is 27, it does not mean that one of the values is 27 because the median value could be the mean of two values in the data set.

   iii. Half of the cars get less than 34 mpg.
      False, 75% of the cars get less than 34 mpg. A student may also state that half of the cars get less than 27 mpg.

   iv. Half of the cars get between 25 and 34 mpg.
      True, 50% of the data falls within the box or the IQR.

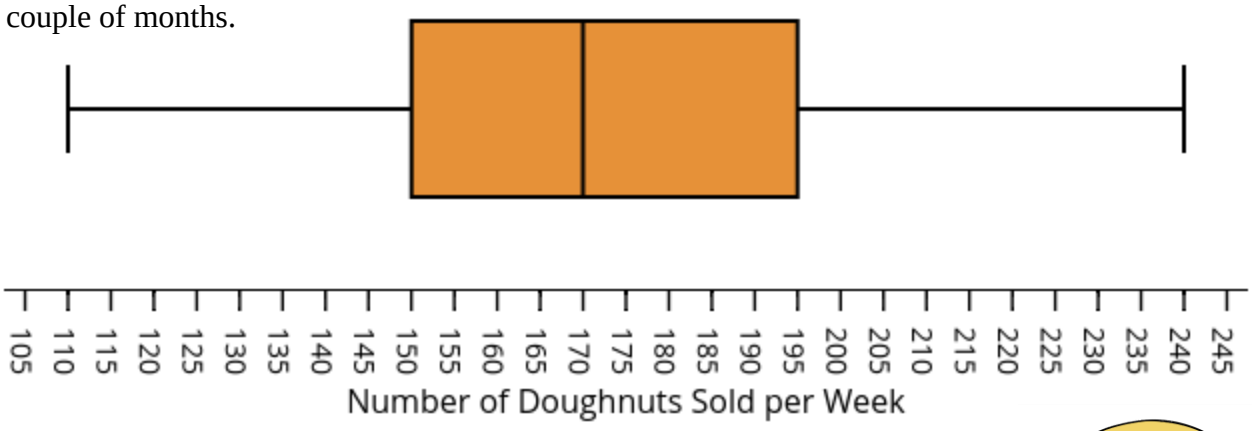
   v. More data values fall within the top section of the box than the bottom section of the box because the median line is further to the left.
      False, 25% of the data fall within each section, the interval for the bottom section is just farther to the left, meaning that the data values are closer together.

   vi. A car that gets 42 mpg is comparable to the gas mileage of most cars.
      False, a gas mileage of 42 mpg is better than most cars.

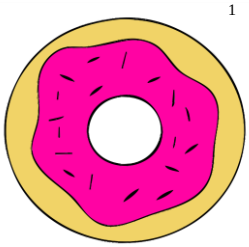
1 evatipp. "Piano Practicing" Plotly Technologies Inc., https://plot.ly/486/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Heart Rate (1)" Plotly Technologies Inc., https://plot.ly/462/~evatipp/ (Last Accessed: March 2016)
3 evatipp. "Gas Mileage (2)" Plotly Technologies Inc., https://plot.ly/447/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. Cindy has made a box plot of the number of doughnuts she has sold each week in her bakery for the last couple of months.



[1]



a. Make a statement about the data that discusses the center.
The median of this data is 170 doughnuts.

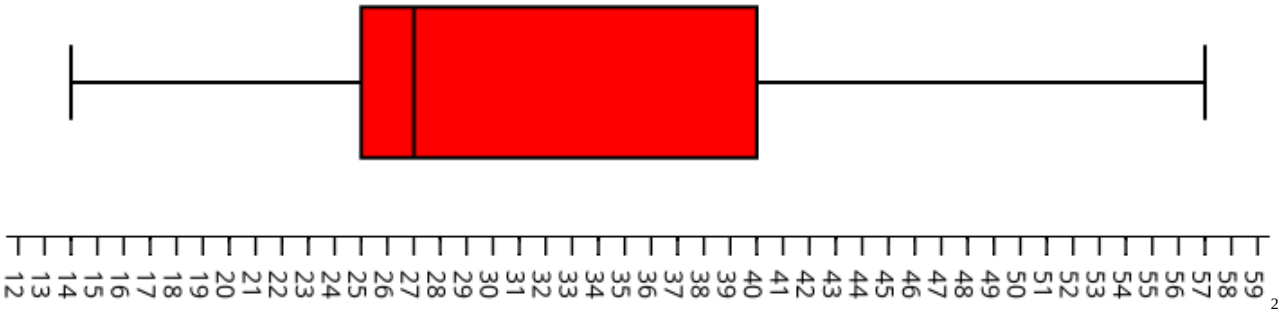
b. Make a statement about the data that discusses the shape.
The shape of the data is fairly symmetrical.

c. Make a statement about the data that discusses the spread or variability
The range of the data is 130. However, the IQR is about 45; this means the middle 50% of the data that is surrounding the median varies by 45 doughnuts.

d. How can Cindy use this information in the future?
Cindy now knows that she can sell around 170 doughnuts per week. However, she must take into account the variability. 50% of her data varies around the median by 45 doughnuts.

5. Suppose you know the following about a data set: the maximum value is 57, the 3rd quartile is 40, the IQR is 15, half of the data is below 27 and the minimum value is 14.

a. Make a box plot of the data.



[2]

b. Are there more data values above or below the median?
The number of values on each side of the median is the same, 50% above it and 50% below it.

c. Think of a context for which these values make sense.
See student answer.

1 evatipp. "Doughnuts (1)" Plotly Technologies Inc., https://plot.ly/472/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "5 number summary plot" Plotly Technologies Inc., https://plot.ly/481/~evatipp/ (Last Accessed: March 2016)

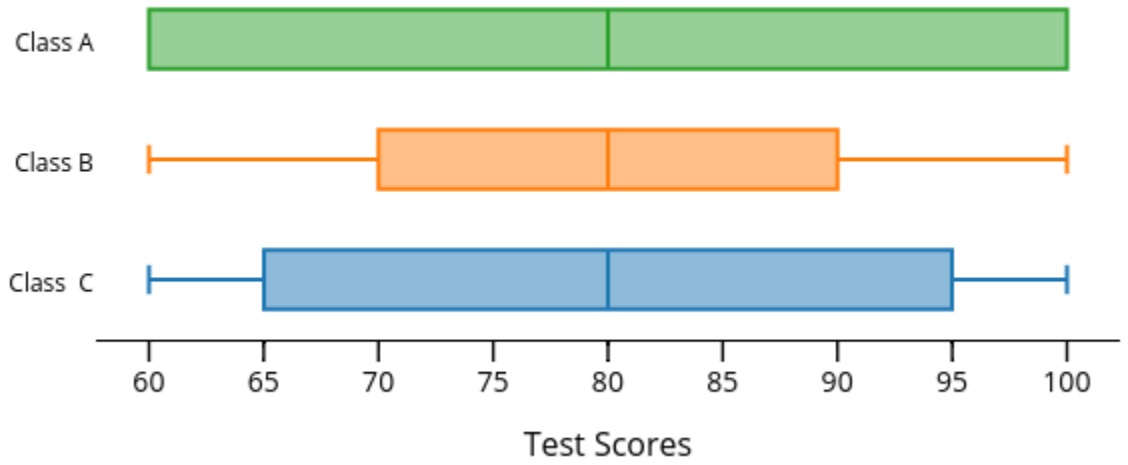
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3d Class Activity: Analyzing Box Plots

As students analyze and make comparisons between data distributions presented in box plots, they must use the features of the plot to construct arguments that explain the relationship of the data values and how they relate to the context which they represent.

1. The three box plots below represent the test scores for three different classes. Examine each plot and then discuss the questions that follow.

These questions are meant to lead a discussion about the distribution of the data in each plot; it is not expected that students write down the answer to each question. Perhaps encourage students to write down their ideas as bulleted points.



[1]

a. What is the same about these box plots and what is different?
All of the plots have a median score of 80. They also all have a range of 40. The variability, as measured by the IQR, of each set of scores is different.

b. Find the IQR for each plot and use it to compare the variability of each set of class scores.
$Class\ A\ IQR = 40$, this means that 50% of the middle scores vary by 40 points. Be sure to discuss that the value of Q1 and the min are the same in this plot. Likewise the value of Q3 and the max are the same.
$Class\ B\ IQR = 20$, this mean that 50% of the middle scores vary by 20 points.
$Class\ C\ IQR = 30$, this means that 50% of the middle scores vary by 30 points.
Class B has the least amount of variability around the median followed by class C. Class A has the most variability around the median and since the value of the 1st and 3rd quartiles are the same as the max and min values we can assume that a lot of the values are clustered at the ends of the range.

c. Make an argument for each class that supports the claim that this class performed the best of the test.
- Class A performed the best because several students must have received full credit in order for the third quartile to be equal to the maximum value.
- Class B performed the best because it has the least amount of variability around the median. This means that the scores were more consistent with each other.
- Class C performed the best because 25% of the students scored higher than 95%. This is better than Class B where 25% of the kids scored higher than 90%. It also has less variability around the median than Class A.

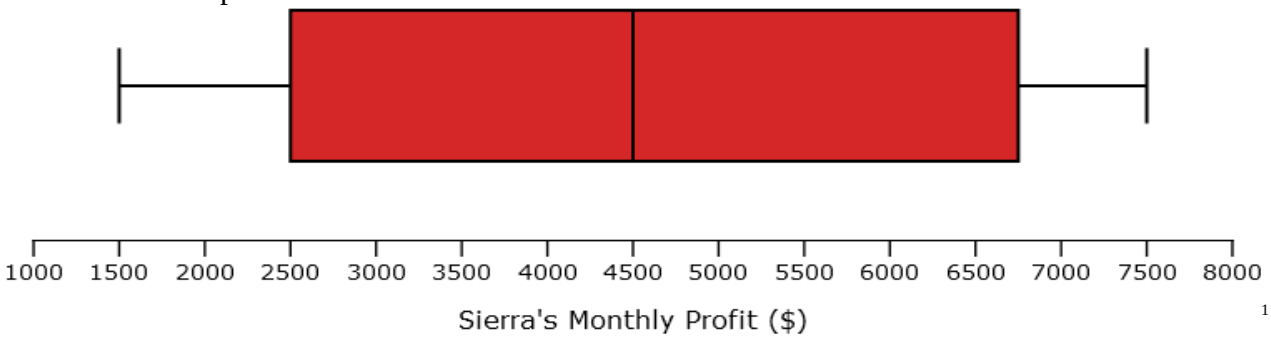
1 evatipp. "Test Scores Box Plot (2)" Plotly Technologies Inc., https://plot.ly/447/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. Sierra has a small landscaping business. Her monthly profits for the last 12 months are shown in the table.

Month	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Profit $	$1500	$1500	$3000	$4000	$5500	$7000	$7500	$7000	$6500	$5000	$3000	$2000

a. Make a box plot of the data.

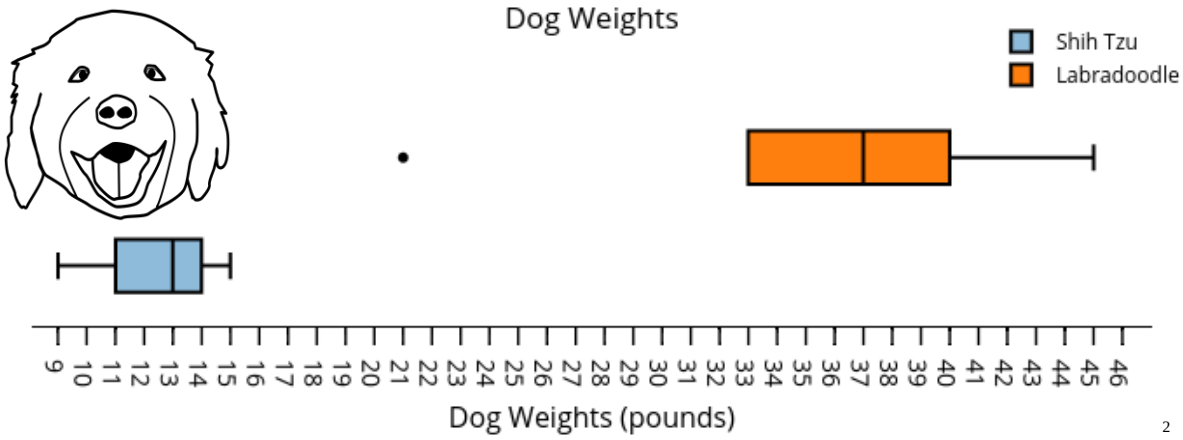


[1]

b. Write at least 3 bullet points that describe Sierra's profit for her landscaping business. Be sure to discuss the shape, center, and spread of the data.

- Sierra's median income for the year was $4500. This means that most of the time she makes around $4500.
- Since the median line is near the center of the box and the lines extending to the extreme values are close to the same length the data is symmetrical.
- The range of the data is $6000 with an IQR of $4250. This indicates that the data varies quite a bit with the middle 50% of the data falling within $4250.
- This data distribution suggests that there is quite a lot of variability in how much money Sierra makes throughout the year. This makes sense because she owns a landscaping business and people do not have as much of a need for landscaping during winter months.

3. The weight, in pounds, of two different breeds of dogs is shown in the box plots.



[2]

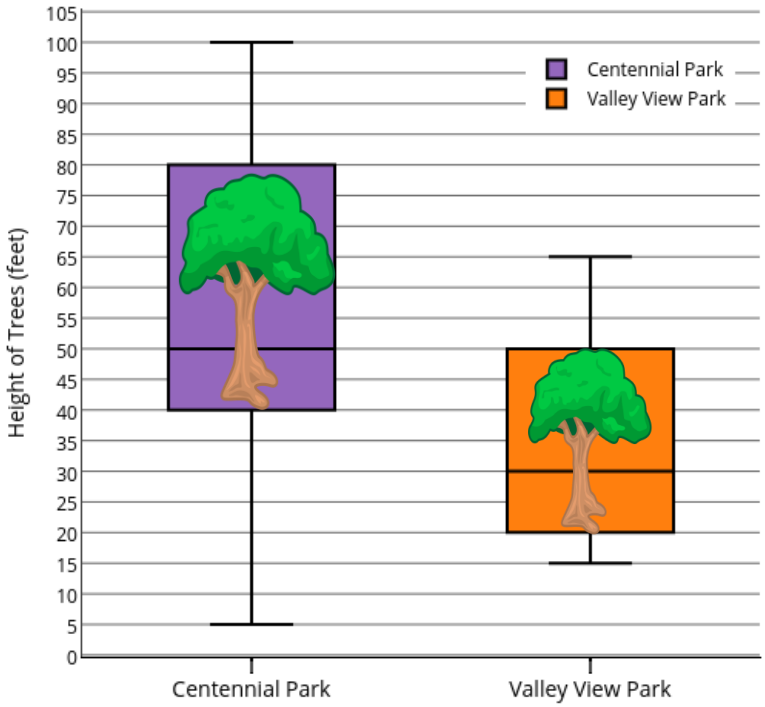
Write at least 3 bullet points that describe how the weights of the Shih Tzu compare to the weights of Labradoodles?

- The Shih Tzu weights vary from 9 to 15 pounds with most dogs weighing around 13 lbs.
- The weights of the Shih Tzu have very little variability.
- The Labradoodles median weight is 37 lbs with 50% of the weights between 33 and 40 lbs.
- The minimum weight for the Labradoodles is 21 and is an outlier.
- For the most part, Labradoodles weigh more than double the Shih Tzus and have more variability in their weights.

1 evatipp. "Landscaping Profit" Plotly Technologies Inc., https://plot.ly/476/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Dog Weights (2)" Plotly Technologies Inc., https://plot.ly/484/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. The box plots show the height of all the trees in two different city parks.



[1]

This is a vertical box plot, this is another way that box plots are often shown.

a. How tall is the smallest tree at Centennial Park?
5 feet tall

b. How tall is the tallest tree in at Valley View Park?
65 feet

c. Which park has trees that are more varied in height? Justify your answer.
Centennial Park's trees vary more in height. You can see his by looking at the size of the box and the length of the line segments extending to the max and min values. It has a greater range and IQR than the data for Valley View Park.

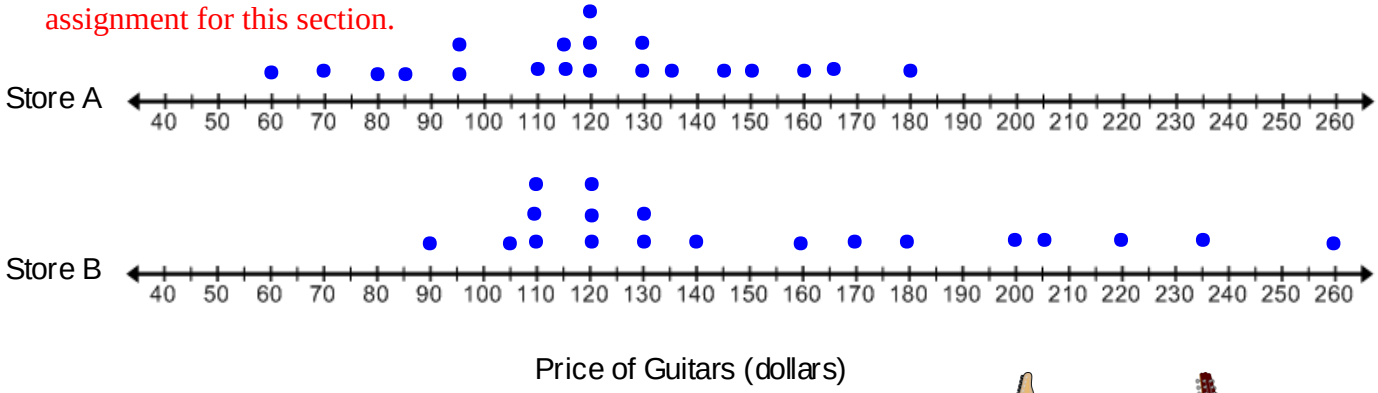
d. How does the median height of trees in Centennial Park compare to the median height of trees in Valley View Park?
Centennial Park's median tree height it 50 feet while Valley View Park's median tree height is 30 ft.

e. Give an explanation that could account for the different medians and variability between the heights of trees in these parks.
Perhaps the trees in Centennial Park are older than the ones in Valley View Park. This would account for a higher median height and variability since trees grow at different rates. You could also argue that the type of trees at Centennial Park are ones that generally grow taller and have more variability in their heights.

1 evatipp. "Height of Trees" Plotly Technologies Inc., https://plot.ly/479/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. Use the dot plots to answer each question. If desired you could assign this problem with the homework assignment for this section.

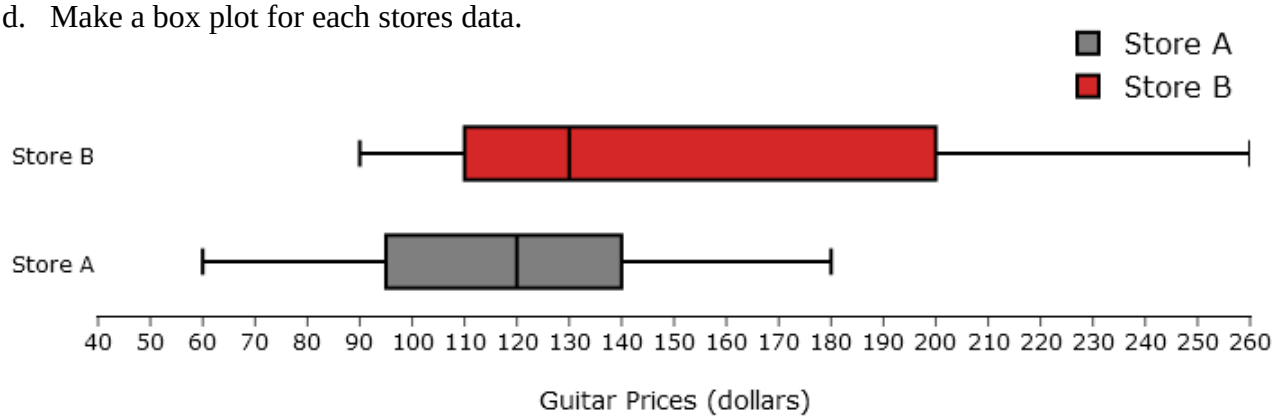


a. What does each data point represent?
Each data point represents the price of a guitar.

b. How many guitars does Store B have in their store?
Store B has 19 guitars.

c. Based off of the dot plots, which store appears to have more variability in prices of guitars. How would this be reflected in their IQR and box plot?
Store B appears to have a little more variability in the prices of guitars at their store. The IQR for store B's data should be bigger than the IQR for store A's data. The box plot for store B should be longer than the box plot for Store A with a long right whisker.

d. Make a box plot for each stores data.



[1]

e. Find the Range and IQR for each store and use them to verify which store has more variability in its guitar pricing.
The Range of Store A is 120 and its IQR is 45, the Range of Store B is 170 and its IQR is 90. Since the IQR of Store B is larger and its range is also larger, we know that Store B has more variability in its guitar prices.

f. Write a few sentences that compare the distribution of data for each store. Be sure to discuss the shape, center, and spread.
The prices of guitars at Store B are more varied than the prices for Store A. However the median price of a guitar at both stores is pretty close with a median price of $120 at Store A and a median price of $130 at store B. The data for Store A is symmetrical, you can see this in the box plot because the median line is near the middle of the box and the line segments or whiskers are about the same length. The data for Store B is skewed right, you can see this in the box plot because the median line is to the left inside the box and the right line segment or whisker is longer than the one of the left.

1 evatipp. "Guitar Prices (1)" Plotly Technologies Inc., https://plot.ly/491/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Spiral Review

Directions: Use a visual model to solve each problem.

1. $\frac{2}{3}+\frac{1}{2}$ $1\frac{1}{6}$	
2. $\frac{3}{4}+\frac{1}{2}$ $1\frac{1}{4}$	
3. $1\frac{3}{4}-\frac{1}{3}$ $1\frac{5}{12}$	
4. $1\frac{1}{2}+\frac{1}{5}$ $1\frac{7}{10}$	
5. Samantha is cutting fabric. She has a piece that is $\frac{7}{12}$ of a foot long. She cuts off $\frac{1}{3}$ of a foot. How much fabric does she have left on the original piece? $\frac{1}{4}$	

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3d Homework: Analyzing Box Plots

1. Create a statistical question of your own that you can answer by surveying people in your class. Then ask at least 15 people in your class your question and record your data below.

   a. Write your question here:____Answers will vary_____

   b. Record your data in the table below.

Student	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Response															

   c. What is the attribute that is being measured from your question?

   d. What is the unit of measurement for your attribute?

   e. Display your data in a box plot, be sure to label your number line and give it a title.

   f. What is the median of your data?

   g. What is the overall shape of the data?

   h. Discuss the variability or spread of the data.

   i. Use the distribution of the data in your box plot to answer your question.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.3e Self Assessment: Section 3.3

Consider the following skills/concepts. Rate your comfort level with each skill/concept by checking the box that best describes your progress in mastering each skill/concept. Corresponding sample problems, referenced in brackets, can be found on the following page.

Skill/Concept	Minimal Understanding 1	Partial Understanding 2	Sufficient Mastery 3	Substantial Mastery 4
1. Find and interpret the range for a data set.	I can find the range of a data set from a list or dot plot. I do not know how to interpret the range or find it in a box plot.	I can find the range of a data set that is presented in a list, dot plot, or box plot. I struggle to interpret it from the given context.	I can find and interpret the range of a data set that is presented in a list, dot plot, or box plot. I do not understand why you cannot find the exact range from a histogram.	I can find and interpret the range of a data set that is presented in a list, dot plot, or box plot. I understand and can explain why you cannot find the exact range from a histogram.
2. Find and interpret the mean absolute deviation (MAD) for a data set.	I can sometimes find the MAD but I don't really know why it is useful.	I know how to find the MAD for a data set but cannot interpret its meaning.	I know how to find the MAD for a data set and can interpret its meaning for a given context.	I know how to find the MAD for a data set and can interpret its meaning for a given context. I understand that the MAD is a measure of variability that relates to the mean.
3. Find and interpret values of the 5-number summary for a data set.	I can find some of the 5-number summary values. I don't really know what they mean.	I can find and interpret at least 3 of the 5-number summary values.	I can find and interpret at least 4 of the 5-number summary values.	I can find and interpret all of the 5-number summary values.
4. Find the interquartile range (IQR) for a data set and interpret its meaning.	I can sometimes find the IQR but I don't really know why it is useful.	I know how to find the IQR for a data set but cannot interpret its meaning.	I know how to find the IQR for a data set and can interpret its meaning for a given context.	I know how to find the IQR for a data set and can interpret its meaning for a given context. I understand that the IQR is a measure of variability that relates to the median.

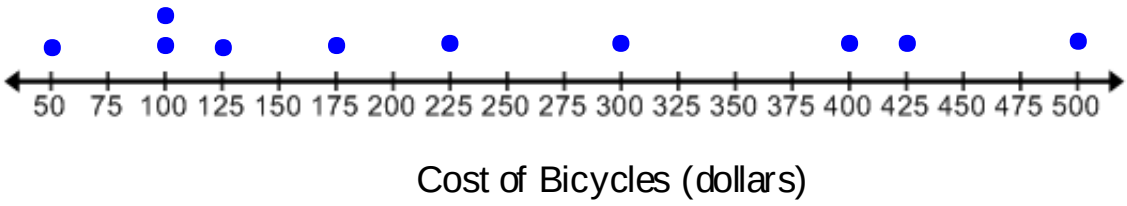
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

5. Display numerical data in a box plot, use key terms to describe its shape, and interpret the distribution of data.	I can display data in a box plot but often don't know how to label it appropriately. I don't know how to describe the shape or interpret the distribution of data.	I can display data in a box plot that is accurately labeled. I can identify some of the terms used to describe its shape. I struggle to interpret the distribution of data.	I can display data in a box plot that is accurately labeled. I can identify and use terms such as max, min, quartiles, inter-quartile range, symmetrical and outliers to describe the shape and interpret the distribution of data. I cannot determine if the data is skewed.	I can display data in a box plot that is accurately labeled. I can identify and use terms such as max, min, quartiles, inter quartile range, skewed, symmetrical and outliers to describe the shape and interpret the distribution of data.
6. Draw conclusions about two or more data sets by comparing their box plots.	I can draw conclusions about the center when comparing box plots but I don't know how to compare the shape or variability.	I can draw conclusions about the shape, center, and spread when comparing box plots.	I can draw conclusions about the shape, center, and spread when comparing box plots. I take into account the context in which the data is given.	I can draw conclusions about the shape, center, and spread when comparing box plots. I take into account the context in which the data is given. I understand that when comparing box plots I must take into account how they are each scaled.

Sample Problems for Section 4.3

Square brackets indicate which concept/skill the problem aligns to.

1. The dot plot below shows the prices of several bikes as a bike shop. [1][2]



   a. Find and interpret the range for this data distribution.
   The range is $450. This means that the bikes at this store vary in price by $450.

   b. Find and interpret the mean.
   The mean is $240. This means that the average cost of a bike at this store is $240.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

c. Based off of the shape of the data would you say that this distribution has a lot of variability or a little bit of variability? Justify your answer.
It appears that there is a lot of variability for the cost of bikes at this store. The data values are very spread out.

d. Find and interpret the mean absolute deviation (MAD).
The MAD is $133. This means that the average deviation of data values from the mean is $133. This indicates a lot of variability.

e. At a different bike store the mean price for a bike is also $240 and the MAD is $50. Do the prices for bikes at the second store have more or less variability than the prices for bikes at the first store? Justify your answer.
The prices at the second store have less variability because their MAD is $50. This means that the average deviation of data values from the mean of $240 is $50.

2. A swimming pool maintenance worker is interested in average number of gallons of water that the backyard pools that he services holds. The amount of water that each pool holds is recorded below. [1][3][4]

Amount of water in gallons: 3105, 4075, 1100, 2050, 2325, 4400, 2475, 1500, 3520

a. Find the values for the 5-Number Summary for this data set.
Max- 4400 gallons

Min- 1100 gallons

Median- 2475 gallons

Lower Quartile (1st quartile)- 1775 gallons

Upper Quartile (3rd quartile)- 3797.5 gallons

b. What percent or fraction of the swimming pools have more than 2475 gallons of water?
50% or $\frac{1}{2}$

c. $\frac{1}{4}$ or 25% of the pools have less than how many gallons of water?
1775 gallons

d. What percent of the swimming pools lie between the first quartile and the third quartile?
50%

e. Find and interpret the range for this data.
The range is 3300 gallons of water. This means that the number of gallons of water that the swimming pools hold are within 3300 gallons of each other.

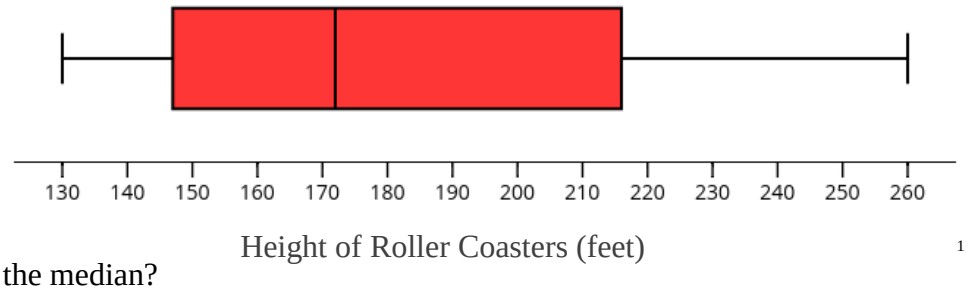
f. Find and the interquartile range (IQR) for this data distribution.
The IQR is 2022.5 gallons. This means that the middle 50% of the swimming pools hold within 2022.5 gallons of each other.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. The data below shows the heights of all of the roller coasters at an amusement park. [1][4][5]

Height of Roller Coasters (feet)												
203	163	130	260	144	208	150	161	210	222	135	172	230

a. Make a box plot of the data



[1]

b. What is the median?

The median height of a roller coaster at this park is 172 feet.

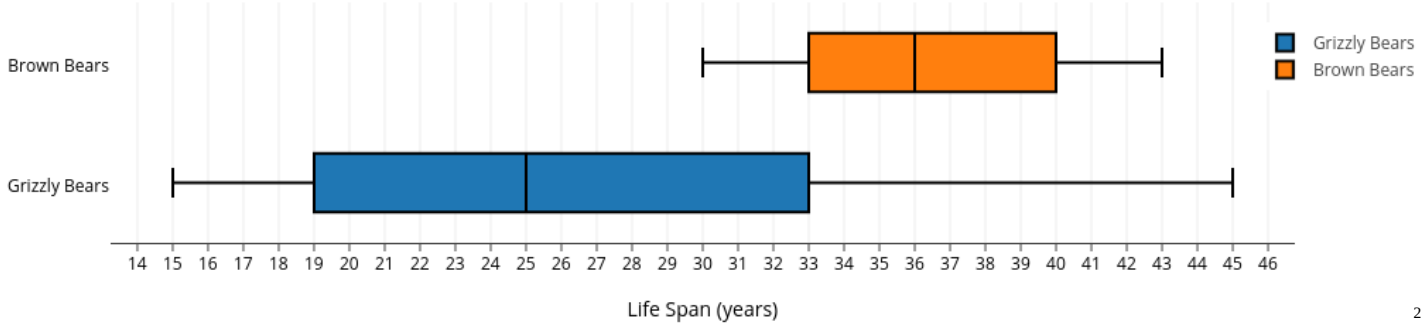
c. Use the box plot to describe the shape of the data distribution.

The data is slightly skewed right with a median of 172. Students might argue that the height of 260 feet is an outlier.

d. Find and interpret the range and IQR.

The range is 130 feet. This means that the height of roller coasters at this park vary by 130 feet. The IQR is 69 feet. This means that the middle 50% of the roller coaster heights vary by 69 feet.

4. The box plots below show the life spans of two different species of bears in a nature preserve. Compare and contrast the ages of the bears. Be sure to discuss shape, center, and variability. [4][6]



[2]

The data for the Grizzly Bears is slightly skewed right while the data for the Brown Bears is fairly symmetrical. The median life span for Grizzly Bears is 25 years. This is less than the median life span for Brown Bears which is 36 years. The life span of Grizzly Bears has a lot more variability than that of Brown Bears. This is true because the IQR for Grizzly Bears is 14 years and the IQR for Brown Bears is 7 years.

1 evatipp. "Price of Bikes" Plotly Technologies Inc., https://plot.ly/plot# (Last Accessed: March 2016)
2 evatipp. "Life Span of Bears" Plotly Technologies Inc., https://plot.ly/522/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Section 4.4: Intrepret and Draw Conclusions about Data

Section Overview:

This section acts as a culmination of the skills of statistical analysis that students have learned throughout the chapter. Students begin to learn how to choose which statistical models and measures will be useful in interpreting data and answering questions. Depending on what is desired, they must choose whether a dot plot, histogram, or box plot best represents the data. They also learn that different information can be obtained from different displays. Understanding when a measure of center or a measure of variability is useful in answering a question is also important. Finally, students are required to develop a deeper understanding of these statistical measures as they compare, interpret, and draw conclusions about several data sets in a variety of contexts, and in turn, collect and analyze data of their own in order to answer a statistical question.

Primary Concepts and Skills to Master in this Section:

By the end of this section, students should be able to:

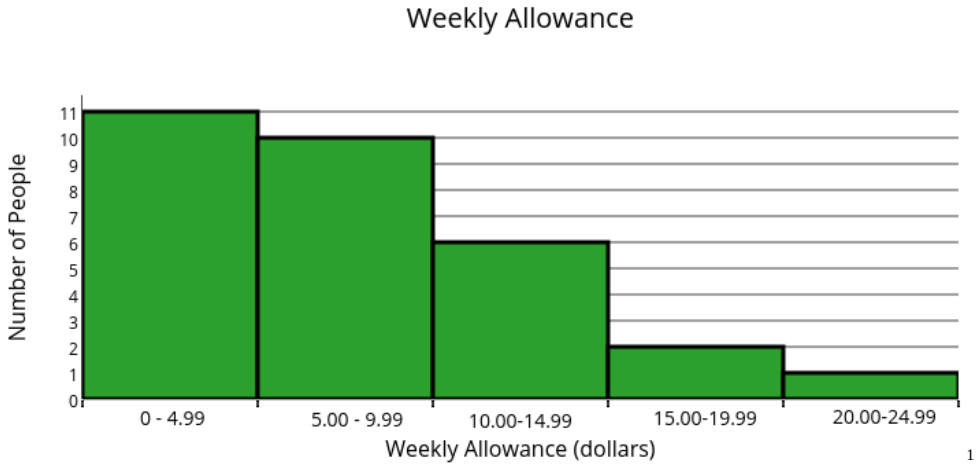
1. Determine whether a measure of center or a measure of variability should be used to answer a statistical question.
2. Interpret, compare, and draw conclusions about data distributions by using statistical models and appropriate measures of shape, center and spread.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

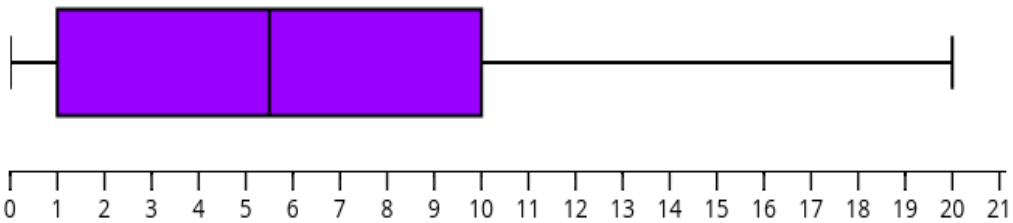
4.4a Class Activity: Make, Interpret and Draw Conclusions About Data Part 1

Different data displays and numerical measures act as tools that can be used to answer different questions that relate to the data distribution. Students learn how to choose these tools appropriately for a given situation.

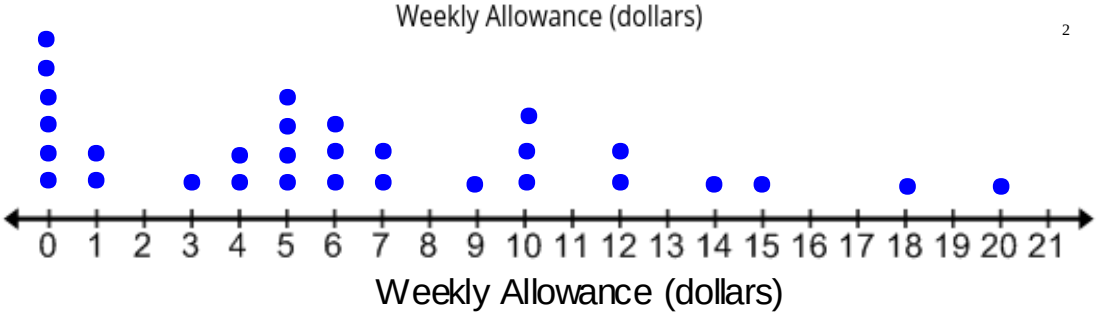
Donovan, Marcus, and Luther have collected data together on the amount of allowance other students in their class get per week. Each of them displayed the data in a different way. Use their data displays to answer the questions that follow.



1



2



1. What is the shape of the data? Which data display did you use to determine this answer? Explain why you chose this display.
   The data is clustered around $0 and $5 with peaks at these values, it is slightly skewed right. The students might argue that it is easiest to see that the data is skewed in the histogram or the dot plot. However, in order to really see the clusters and peaks you would have to use the dot plot. Also the intervals of the histogram will affect how you see the shape of the data. They could also argue that they can see that the data is skewed right in a box plot because of the tail on the box plot but there is really no way to know if that the long tail is only due to a single outlier.

1 evatipp. "Weekly Allowance" Plotly Technologies Inc., https://plot.ly/512/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Weekly Allowance Box Plot" Plotly Technologies Inc., https://plot.ly/513/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. What is the minimum? Which data display did you use and why?
   The minimum is 0. This is very easy to see on a box plot, it is the value at the end of the line segment (whisker) on the left of the box. You can also see the minimum value on a dot plot. You cannot determine the minimum from a histogram because you do not know the exact values in the 0 – 4.99 interval.
3. How many people were surveyed? Which data display did you use and why?
   30 people were surveyed. You can see this in the histogram by summing the frequencies from each interval. You can determine this from the dot plot by counting the number of dots.

4. What is the median of the data? Which data display did you use and why?
   The median of the data is $5.50. This can most easily be seen on the box plot, it is the value of the line inside the box. Students may also argue that they can calculate the median from the data values in the dot plot. You cannot determine the median from the histogram, although you may be able to determine which interval it falls in.

5. Describe the variability of the data set. Which data display did you use and why?
   The range of the data is $20.00 and the IQR is $9.00. The range is found by subtracting the max and min which can easily be found in the dot plot or box plot. The IQR is found by finding the difference of Q3 and Q1, this is most easily found in the box plot. However you can find the quartiles from the dot plot as well. A student may give the MAD which is $4.40 and can be calculated from the values in the dot plot.

6. What is the mean? Which data display did you use and why?
   6.5; the only data display that you can use to find the mean is the dot plot because it is the only one that shows every exact data value.

As a class you may want to summarize the pros and cons of each data display before moving on.

For each of the five statistical questions below (Questions A through E), decide if you would answer the question by considering center or considering variability in the data distribution.

7. A car repair shop keeps track of the number of days it takes to service a car at their shop.

   Question A: For cars sent to this shop, what is a typical number of days that it takes for a car to be repaired?

   Measure of center; this question asks about a typical value it would be answered by considering a measure of center (mean or median).

8. Two different cities record their average rainfall for each month during the year.

   Question B: On average, which cities receive more rainfall per year?

   Measure of center; this question asks which city has a higher average which would be answered by comparing the centers of each distribution.

   Question C: Which city has monthly rainfalls that are more consistent (more similar to one another) from month to month?

   Measure of variability; this question asks about the consistency of rainfall from month to month within a city. It would be answered by comparing the variability of each data distribution. The city with less variability would be more consistent.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

9. Bags of apples are often measured by weight rather than the number of apples each bag contains. Suppose that you own a grocery store and you carry two different types of apples in your store, Red Delicious and Granny Smiths. For the past year, for each brand of apple, you have kept track of the number of apples contained in each bag delivered to you.

   Question D: If you wanted to ensure that your customers purchased bags of apples that each have the same number of apples in them, should you buy Red Delicious apples or Granny Smith apples in the future?

   Measure of variability; in order for each customer to get the same number of apples you would want to consider the brand with the smallest variability.

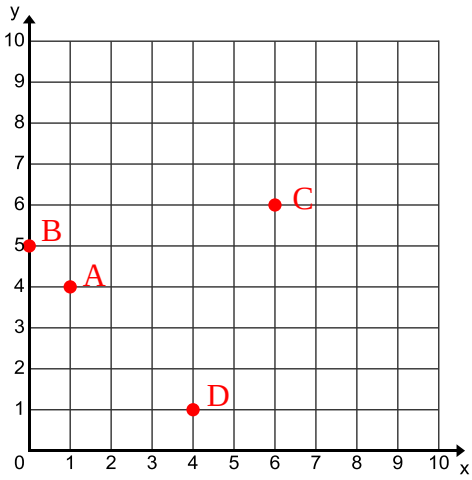
   Question E: If you wanted to ensure that your customers purchased bags of apples that contain the greatest number of apples, should you buy Red Delicious apples or Granny Smith apple in the future?

   Measure of center; if you wanted to ensure that your customers purchased bags with the greatest number of apples you would want to consider the brand with the greatest measure of center.

Spiral Review

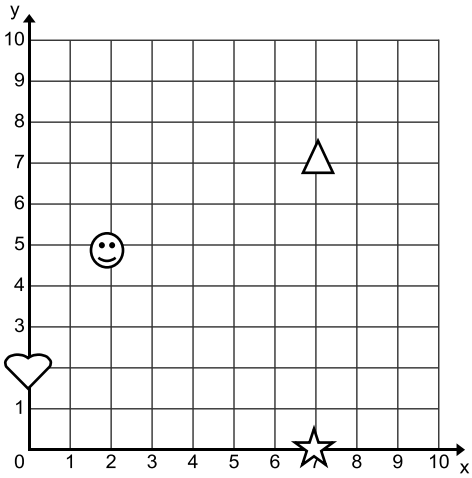
1. Graph and label each set of points on the coordinate plane given below.

   a. A (1,4)
   b. B (0,5)
   c. C (6,6)
   d. D (4,1)



2. Name the ordered pair for each shape shown on the coordinate plane below.

   a. Star (7,0)
   b. Triangle (7,7)
   c. Smiley (2,5)
   d. Heart (0,2)



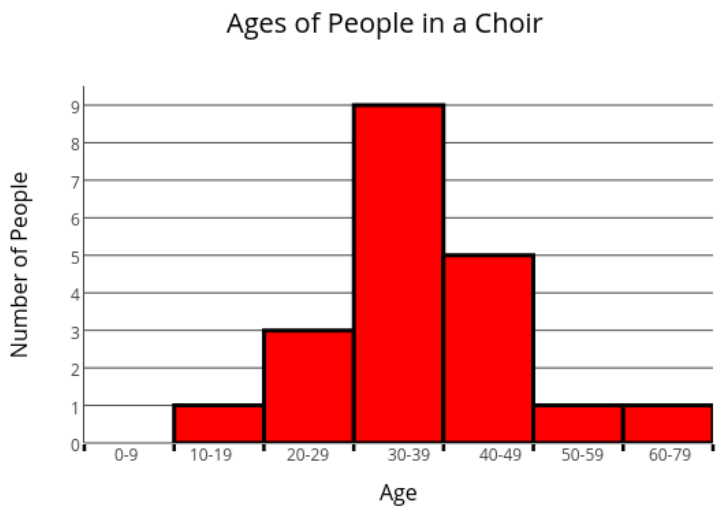
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.4a Homework: Make, Interpret, and Draw Conclusions About Data Part 1

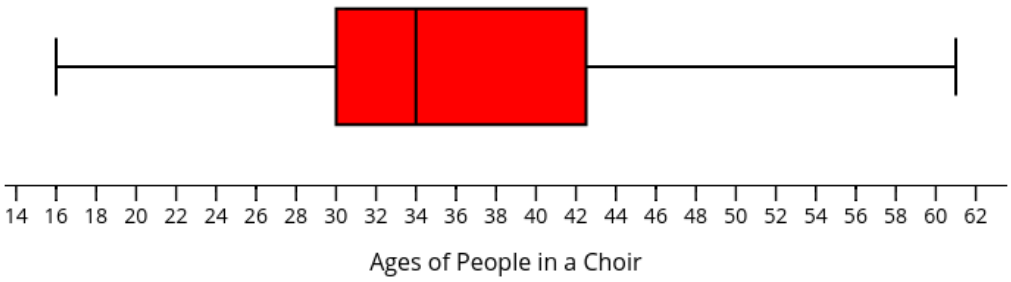
1. Below are the ages of 20 people that sing in a choir together.

   16, 21, 24, 29, 30, 30, 32, 32, 32, 34, 34, 35, 39, 42, 42, 43, 45, 47, 52, 61

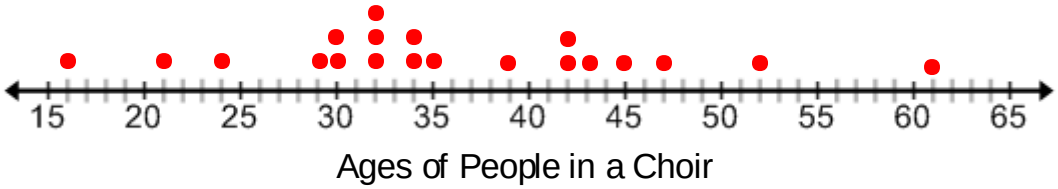
   a. Make an appropriate graph to summarize these ages.
   Students may choose any of the following graphs to represent the data. However if they choose to represent the data with a histogram they will not be able to find exact number numerical measures of center and variability to answer the following questions. They could answer them generally though.



1



2



1 evatipp. "Ages of People in a Choir (1)" Plotly Technologies Inc., https://plot.ly/509/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Ages of People in a Choir" Plotly Technologies Inc., https://plot.ly/508/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

b. Describe the distribution of ages for people in this choir. Be sure to describe shape, center and variability.
The distribution of ages is centered at approximately 35 years (the median is 34 and the mean is 36). The range is 45 years, the interquartile range is 12.5 years, and the MAD is 8.3 years. The shape of the data is slightly skewed right and peaks between 30 and 35 years.

c. What is the typical age for people in this choir? State which kind of numerical measure you used to answer this question and explain your choice.
A typical age is around 34 years if you use the median and 36 years if you use the mean. A measure of center would best answer this question because it is asking for a typical age. You can use either the mean or median as a measure of center because there are no major outliers and the shape is fairly symmetrical.

c. Are most of the people in the choir around the same age? State which kind of numerical measure you used to answer this question and explain your choice. For the median of 34 years, the range is 45 years and the IQR is 12.5 years. This means that all the people in the choir are within 45 years of each other and the middle 50% of people are within 12.5 years of each other. For the mean of 36 years, the MAD is 8.3 years. This means that the ages of the people in the choir are on average within 8.3 years of 36 years. A measure of variability would best answer this question because it is asking how consistent the ages are in the choir. You would need to choose your measure of variability based off of the measure of center that was chosen in the part a.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.4b Class Activity: Make, Interpret, and Draw Conclusions About Data Part 2

In this lesson students need to gather all their statistical tools in order to appropriately model and make statistical arguments about data.

1. Choose the best term from the box to complete each sentence.

Mean	Median	Statistical Question	Mean Absolute Deviation (MAD)	Quartiles	Interquartile Range (IQR)

a. The _____Quartiles_____ are the values of the points that split the data into four equal parts.

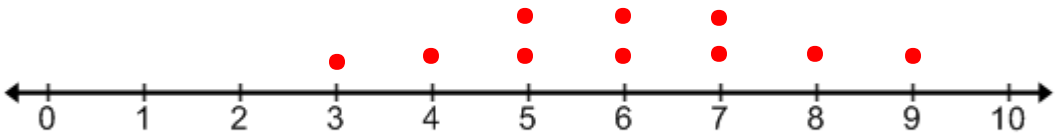
b. The_____Mean Absolute Deviation (MAD)_____ is a measure of variation. It is computed by finding the *mean* of the absolute deviations in the data set.

c. The_____Median_____ is found by ordering the data from least to greatest and then finding the middle number. If there are two numbers in the middle, then the median is the mean of those two middle numbers.

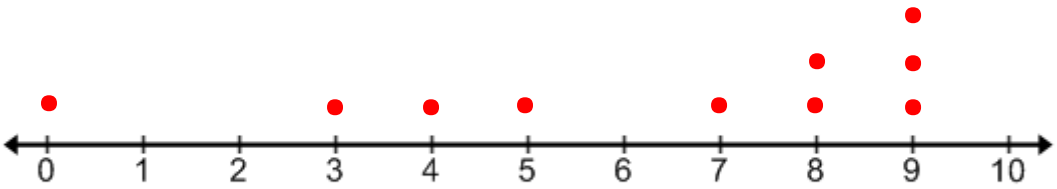
d. A_____Statistical Question_____ generates a variety of answers rather than a single answer.

2. Draw a dot plot with at least 10 data points to match the given conditions.
Answers may vary, sample answers are given.

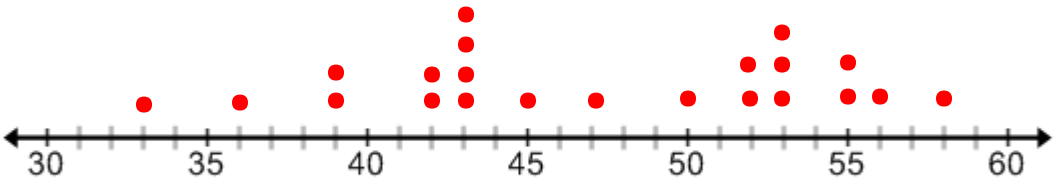
a. A dot plot that is symmetrical and has a mean of 6.



b. A dot plot that is skewed left and has an outlier.

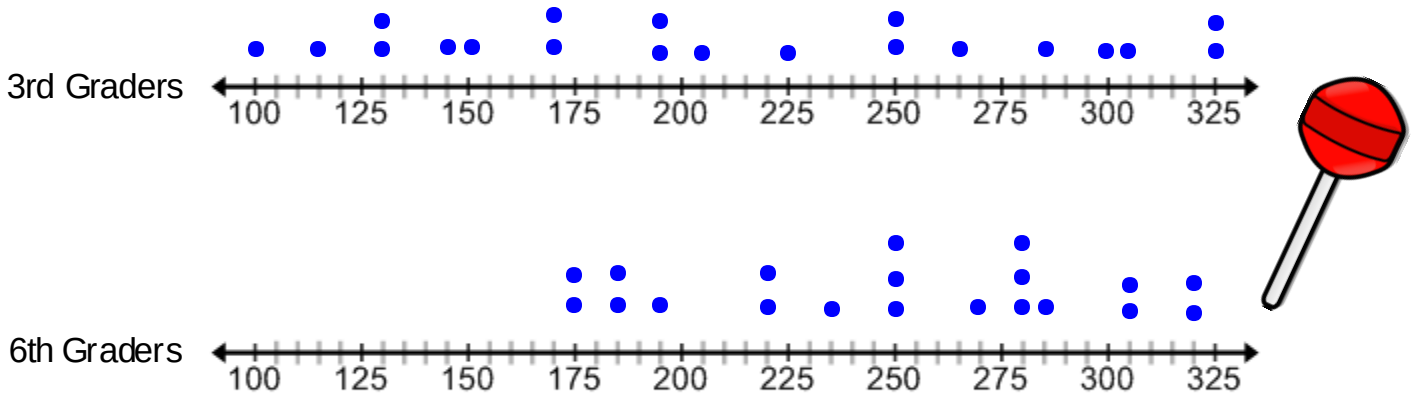


c. A dot plot that has two peaks and a range of 20.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. The dot plot below shows the number of licks it takes for 20 third graders and 20 sixth graders to get to the center of the Tootsie Pop sucker.



Number of Licks to Reach the Center of a Tootsie Pop

If you feel that your students struggle to give strong statistical analysis in writing, ask them to write down key ideas in a bulleted list or verbally give their arguments aloud to a neighbor or to you.

a. Describe the data distribution for the 3rd graders. Be sure to discuss the shape, center, and spread.
The shape of the data for the 3rd graders appears to be fairly symmetrical. The center of the data is around 200 licks, the median is exactly 200. The data is pretty spread out within the range of 225 licks and an IQR of 127.5 licks.

b. In general does it take a 3rd grader or 6th grader more licks to reach the center of a tootsie pop? Justify your answer with statistical measures.

Students must first identify that a measure of center should be used to answer this question. They must then identify which measure of center would be best. Ask students to justify their choices and share with the class.

The center for the 6th graders appears to be greater. Based off of the shape of both data sets you can use either measure of center to compare the average number of licks. The median for 3rd graders is 200 licks and the median for 6th graders is 250 licks. These measures of center show that, statistically, it takes 6th graders more licks to get to the center of a tootsie pop. Students can make a similar argument by comparing the means.

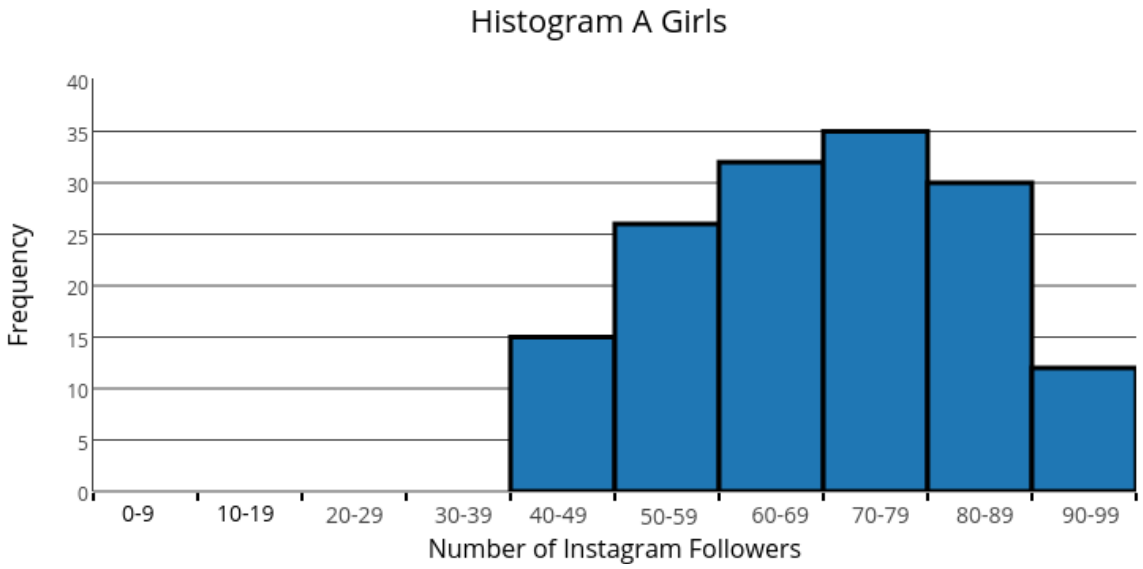
c. Which group of students is more consistent in the number of licks that it takes them to get to the center of a tootsie pop? Justify your answer with statistical measures.

Students must first identify that a measure of variability should be used to answer this question. They must then identify which measure of variability would be best. Ask students to justify their choices and share with the class.

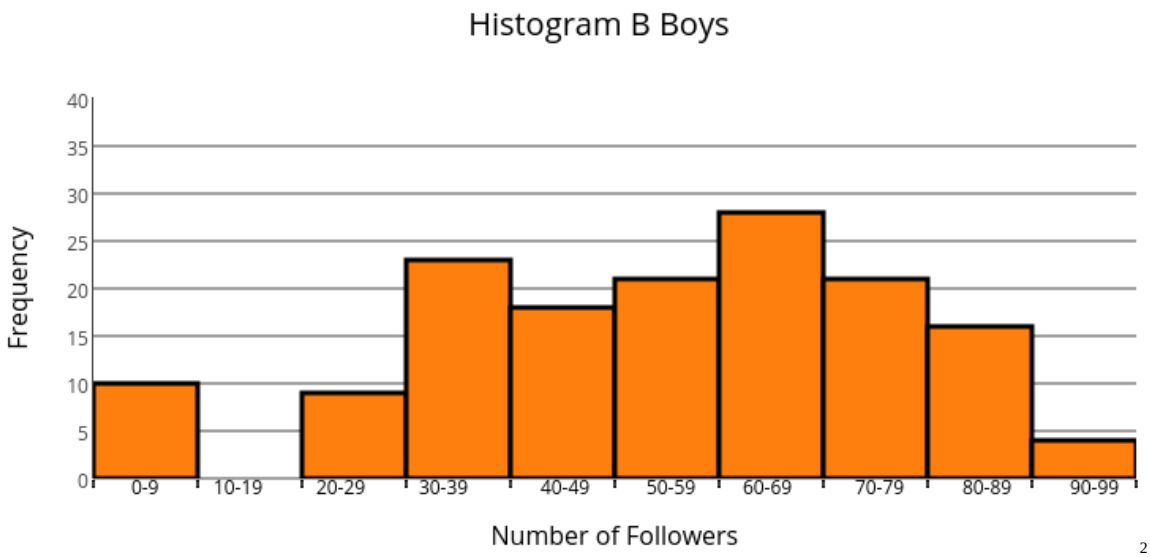
The 3rd graders appear to have a greater spread for their data. The range for the 3rd graders is 225 licks and the IQR is 127.5 licks. This means that for the 3rd graders the middle 50% of students were within 127.5 licks of each other. The range for the 6th graders is 145 licks and the IQR is 75 licks. This means that for the 6th graders the middle 50% of students are within 75 licks of each other. Statistically the number of licks that it takes a 3rd grader to get to the center of a Tootsie Pop has more variability. In other words, the 6th graders were more consistent with the number of licks it took them to get to the center of the Tootsie Pop.
Students can make a similar argument by comparing the MAD.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

2. Histogram A shows the number of Instagram followers that 150 teenage girls have. Histogram B shows the number of Instagram followers that 150 teenage boys have.



1



2

a. Describe the data distribution for the girls. Be sure to discuss the shape, center, and spread.
The shape of the data is fairly symmetrical; the center of the data appears to be around 70 – 79 followers. The spread of the data is between 40 and 99 followers.

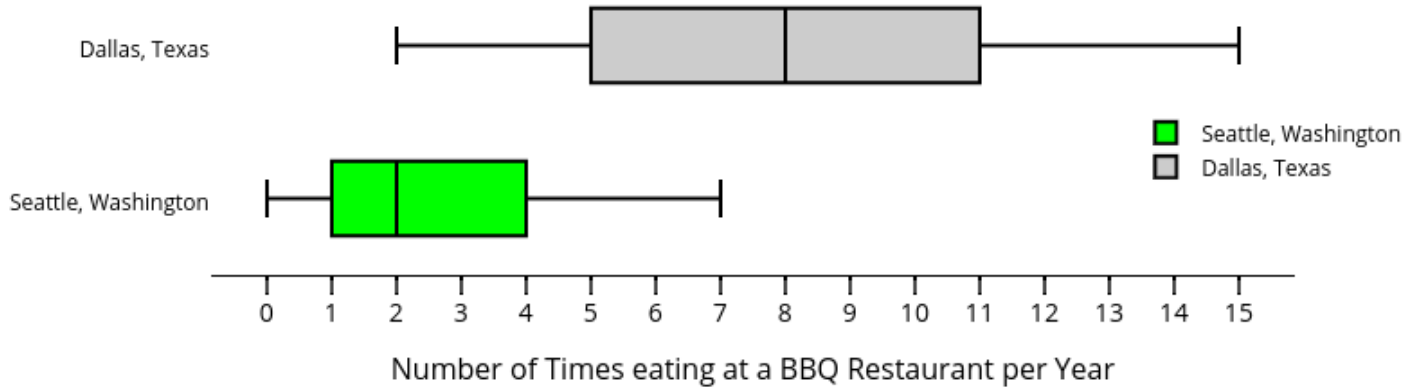
b. On average do girls or boys have more Instagram followers? Is the data for girls and the data for boys centered about the same place? If not, which one has the greater center?
It is not possible to use numerical measures of center to compare the averages from data in a histogram. However, you can still make inferences based off of the shape of the data. It appears as though the data set for the girls has a greater center.

c. Which group of people has more consistency in the number of Instagram followers that they have?
Again it is not possible to compare the variability with a numerical measure from data in a histogram. However, you can still make inferences based off of the shape of the data. The boy's data has greater spread or variability. This means that the number of followers that boys have on Instagram is less consistent than the number of follower that girls have.

1 evatipp. "Histogram A Girls Instagram" Plotly Technologies Inc., https://plot.ly/500/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Histogram B Boys" Plotly Technologies Inc., https://plot.ly/503/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. The box plots below consist of data on the number of times that 50 people who live in two major cities eat at a BBQ restaurant each year.



[1]

a. Describe the data distribution for the Seattle Washington. Be sure to discuss the shape, center, and spread.
The number of times that people eat at a BBQ restaurant per year in Seattle is centered around 2 times per year. The range of the data is 7 with an IQR of 3. This means that 50% of the people eat at a BBQ restaurant within 3 times of each other. The data is slightly skewed right; the median line is to the left of the center of the box. This means that there may be a few people that are outliers, eating at a BBQ restaurant 7 times per year.

b. Do more people typically eat at BBQ restaurants that live in Dallas, Texas or that live in Seattle, Washington? Justify your answer with statistical measures. Is the data for Seattle and the data for Dallas centered in about the same place? If not, which one has the greater center?
The median number of visits for Dallas is higher with 8 visits per year than Seattle which is 2 visits per year. This means that statistically more people that live in Dallas eat at BBQ restaurants than people that live in Seattle.

c. Which city has more consistency in the number of people that eat at BBQ restaurants?
The range for Dallas is 13 with an IQR of 6. The range for Seattle is 7 with an IQR of 3. Thus the spread for Dallas is greater. This means there is much more variability in the number of times that people who live in Dallas go to a BBQ restaurant.

Remind students that while the range and IQR for Dallas may be rather small numbers, (13 and 6), they are rather large for this context. We are talking about the number of times that people eat at a BBQ restaurant per year.

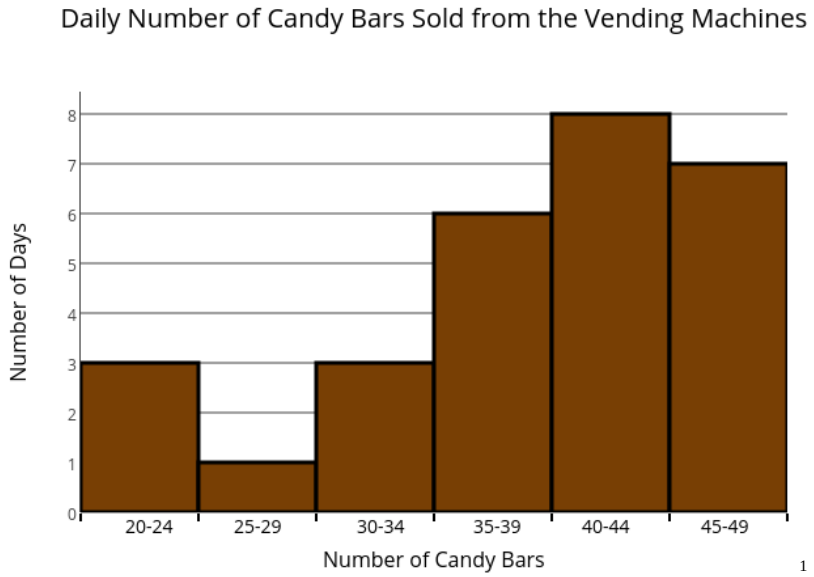
d. Why might this information be useful?

This information could be useful to someone that might be considering in which city to open a BBQ restaurant. They would want to choose a city where people eat BBQ more often.

1 evatipp. "BBQ Restaurants" Plotly Technologies Inc., https://plot.ly/505/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4. Roman thinks that his school needs to get another vending machine in the school cafeteria because there is always a long line to use the machine at lunch time. The principal has told everyone that they cannot get another vending machine unless they can show that on average at least 40 candy bars are sold each day. Roman conducts a survey and recorded his results in the histogram below.

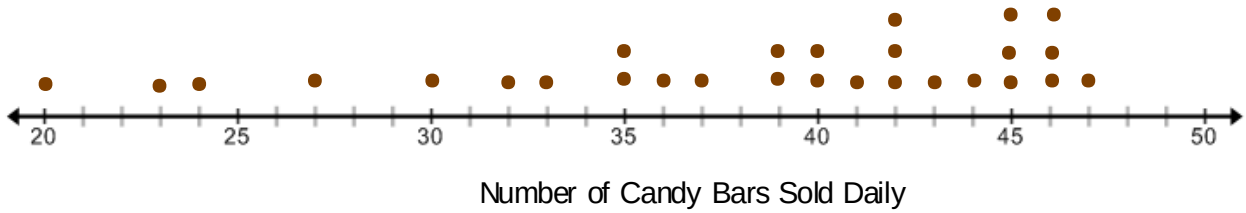


[1]

a. What possible arguments could Roman give to his principal that on average at least 40 candy bars are sold from the vending machine each day?

Assess how much guidance your students might need to answer the question above. You can provide them with a series of discussion questions to scaffold their thinking. Possible discussion questions are below.
What statistical question could be answered based off of this data distribution?
How do you think Roman collected this data?
How many days did Roman collect data on the number of candy bars sold?
Can you determine on how many days exactly 40 candy bars were sold?
On how many days were at least 40 candy bars sold? What percentage of the days is this?
How would you describe the shape of this data?
Use the histogram to estimate the median number of candy bars sold.
Based off of the shape of the data, do you think that the mean number of candy bars sold will be less than, greater than, or equal to the median number of candy bars sold?

b. Roman decides to present his data in a dot plot in addition to the histogram. Use the dot plot to find the exact *mean* and median number of candy bars sold.



$Mean = 38\ bars\ ;\ Median = 40\ bars$

Possible discussion questions might include: What is the benefit of displaying data in a dot plot over a histogram? How would you describe the shape of the data based off of the dot plot? Are there any peaks or clusters? What about outliers?

1 evatipp. "Daily Number of Candy Bars Sold from Vending Machines" Plotly Technologies Inc., https://plot.ly/516/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

c. In addition to looking at measures of center, Roman wants to consider variability. Find numerical summaries that can help you discuss the variability of the data. Interpret what these numerical summaries mean.

Students may find the range, the MAD, or the IQR. The range equals 27, this means that the number of candy bars sold varied by 27 over the course of Roman's survey. The MAD is approximately 6.14, this means that the average deviation of data values from the mean is 6.14 candy bars; $IQR = 10.5$, this means that the middle 50% of the data is within 10.5 candy bars of each other.

Possible discussion points might be: Why does Roman want to consider variability? Why is this important? Use the dot plot to find the 5-number summary for the data. Make a box plot of the data. What does the box plot tell you about the data? Which measure of center and measure of variability do you think Roman should use to represent this data and why?

d. Use the information found in the previous questions to refine your argument about convincing Roman's principal that in general at least 40 candy bars are sold each day.

Sample argument is given.

While the mean number of candy bars sold each day is 38, the median number of bars sold is 40. It can be argued that the median is a better measure of center for this data because the data is skewed left. The IQR for this data is 10.5, this means that 50% of the data around the median of 40 is within 10.5 bars of each other.

Before assigning the homework you may want to brainstorm ideas of statistical questions that students in your class would be interested in answering by conducting a survey.

Spiral Review

Mark each statement as true or false. If the statement is false, rewrite it so that it is true.

	Statement	True	False
1.	All trapezoids are quadrilaterals.	x	
2.	All rectangles are squares. All square are rectangles.		x
3.	Kites are never rhombuses. Kites are sometimes rhombuses.		x
4.	Rectangles are always parallelograms.	x	
5.	All squares are kites.	x	

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.4b Homework: Make, Interpret, and Draw Conclusions About Data Part 2

1. Choose the best term from the box to complete each sentence

Mean	Histogram	Range	Interquartile Range (IQR)	Dot Plot	Box Plot

   a. A____Box Plot_____ is a data display that divides a set of data into four parts using the median and quartiles

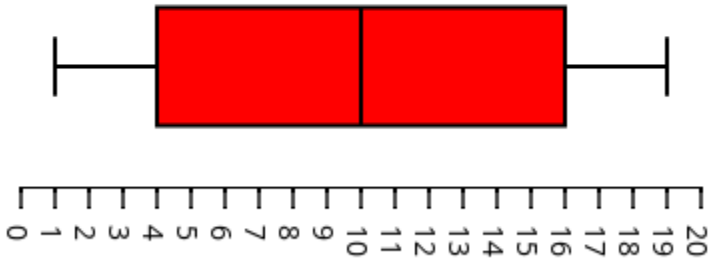
   b. The____Interquartile Range (IQR)_____ is the difference of the 3rd quartile and the 1st quartile.

   c. The_____Mean_____is a measure of center for a data set found by taking the sum of the data divided by the number of data values.

   d. A _____Dot Plot_____use a number line to show the number of times each value in a data set occurs.

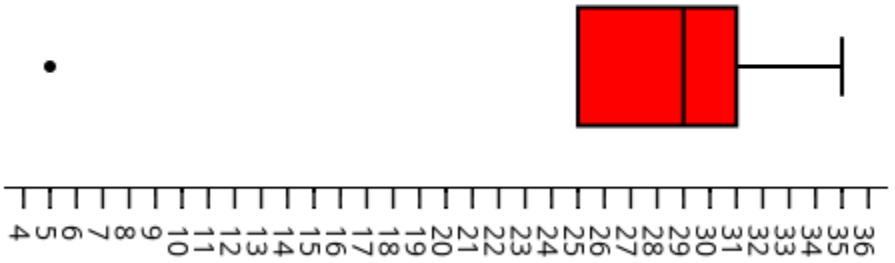
2. Draw a box plot to match the given conditions.
   Answers may vary, sample answers are given.

   a. A box that is symmetrical and has a median of 10.



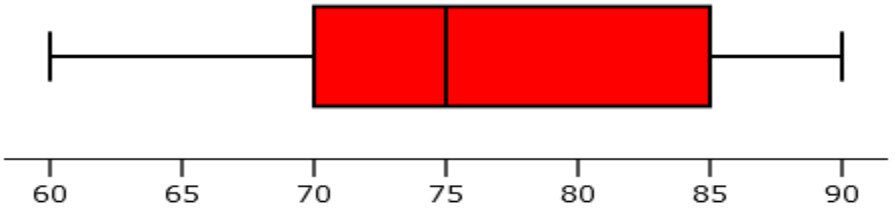
1

   b. A box plot that is skewed left and has an outlier at 5.



2

   c. A box plot that has and IQR of 15 and a range of 30.



3

1 evatipp. "Symmetrical Box med 3" Plotly Technologies Inc., https://plot.ly/518/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Box Plot Skewed Left with an outlier at 5" Plotly Technologies Inc., https://plot.ly/520/~evatipp/ (Last Accessed: March 2016)
3 evatipp. "Box with IQR 15, range 30 (1)" Plotly Technologies Inc., https://plot.ly/522/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

3. Create a statistical question of your own that you can answer by surveying people in your class. Then ask at least 15 people in your class your question and record your data below.

   a. Write your question here:____Answers will vary_____

   b. Record your data in the table below.

Student	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Response															

   c. What is the attribute that is being measured from your question?

   d. What is the unit of measurement for your attribute?

   e. Summarize your data will a graphical display (histogram, dot plot, or box plot).

   f. What is the overall shape of the data?

   g. What is the center of your data?

   h. Discuss the variability or spread of the data.

   i. Use the numerical summaries and graphical display of your data to answer your statistical question.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

4.4c Self Assessment: Section 3.4

Consider the following skills/concepts. Rate your comfort level with each skill/concept by checking the box that best describes your progress in mastering each skill/concept. Corresponding sample problems, referenced in brackets, can be found on the following page.

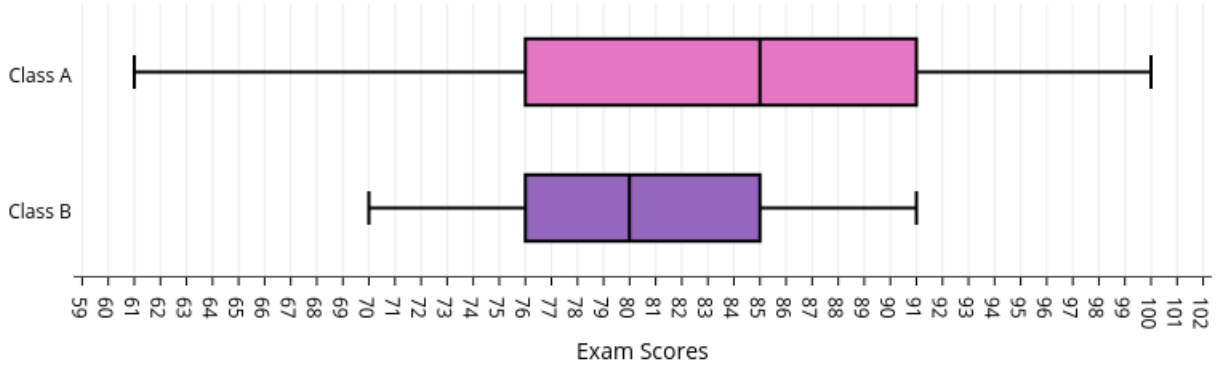
Skill/Concept	Minimal Understanding 1	Partial Understanding 2	Sufficient Mastery 3	Substantial Mastery 4
1. Determine whether a measure of center or a measure of variability should be used to answer a statistical question.	I can determine whether a measure of center should be used to answer a statistical question.	I can determine whether a measure of variability should be used to answer a statistical question.	I can determine whether a measure of center or measure of variability should be used to answer a statistical question.	I can determine whether a measure of center or measure of variability should be used to answer a statistical question and in turn use this measure to answer the statistical question.
2. Interpret, compare, and draw conclusions about data distributions by using statistical models and appropriate measures of shape, center and spread.	I can find some measures of shape, center, and spread but I don't really know how to use them to interpret data.	I know how to make some conclusions about a data distribution using statistical models but I struggle to identify and find all the appropriate measures of shape, center, and spread.	I know how to interpret and draw conclusions about a single data distribution by using statistical models and appropriate measures of shape, center, and spread.	I know how to interpret, compare, and draw conclusions about data distributions by using statistical models and appropriate measures of shape, center, and spread.

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

Sample Problems for Section 4.4

Square brackets indicate which concept/skill the problem aligns to.

1. The data below shows the recent exam scores for two different English classes. For each of the three statistical questions below (Questions A, B, and C), decide if you would answer the question by considering center or considering variability in the data distribution. Then use this measure to answer the statistical question. [1]



[1]

Question A: What is the typical test score for all of the students on this exam?
Measure of center; this question asks about a typical value that would be answered by averaging the measures of center for both classes.
Median for Class A=85, median for Class B=80, overall median =82.5

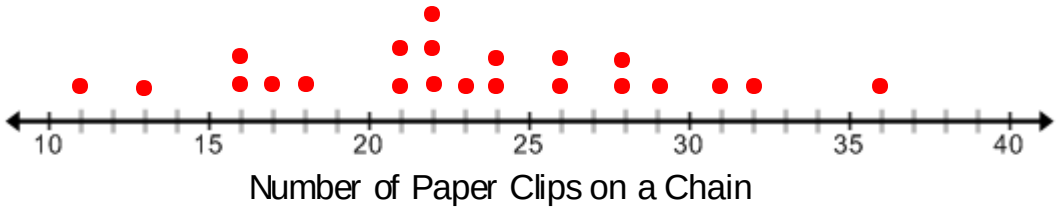
Question B: Which class has the highest median score for this exam?
Measure of center; this question asks which class has a higher average. This would be answered by comparing the centers for each distribution.
Class A has a higher median, it is 85 and Class B's median is 80.

Question C: Which class has scores that are more consistent (more similar) with one another?
Measure of Variability; this question asks about the consistency of scores within a class. It would be answered by comparing the variability of each data distribution.
The IQR for Class B is 9 and the IQR for Class A is 15. This means that the scores for Class B vary less, or are more consistent with one another, than the scores for Class A.

2. A group of 6th graders were each given one minute to make a paper clip chain as long as possible. The number of paper clips on each person's chain is shown below. [2]

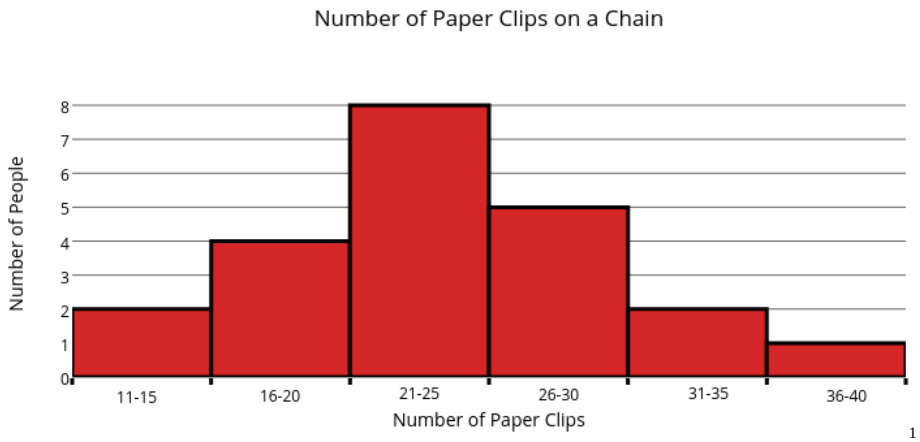
$$11, 13, 16, 16, 17, 18, 21, 21, 22, 22, 22, 23, 24, 24, 26, 26, 28, 28, 29, 31, 32, 36$$

a. Choose an appropriate graphical display for your data. Represent the data in this display and describe the shape of the data.
Students may choose to display the data in a histogram, box plot, or dot plot.

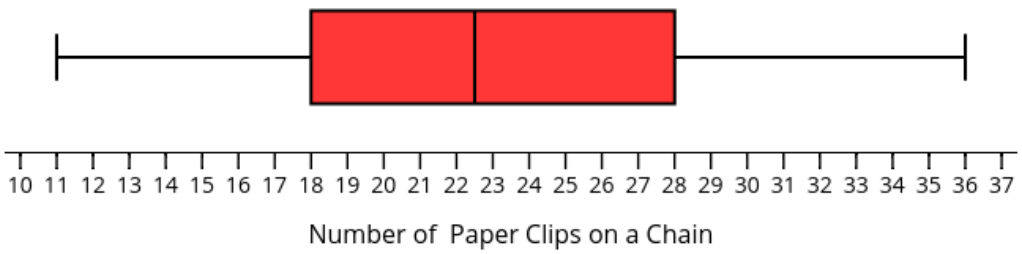


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

1 evatipp. "Comparing Exam Scores (1)" Plotly Technologies Inc., https://plot.ly/552/~evatipp/ (Last Accessed: March 2016)



[1]



[2]

The data appears to be fairly symmetrical with most of the data between 16 and 35 clips. The number of paper clips on a chain peaks between 21-25 clips. There are no major outliers.

b. Choose and find an appropriate measure of center for the data. Explain why you chose this measure of center to represent the data and interpret its meaning given the context.

Since the data is symmetrical and does not appear to have any major outliers students may choose either the mean or median.

The mean is 23; this means that the average number of paper clips in chain is 23. The median is 22.5; this means that half the chains have more than 22.5 clips and half of the chains have less than 22.5 clips.

c. Choose and find an appropriate measure of variability. Explain why you chose this measure of variability to represent the data and interpret its meaning given the context.

Students should base their measure of variability off of their measure of center. If they chose the mean as a measure of center they should use the MAD to describe the variability. If they chose the median as the measure of center then they should use the IQR to describe the variability.

The MAD is about 4.9; this means that each chain has an average deviation of 4.9 clips away from the mean of 23clips.

The IQR is 10; this means that the middle 50% of the data clustered around the median vary by 10 clips.

1 evatipp. "Paper Clip Chain" Plotly Technologies Inc., https://plot.ly/555/~evatipp/ (Last Accessed: March 2016)
2 evatipp. "Paper Clips Box Plot" Plotly Technologies Inc., https://plot.ly/558/~evatipp/ (Last Accessed: March 2016)

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.