CCGPS Frameworks Teacher Edition

Mathematics

6th Grade
Unit 6: Statistics

Dr. John D. Barge, State School Superintendent
"Making Education Work for All Georgians"

Unit 6
Statistics

TABLE OF CONTENTS

OVERVIEW

In this unit students will:

- Analyze data from many different sources such as organized lists, box-plots, bar graphs and stem-and-leaf plots
- Understand that responses to statistical questions may vary
- Understand that data can be described by a single number
- Determine quantitative measures of center (median and/or mean)
- Determine quantitative measures of variability (interquartile range and/or mean absolute deviation)

Although the units in this instructional framework emphasize key standards and big ideas at specific times of the year, routine topics such as estimation, mental computation, and basic computation facts should be addressed on an ongoing basis. Ideas related to the eight process standards should be addressed constantly as well. To assure that this unit is taught with the appropriate emphasis, depth, and rigor, it is important that the tasks listed under "Evidence of Learning" be reviewed early in the planning process. A variety of resources should be utilized to supplement this unit. This unit provides much needed content information, but excellent learning activities as well. The tasks in this unit illustrate the types of learning activities that should be utilized from a variety of sources.

STANDARDS ADDRESSED IN THIS UNIT

Mathematical standards are interwoven and should be addressed throughout the year in as many different units and activities as possible in order to emphasize the natural connections that exist among mathematical topics. Teaching of the CCGPS should embed the eight mathematical practices. The Standards for Mathematical Practice *(MP)* describe varieties of expertise that mathematics educators at all levels should seek to develop in their students.

KEY STANDARDS

Apply and extend previous understandings of measurement and interpreting data.

MCC6.SP.1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.

MCC6.SP.2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

MCC6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

MCC6.SP.4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

MCC6.SP.5. Summarize numerical data sets in relation to their context, such as by:

MCC6.SP.5.a. Reporting the number of observations.

MCC6.SP.5.b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement

MCC6.SP.5.c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

MCC6.SP.5.d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Standards for Mathematical Practice:

The Standards for Mathematical Practice describe varieties of expertise that mathematics educators at all levels should seek to develop in their students. These practices rest on important "processes and proficiencies" with longstanding importance in mathematics education. The first of these are the NCTM process standards of problem solving, reasoning and proof, communication, representation, and connections. The second are the strands of mathematical proficiency specified in the National Research Council's report *Adding It Up*: adaptive reasoning , strategic competence, conceptual understanding (comprehension of mathematical concepts, operations and relations), procedural fluency (skill in carrying out procedures flexibly, accurately, efficiently and appropriately) and productive disposition (habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy).

1. **Make sense of problems and persevere in solving them.** In grade 6, students solve problems involving ratios and rates and discuss how they solved them. Students solve real world problems through the application of algebraic and geometric concepts. Students seek the meaning of a problem and look for efficient ways to represent and solve it. They may check their thinking by asking themselves, "What is the most efficient way to solve the problem?", "Does this make sense?", and "Can I solve the problem in a different way?"

2. **Reason abstractly and quantitatively.** In grade 6, students represent a wide variety of real world contexts through the use of real numbers and variables in mathematical

expressions, equations, and inequalities. Students contextualize to understand the meaning of the number or variable as related to the problem and decontextualize to manipulate symbolic representations by applying properties of operations.

3. **Construct viable arguments and critique the reasoning of others.** In grade 6, students construct arguments using verbal or written explanations accompanied by expressions, equations, inequalities, models, and graphs, tables, and other data displays (i.e. box plots, dot plots, histograms, etc.). They further refine their mathematical communication skills through mathematical discussions in which they critically evaluate their own thinking and the thinking of other students. They pose questions like "How did you get that?", "Why is that true?" "Does that always work?" They explain their thinking to others and respond to others' thinking.

4. **Model with mathematics.** In grade 6, students model problem situations symbolically, graphically, tabularly, and contextually. Students form expressions, equations, or inequalities from real world contexts and connect symbolic and graphical representations. Students begin to explore covariance and represent two quantities simultaneously. Students use number lines to compare numbers and represent inequalities. They use measures of center and variability and data displays (i.e. box plots and histograms) to draw inferences about and make comparisons between data sets. Students need many opportunities to connect and explain the connections between the different representations. They should be able to use all of these representations as appropriate to a problem context.

5. **Use appropriate tools strategically.** Students consider available tools (including estimation and technology) when solving a mathematical problem and decide when certain tools might be helpful. For instance, students in grade 6 may decide to represent similar data sets using dot plots with the same scale to visually compare the center and variability of the data. Additionally, students might use physical objects or applets to construct nets and calculate the surface area of three dimensional figures.

6. **Attend to precision**. In grade 6, students continue to refine their mathematical communication skills by using clear and precise language in their discussions with others and in their own reasoning. Students use appropriate terminology when referring to rates, ratios, geometric figures, data displays, and components of expressions, equations or inequalities.

7. **Look for and make use of structure.** Students routinely seek patterns or structures to model and solve problems. For instance, students recognize patterns that exist in ratio tables recognizing both the additive and multiplicative properties. Students apply properties to generate equivalent expressions (i.e. $6 + 2x = 3(2 + x)$ by distributive property) and solve equations (i.e. $2c + 3 = 15$, $2c = 12$ by subtraction property of equality), $c=6$ by division property of equality). Students

compose and decompose two- and three dimensional figures to solve real world problems involving area and volume.

8. **Look for and express regularity in repeated reasoning.** In grade 6, students use repeated reasoning to understand algorithms and make generalizations about patterns. During multiple opportunities to solve and model problems, they may notice that *a/b* ÷ *c/d* = *ad/bc* and construct other examples and models that confirm their generalization. Students connect place value and their prior work with operations to understand algorithms to fluently divide multi-digit numbers and perform all operations with multi-digit decimals. Students informally begin to make connections between covariance, rates, and representations showing the relationships between quantities.

RELATED STANDARDS

Apply and extend previous understandings of multiplication and division to divide fractions by fractions.

MCC6.NS.1 Interpret and compute quotients of fractions, and solve word problems involving division of fractions by fractions, e.g., by using visual fraction models and equations to represent the problem. For example, create a story context for (2/3) ÷ (3/4) and use a visual fraction model to show the quotient; use the relationship between multiplication and division to explain that (2/3) ÷ (3/4) = 8/9 because ¾ of 8/9 is 2/3. (In general, (a/b) ÷ (c/d) = ad/bc.) How much chocolate will each person get if 3 people share ½ lb. of chocolate equally? How many ¾ - cup servings are in 2/3 of a cup of yogurt? How wide is a rectangular strip of land with length ¾ mi and area ½ square miles?

Compute fluently with multi-digit numbers and find common factors and multiples

MCC6.NS.2 Fluently divide multi-digit numbers using the standard algorithm.

MCC6.NS.3 Fluently add, subtract, multiply, and divide multi-digit decimals using the standard algorithm for each operation.

MCC6.NS.4 Find the greatest common factor of two whole numbers less than or equal to 100 and the least common multiple of two whole numbers less than or equal to 12. Use the distributive property to express a sum of two whole numbers 1-100 with a common factor as a multiple of a sum of two whole numbers with no common factor. For example, express 36 + 8 as 4(9 + 2).

ENDURING UNDERSTANDINGS

- Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers

- 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.
- 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.
- Understand that numerical data can be displayed in plots on a number line, including dot plots, histograms, and box plots.
- Summarize numerical data sets in relation to their context, such as by:
    - Reporting the number of observations.
    - Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
    - 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.
    - 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

ESSENTIAL QUESTIONS

- What is the best way to organize a set of data?
- What kinds of graphs will best represent a given set of data?
- How can I describe the center of a set of data?
- How can I describe the spread of a set of data?
- How can I use data to compare different groups?
- How do I choose and create appropriate graphs to represent data?
- What conclusions can be drawn from data?

CONCEPTS & SKILLS TO MAINTAIN

In order for students to be successful, the following skills and concepts need to be maintained

- Analyzing patterns and seeing relationships
- Fluency with operations on multi-digit numbers and decimals

SELECTED TERMS AND SYMBOLS

The following terms and symbols are often misunderstood. These concepts are not an inclusive list and should not be taught in isolation. However, due to evidence of frequent difficulty and misunderstanding associated with these concepts, instructors should pay particular attention to them and how their students are able to explain and apply them.

The definitions below are for teacher reference only and are not to be memorized by the students. Students should explore these concepts using models and real life examples. Students should understand the concepts involved and be able to recognize and/or demonstrate them with words, models, pictures, or numbers.

The websites below are interactive and include a math glossary suitable for middle school children. **Note – At the middle school level, different sources use different definitions. Please preview any website for alignment to the definitions given in the frameworks.**

http://www.amathsdictionaryforkids.com/
This web site has activities to help students more fully understand and retain new vocabulary

http://intermath.coe.uga.edu/dictnary/homepg.asp
Definitions and activities for these and other terms can be found on the Intermath website. Intermath is geared towards middle and high school students.

- **Box and Whisker Plot**- A diagram that summarizes data using the median, the upper and lowers quartiles, and the extreme values (minimum and maximum). Box and whisker plots are also known as box plots. It is constructed from the five-number summary of the data: Minimum, Q1 (lower quartile), Q2 (median), Q3 (upper quartile), Maximum.

- **Frequency**- the number of times an item, number, or event occurs in a set of data

- **Grouped Frequency Table**- The organization of raw data in table form with classes and frequencies

- **Histogram**- a way of displaying numeric data using horizontal or vertical bars so that the height or length of the bars indicates frequency

- **Inter-Quartile Range (IQR)**- The difference between the first and third quartiles. (Note that the first quartile and third quartiles are sometimes called upper and lower quartiles.)

- **Maximum value**- The largest value in a set of data.

- **Mean Absolute Deviation**- the average distance of each data value from the mean. The MAD is a gauge of "on average" how different the data values are form the mean value.

- **Mean**- The "average" or "fair share" value for the data. The mean is also the balance point of the corresponding data distribution.

$$arithmetic\ mean = \bar{x} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n}$$

- **Measures of Center**- The mean and the median are both ways to measure the center for a set of data.

- **Measures of Spread**- The range and the Mean Absolute Deviation are both common ways to measure the spread for a set of data.

- **Median**- The value for which half the numbers are larger and half are smaller. If there are two middle numbers, the median is the arithmetic mean of the two middle numbers. Note: The median is a good choice to represent the center of a distribution when the distribution is skewed or outliers are present.

- **Minimum value**- The smallest value in a set of data.

- **Mode**- The number that occurs the most often in a list. There can be more than one mode, or no mode.

- **Outlier**- A value that is very far away from most of the values in a data set.

- **Range**- A measure of spread for a set of data. To find the range, subtract the smallest value from the largest value in a set of data.

- **Stem and Leaf Plot**- A graphical method used to represent ordered numerical data. Once the data are ordered, the stem and leaves are determined. Typically the stem is all but the last digit of each data point and the leaf is that last digit.

CLASSROOM ROUTINES

The importance of continuing the established classroom routines cannot be overstated. Daily routines must include such obvious activities as estimating, analyzing data, describing patterns, and answering daily questions. They should also include less obvious routines, such as how to select materials, how to use materials in a productive manner, how to put materials away, how to access classroom technology such as computers and calculators. An additional routine is to allow plenty of time for children to explore new materials before attempting any directed activity with these new materials. The regular use of routines is important to the development of students' number sense, flexibility, fluency, collaborative skills and communication. These routines contribute to a rich, hands-on standards based classroom and will support students' performances on the tasks in this unit and throughout the school year.

STRATEGIES FOR TEACHING AND LEARNING

The purpose of this unit is to begin the study of statistics, beginning with examples of one variable (numerical) data sets, and evaluating the mean, median, and mode of these data sets. These are all ways of describing a data set numerically. Students should become aware of different methods of organizing data, beginning with stem and leaf plots. Follow stem and leaf plots with frequency tables, which will then lead into histograms. Guide students to see the similarity between the shape of the histogram and the shape of the stem and leaf plot if you turn the stem plot sideways for the same data set. Teachers should help set the frequency table intervals (histogram intervals) so that a skyscraper or pancake effect does not occur. Encourage students to mark the mean and the median on the histograms so that they understand that both of these values describe the center of the data. Students should begin to use statistical language, such a "distribution" to describe the histogram of the data. Students should also get into the habit of describing the shape of the distribution as "single peaked, double peaked, roughly symmetric, or skewed."

Next, students will find the range and Mean Absolute Deviation (MAD) of the data sets. This will require a strong conceptual foundation as to what these terms represent and the use of mathematical operations, such as addition and division. Teachers should use whole numbers to help students evaluate the MAD. The range and the MAD help describe the spread of the data, which describes data sets in more detail than just using measures of center. Using the task for MAD will help the students understand the importance of knowing this measure of spread, instead of only using the mean or median when describing a data set.

Students will also learn about box plots (often called box and whisker plots). Box plots are a visual display of the 5 Number Summary: minimum, lower quartile Q1, median Q2, upper quartile Q3, and maximum. Side-by-side box plots can be used on the same scale to compare and contrast two sets of data. It should be emphasized that the "box" holds the middle 50% of the data, known as the Inter Quartile Range (IQR). Students should recognize "outliers" on the box plot by using the $1.5 \times IQR$ rule, which states that if a value from the data set lies 1.5 x IQR above or below the median, it is considered an outlier. Students should also recognize that outliers affect the mean of the data set, but not the mode.

The following strategies should be used to help students be successful in this unit.

- Students should be actively engaged by developing their own understanding.
- Mathematics should be represented in as many ways as possible by using graphs, tables, pictures, symbols and words.

- Interdisciplinary and cross curricular strategies should be used to reinforce and extend the learning activities.
- Appropriate manipulatives and technology should be used to enhance student learning.
- Students should be given opportunities to revise their work based on teacher feedback, peer feedback, and metacognition which includes self-assessment and reflection.
- Students should write about the mathematical ideas and concepts they are learning.
- Consideration of all students should be made during the planning and instruction of this unit. Teachers need to consider the following:
    - What level of support do my struggling students need in order to be successful with this unit?
    - In what way can I deepen the understanding of those students who are competent in this unit?
    - What real life connections can I make that will help my students utilize the skills practiced in this unit?
- Graphic organizers should be used to help students arrange ideas and tie mathematical thinking to situations involving real-world situations

INSTRUCTIONAL RESOURCES AND TOOLS

- Newspaper and magazine graphs for analysis of the spread, shape and variation of data

- Hollywood Box Office This rich problem focuses on measures of center and graphical displays.

- Wet Heads In this lesson, students create stem-and-leaf plots and back-to-back stem-and-leaf plots to display data collected from an investigative activity.

- Stella's Stumpers Basketball Team Weight This problem situation uses the mean to determine a missing data element.

- Learning Conductor Lessons. Use the interactive applets in these standards-based lessons to improve understanding of mathematical concepts. Scroll down to the **statistics** section for your specific need.

- From the National Council of Teachers of Mathematics, Illuminations: Height of Students in our Class. This lesson has students creating box-and-whisker plots with an extension of finding measures of center and creating a stem-and-leaf plot.

- National Library of Virtual Manipulatives. Students can use the appropriate applet from this page of virtual manipulatives to create graphical displays of the data set. This provides an important visual display of the data without the tediousness of the student hand drawing the display.

- Kader, Gary D. "Means and MADs." *Mathematics Teaching in the Middle School* 4.6 (1999): 398-403. Print.

- Franklin, C., Kader, G., Mewborne, D., Moreno, J., Peck, R., Perry, M., Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (gaise) report: A pre-k- 12 curriculum framework.* Alexandria, Virginia: American Statistical Association. Print.

EVIDENCE OF LEARNING

By the conclusion of this unit, students should be able to demonstrate the following competencies:

- Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.
- 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.
- Understand 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.
- Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- Summarize numerical data sets in relation to their context, such as by:
    - Reporting the number of observations.
    - Describing the nature of the attribute under investigation, including how it was measured and its units of measurement
    - 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.

TASKS:

The following tasks represent the level of depth, rigor, and complexity expected of all sixth grade students. These tasks or tasks of similar depth and rigor should be used to demonstrate evidence of learning. It is important that all elements of a task be addressed throughout the learning process so that students understand what is expected of them. While some tasks are identified as a performance task, they may also be used for teaching and learning (learning task).

Task Name	Task Type *Grouping Strategy*	Content Addressed
Who was the Greatest Yankee Home Run Hitter?	Learning / Scaffolding Task *Individual/Partner*	Mean, Median, Mode, Stem and Leaf Charts, Frequency lists, Histograms, Range
How Long is a Minute?	Learning / Scaffolding Task *Individual/Partner*	Box Plots, Inter Quartile Range (IQR), Minimum, Maximum, 5 number summary
Where's Waldo?	Learning / Scaffolding Task *Individual/Partner* Differentiation Option	Side by Side Box Plots; A deeper look into percentages that are present in box plots, discussion of outliers. Differentiation option at end of task
How Many People are in Your family?	Learning / Scaffolding Task *Class Activity*	Mean, Mean Absolute Deviation, Mode, Range
Culminating Task: Order Up! Fast Food Frenzy Part I: Burgers~ I'm Lovin' It! Part II: Biggie Size those Fries~ or Maybe Not! Part III: I'm Thirsty!	Performance Task *Individual Task*	Part I: Measures of Center, measures of spread, overall shape, outliers Part II: Side by Side Box Plots, Spread, Shape, Variability, MAD Part III: Histograms, measures of spread, graphing calculator extension

Task: Who Was the Greatest Yankee Home Run Hitter?

ESSENTIAL QUESTIONS

- What is meant by the center of a data set, how is it found and how is it useful when analyzing data?
- How do I choose and create appropriate graphs to represent data?
- How can I describe the spread of a set of data?
- How can I use data to compare different groups?
- What conclusions can be drawn from data?

STANDARDS ADDRESSED:

MCC6.SP.2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
MCC6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
MCC6.SP.4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
MCC6.SP.5. Summarize numerical data sets in relation to their context, such as by:
MCC6.SP.5.a. Reporting the number of observations.
MCC6.SP.5.b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement
MCC6.SP.5.c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
MCC6.SP.5.d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.
The following table lists four of the greatest New York Yankees' home run hitters with the number of homeruns each hit while a Yankee.

Adapted from : James M. Landwehr and Ann E. Watkins, Dale Seymour Publications, Mathematics, 1986, Pg. 160

Babe Ruth			Lou Gehrig			Mickey Mantle			Roger Maris	
Year	Home runs		Year	Home runs		Year	Home runs		Year	Home runs
1920	54		1923	1		1951	13		1960	39
1921	59		1924	0		1952	23		1961	61
1922	35		1925	20		1953	21		1962	33
1923	41		1926	16		1954	27		1963	23
1924	46		1927	47		1955	37		1964	26
1925	25		1928	27		1956	52		1965	8
1926	47		1929	35		1957	34		1966	13
1927	60		1930	41		1958	42			
1928	54		1931	46		1959	31			
1929	46		1932	34		1960	40			
1930	49		1933	32		1961	54			
1931	46		1934	49		1962	30			
1932	41		1935	30		1963	15			
1933	34		1936	49		1964	35			
1934	22		1937	37		1965	19			
			1938	29		1966	23			
			1939	0		1967	22			
						1968	18			

Source: Macmillan Baseball Encyclopedia, 4th edition

1. Find the mean, median, and mode, and number of observations for each player.

	Ruth		***Gherig***		***Mantle***		***Maris***
Mean	***43.9***		***29***		***29.7***		***29***
Median	***46***		***32***		***28.5***		***26***
Mode	***46***		***0, 49***		***23***		***no mode***
n	***15***		***17***		***18***		***7***

Of the three values you computed for each player, which do you think best describes the performance of each player? Why?

The median is not affected by extreme high or low values, so the median would best describe the performance of each player.

2. Make a stem and leaf plot for each player.

Comment

Shown below are ordered stem and leaf plots for each player. An ordered stem and leaf plot makes finding the mean and quartiles much easier. Having the data ordered also makes it easier to identify the mode of the data.

It is important to emphasize several points about these plots.

- ***Make sure students include a key with the plots.***
- ***Two of the players, Gehrig and Maris, had years in which their home run totals were single digits. Be sure to point out how single digits are handled in a plot where everything else is a two-digit number.***
- ***With the Maris plot, be sure students know that although there are no data values in the 40's or 50's, the 4 and 5 MUST be included in the plot. Part of the purpose of a stem and leaf plot is to be able to see the distribution of the data and to compare it to a histogram. If the 4 and 5 are not included, then the stem and leaf plot does not have the same shape as the corresponding histogram, and would be an incorrect representation of the data.***

Solutions

a.

Babe Ruth 2 | 5 represents 25 home runs

stem	leaf
2	2 5
3	4 5
4	1 1 6 6 6 7 9
5	4 4 9
6	0

Lou Gehrig 1| 6 represents 16 home runs

stem	leaf
0	0 0 1
1	6
2	0 7 9
3	0 2 4 5 7
4	1 6 7 9 9

Mickey Mantle 1 | 3 represents 13 home runs

stem	leaf
1	3 5 8 9
2	1 2 3 3 7
3	0 1 4 5 7
4	0 2
5	2 4

Roger Maris   1 | 3 represents 13 home runs

stem	***leaf***
0	***8***
1	***3***
2	***3 6***
3	***3 9***
4	
5	
6	***1***

From your graphical display, discuss which player appears to be the greatest home run hitter? Why did you choose this person?

Comment

Some students may choose Roger Maris because he had the highest single season total. Others may choose Babe Ruth because he hit home runs more consistently than Maris. Some students may choose Mickey Mantle because his career lasted the longest. Some may choose Lou Gehrig for a similar reason (longevity) even though he had years where he did not hit any home runs. (If students are familiar with Lou Gehrig's disease, they may think he performed well in spite of his disease.)

3. Create a frequency table for Babe Ruth's home runs. Use the stem and leaf plot of data to help you organize your frequency table.

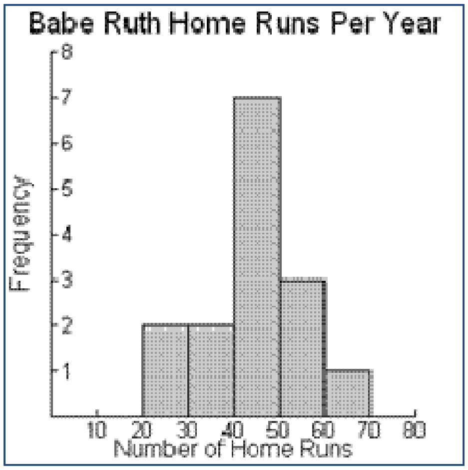
Students should be sure that the frequency table intervals match the intervals on the stem and leaf plots. This is also a good time to remind students that intervals cannot overlap.

Babe Ruth

home runs	***frequency***
20-29	***2***
30-39	***2***
40-49	***7***
50-59	***3***
60-69	***1***

4. Using the frequency table above, create a histogram of Babe Ruth's home runs.

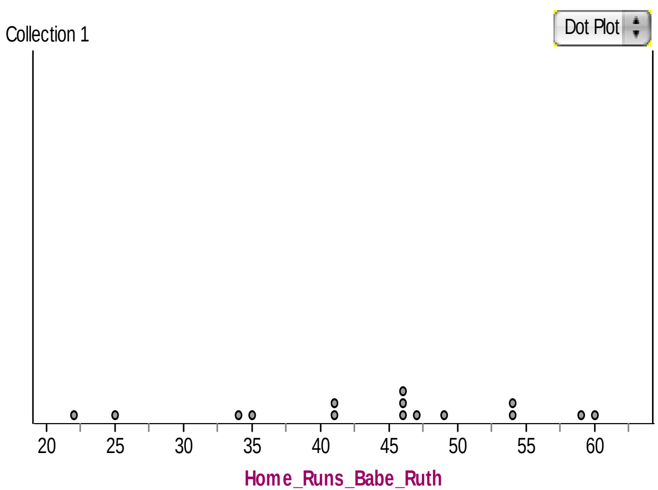
Once students have created the frequency table for Babe Ruth, creating the histogram should not take long. The data have already been organized and students have to put the information from the frequency table into graphical form.



5. Looking at the histogram, determine the range of homeruns Babe Ruth hit while playing for the Yankees.

60-22=38 homerun range

6. Create a dot plot using Babe Ruth's home runs.



7. Describe the similarities between Babe Ruth's histogram and Babe Ruth's dot plot. Are there any differences? Does the range change between the histogram and the dot plot?

The majority of the data is centered between 40-50. The range of data for both graphs is between 22 to 60.

Task: Who Was the Greatest Yankee Home Run Hitter?

The following table lists four of the greatest New York Yankees' home run hitters with the number of homeruns each hit while a Yankee.

Adapted from : James M. Landwehr and Ann E. Watkins, Dale Seymour Publications, Mathematics, 1986, Pg. 160

Babe Ruth		Lou Gehrig		Mickey Mantle		Roger Maris	
Year	Home runs	Year	Home runs	Year	Home runs	Year	Home runs
1920	54	1923	1	1951	13	1960	39
1921	59	1924	0	1952	23	1961	61
1922	35	1925	20	1953	21	1962	33
1923	41	1926	16	1954	27	1963	23
1924	46	1927	47	1955	37	1964	26
1925	25	1928	27	1956	52	1965	8
1926	47	1929	35	1957	34	1966	13
1927	60	1930	41	1958	42		
1928	54	1931	46	1959	31		
1929	46	1932	34	1960	40		
1930	49	1933	32	1961	54		
1931	46	1934	49	1962	30		
1932	41	1935	30	1963	15		
1933	34	1936	49	1964	35		
1934	22	1937	37	1965	19		
		1938	29	1966	23		
		1939	0	1967	22		
				1968	18		

Source: Macmillan Baseball Encyclopedia, 4th edition

1. Find the mean, median, and mode, and number of observations for each player.

	Ruth	***Gherig***	***Mantle***	***Maris***
Mean				
Median				
Mode				
n				

Of the three values you computed for each player, which do you think best describes the performance of each player? Why?

2. Make a stem and leaf plot for each player.

Babe Ruth

stem	*leaf*

Lou Gehrig

stem	*leaf*

Mickey Mantle

stem	*leaf*

Roger Maris

stem	*leaf*

From your graphical display, discuss which player appears to be the greatest home run hitter? Why did you choose this person?

3. Create a frequency table for Babe Ruth's home runs. Use the stem and leaf plot of data to help you organize your frequency table.

Babe Ruth

home runs	***frequency***
20-29	
30-39	
40-49	
50-59	
60-69	

4. Using the frequency table above, create a histogram of Babe Ruth's home runs.

5. Looking at the histogram, determine the range of homeruns Babe Ruth hit while playing for the Yankees.

6. Create a dot plot using Babe Ruth's home runs.

7. Describe the similarities between Babe Ruth's histogram and Babe Ruth's dot plot. Are there any differences? Does the range change between the histogram and the dot plot?

Task: How Long is a Minute?

ESSENTIAL QUESTIONS

- What is meant by the center of a data set, how is it found and how is it useful when analyzing data?
- How do I choose and create appropriate graphs to represent data?
- How can I describe the spread of a set of data?
- What conclusions can be drawn from data?

Materials Needed: Stop watch, white paper, string or yarn, graphing calculator

STANDARDS ADDRESSED:

MCC6.SP.2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
MCC6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
MCC6.SP.4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
MCC6.SP.5. Summarize numerical data sets in relation to their context, such as by:
MCC6.SP.5.c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
MCC6.SP.5.d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Learning about Box Plots Kinesthetically

1. Do you think you can determine how long a minute without looking at a clock? With your partner and your stop watch, you will each attempt to determine when you think a minute has passed without looking at a watch or clock.
Comment: Each group of 2 students will need a stop watch. Many have this on their cell phones if you are willing to allow them the use of their cell phones for this activity. Science teachers may also be able to provide you with stop watches.

2. You and your partner should sit as far as possible from other students in the class. Make sure you cannot see a clock from your position. Remove your watches and hide them! COUNT in seconds (do not convert to minutes and seconds- keep all answers in seconds only).
Comment: Be sure that no watches or cell phones are out. Students can move around and sit in the floor. Be sure that students stay quiet during this lab. The "timer" is the student with

the stop watch (it is okay if they are facing the clock), and the "guesser" is the student who is attempting to guess when a minute has passed.

3. You will take turns measuring and timing. The timer tells the measurer when to begin. When the measurer believes a minute has passed, he should say, "stop" quietly. At that point, the timer should record the time that has passed to the nearest second. Do not tell your partner how much time actually passed! The "timer" needs to record the "guessers" time down on a sheet of paper. Now pass the watch and switch roles.
Comment: Be sure to record each time down on this paper so you do not forget it! Students will re-write these times down on a blank sheet of paper LARGE enough for the class to see in the next step.

4. Once each person in your group has gone, you can share times with your partner. Now, write **your own** time down (in seconds) on a sheet of paper large enough so that everyone can see it (one sheet per student).
5. Stand in line from smallest to largest, and hold the pages in front of you, so that everyone can see them.
Comment: Have students line up from least to greatest. At this point, you may choose to have your students tape their papers to the board, and then sit down in their desks for the remainder of the lesson.

6. Find the median of the data by counting to the middle. Your teacher or a student will put a sticky note where the median is. The median is also known as Quartile 2 (Q2)

7. Find the median of the smaller set of numbers. Have a student put a sticky note on this place. This is Quartile 1 (Q1)~the median of the smaller half of data.
Comment: If Q1 is in between two values, have both the students that are holding the two values it falls between hold the sticky note for Q1.

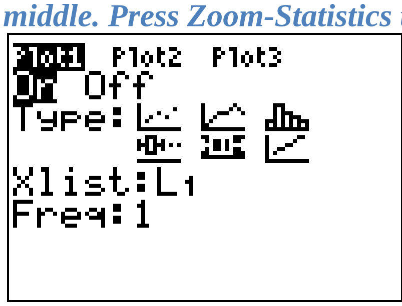
8. Find the median of the larger set of numbers. Have a student put a sticky note on this place. This is Quartile 3 (Q3)~ the median of the larger half of data.
Comment: If Q3is in between two values, have both the students that are holding the two values it falls between hold the sticky note for Q3.

9. Put a string around the people from Q1-Q3. This is the "box" of the box plot. 50% of the data set lies within this area.
Comment: The string can also be taped onto the board around Q1 and Q3.

10.Then, put strings from the ends of the box to the minimum and maximum. These are the "whiskers" of the box and whisker plot. 25% of the data lies in each whisker.

11. Collect each individual persons data from the room. Draw a box plot of the classes data that is the same box plot that you made with your bodies.

12. Input the data in List 1 in the Graphing Calculator. This is the same box plot you have drawn and created with your bodies.
Comment: This is done in the Graphing Calculator (TI-84) by inputting the data in L1 under the Stat Menu. Then go to 2nd-StatPlot-Plot 1. Choose the Box Plot in the bottom row in the middle. Press Zoom-Statistics to show the box plot graph.



Task: How Long is a Minute?

Learning about Box Plots Kinesthetically

Materials Needed: Stop watch, white paper, string or yarn, graphing calculator

1. Do you think you can determine how long a minute without looking at a clock? With your partner and your stop watch, you will each attempt to determine when you think a minute has passed without looking at a watch or clock.

2. You and your partner should sit as far as possible from other students in the class. Make sure you cannot see a clock from your position. Remove your watches and hide them! COUNT in seconds (do not convert to minutes and seconds- keep all answers in seconds only).

3. You will take turns measuring and timing. The timer tells the measurer when to begin. When the measurer believes a minute has passed, he should say, “stop” quietly. At that point, the timer should record the time that has passed to the nearest second. Do not tell your partner how much time actually passed! The “timer” needs to record the “guessers” time down on a sheet of paper. Now pass the watch and switch roles.

4. Once each person in your group has gone, you can share times with your partner. Now, write **your own** time down (in seconds) on a sheet of paper large enough so that everyone can see it (one sheet per student).

5. Stand in line from smallest to largest, and hold the pages in front of you, so that everyone can see them.

6. Find the median of the data by counting to the middle. Your teacher or a student will put a sticky note where the median is. The median is also known as Quartile 2 (Q2)

7. Find the median of the smaller set of numbers. Have a student put a sticky note on this place. This is Quartile 1 (Q1)~the median of the smaller half of data.

8. Find the median of the larger set of numbers. Have a student put a sticky note on this place. This is Quartile 3 (Q3)~ the median of the larger half of data.

9. Put a string around the people from Q1-Q3. This is the “box” of the box plot. 50% of the data set lies within this area.

10. Then, put strings from the ends of the box to the minimum and maximum. These are the “whiskers” of the box and whisker plot. 25% of the data lies in each whisker.

11. Collect each individual persons data from the room. Draw a box plot of the classes data that is the same box plot that you made with your bodies.

12. Input the data in List 1 in the Graphing Calculator. This is the same box plot you have drawn and created with your bodies.

Task: Where's Waldo?

ESSENTIAL QUESTIONS

- What is the best way to organize a set of data?
- What kinds of graphs will best represent a given set of data?
- How can I describe the center of a set of data?
- How can I describe the spread of a set of data?
- How can I use data to compare different groups?
- How do I choose and create appropriate graphs to represent data?
- What conclusions can be drawn from data?

STANDARDS ADDRESSED:

MCC6.SP.1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.
MCC6.SP.2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
MCC6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
MCC6.SP.4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
MCC6.SP.5. Summarize numerical data sets in relation to their context, such as by:
MCC6.SP.5.a. Reporting the number of observations.
MCC6.SP.5.b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement
MCC6.SP.5.c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
MCC6.SP.5.d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Materials Needed: Coin, 1 Waldo picture per group, stop watch/timer

Part 1: Describing and Comparing Side-By-Side Box Plots

Do male and female brains operate at different speeds? Collect data using the steps below to answer this question.

1. Pair up in groups of 2. Boys should be paired with boys, and girls paired with girls.
Comment: If genders cannot be matched up appropriately, that is okay!

2. Flip a coin to see who is going to be Student 1 (heads) and Student 2 (tails). Heads wins.
Comment: Use dice or coin or the random integer generator in the TI-30 calculator or the TI-84 graphing calculator. On the smaller calculators, press PRB, then randint(1,10), then enter. This will give you a random number between 1 and 10. Press enter again for another number. Whoever gets the "highest number" becomes Student 1.

3. The teachers will place a piece of paper on the desk of each pair of students, picture side down. Students are not to touch the paper or turn it over until instructed to do so by the teacher. The picture is a "Where's Waldo" photo.
Comment: See accompanied Waldo photo. You may choose to find your own Waldo image online to print in color for a class set.

4. The teacher will show you a picture of Waldo on the board so you know what to look for.
Comment: Many students have never heard of Where's Waldo. So, show them a picture of Waldo from the internet in color on your board if possible!

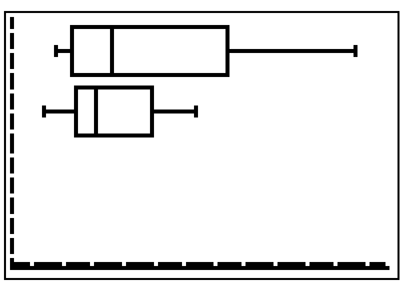
5. When the teachers says, "GO!", Student 1 in each pair is the turn the paper over and find Waldo as quickly as possible. Student 2 is to use a watch/timer on cell phone/clock to time how long it takes for Student 1 to find and point to Waldo. Record your time to the nearest second.
Comment: Watch out for "false finds". They MUST find Waldo before they say they found him!

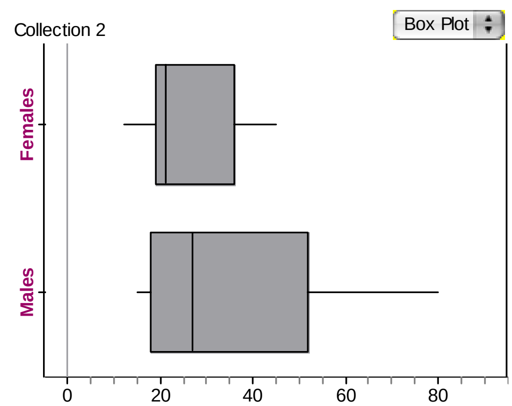
6. Collect data from the class on the board, separating data between boys and girls.
Comment: Have the students organize how to collect this data on the board. Allow them to come to the board to write answers if you choose.

7. Each student must construct two box plots using the steps below to help you. The graph of the 2 box plots must follow. Use graph paper.

	Boys	Girls
a. Put the observations in least to greatest order.		
b. Identify the least number of seconds.		
c. Identify the most number of seconds.		
d. Find the median (Q2).		
e. Find the lower quartile (Q1).		
f. Find the upper quartile (Q3).		

g. Create a number line at the bottom of the graph paper that includes all of the above values in steps (a-f) above. Then, draw your box plots on the same number line together on the graph paper.

Sample Answer: 



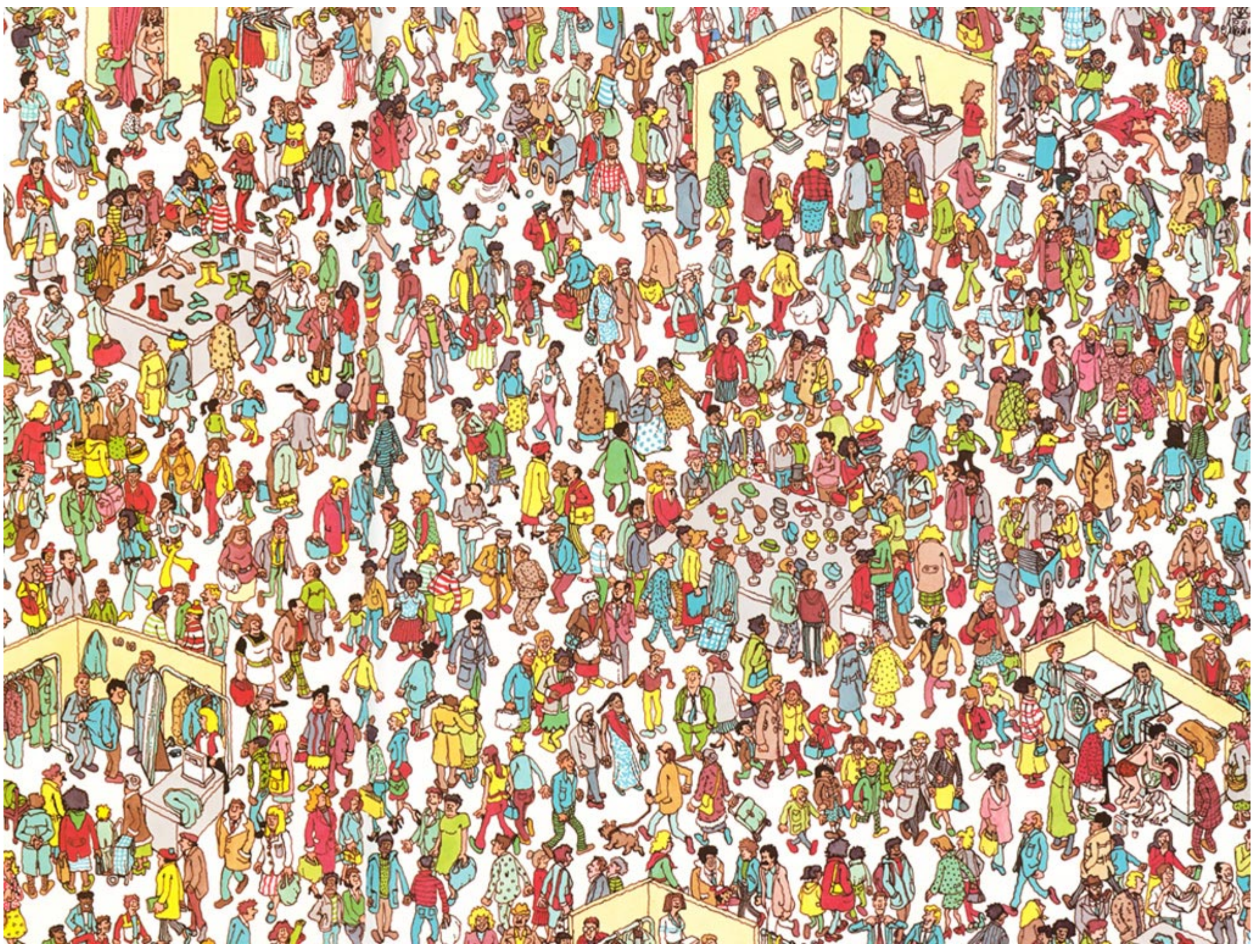
8. Describe the spread of the data for the boys box plot and the girl box plot. Which gender seems to have the largest range?
Sample answer: They may say that the spread of the boys is larger than the spread of the girls. Boys have the largest range.

9. Which gender, overall, found Waldo the quickest? How did you determine this?
Sample Answer: Girls were overall quicker in this sample. The "whiskers" do not reach as far out, and the "box" (the middle 50%) is much smaller and lower on the x-axis.

10. Write a conclusion comparing the speed of boys brain function vs the speed of girls brain function. Use numerical data, such as mean, median, and range to support your answer.

Sample answer: Girls brain function is quicker. The mean boys time is 34.9 with a median of 27. The mean time for girls is 26.7 with a mean of 23.5. The range of boys time was 34 , while the range for girls times was only 17. Therefore, girls tend to have faster brain function than boys, in the data collected for this experiment.

11. Do you think if we did the same experiment with 30 other random people, we would come up with the same conclusion? Why or why not?
Comment: Teachers may want to discuss what would happen if they collect a whole different set of data from a different set of students. Do they feel that it will yield the same conclusions? Why or why not?



Part 2: A Closer Look at Box Plots
Below is a data set of the length of times, in seconds, that it took for nine boys to find Waldo:
7 8 8 10 11 12 13 13 29

12. Find the median of the box plot and circle it.
7 8 8 10 11 12 13 13 29

a. About what percent of the values in a data set are below the median? How do you know this?
Approximately 50 %. This is because the median is the middle of the data set, so the median splits the data set in half.
b. About what percent of values in a data set are above the median? Why?
Approximately 50%. Because ½ of all values lie above the median.

13. Find the Upper (Q3) and Lower (Q1) Quartiles of the box plot. Draw a small vertical line where Q1 and Q3 are on the list of numbers above.
Comment: Remember to find the value of Q1 and Q3 in this data set, you will need to find the mean of the two middle values in Q1 and Q3.

7 8 8 10 11 12 13 13 29

The value of Q1 is 8.
The value of Q3 is 13.

a. About what percent of the data distribution are in each quartile?
Approximately 25%

b. About what percent of the values fall ABOVE the lower quartile?
75%

c. About what percent of the values fall below the upper quartile?
75%

d. The Inter-Quartile Range (IQR) is the "box" of the box plot. In other words, the IQR contains all data between Q1 and Q3. What percent of the data fall between the upper quartile and the lower quartile?
50%. It is the middle 50% of the data set.

e. Find the IQR using the values given above.
13-8=5

f. Why is the IQR important when using it to describe the data?
It means that 50% of our data set is between 8 and 13. So, the middle 50% of the data is within 5 seconds of the median of 11.

g. Were there any outliers? Justify your answer
29 is an outlier. It lies very far outside the pattern for the rest of the data set.

h. Why is it useful to know outliers of a set of data?

Because outliers are prone to have a substantial effect on the mean of the data set, it is good to know if any outliers are present. Knowing that outliers are present might encourage the use of the median over the mean, since the median is not affected by the existence of outliers.

i. Find the mean of the data set. Then, remove the outlier and find the new mean. How does the outlier affect the mean?

$\boldsymbol{Original\ Mean = \frac{111}{9} = 12.33\ seconds\ \ New\ Mean = \frac{82}{8} = 10.25\ seconds}$

The new mean once the outlier has been removed, dropped by over 2 seconds. The original mean is higher because of the "pull" that extreme numbers can have on the mean.

j. What is the median of the data set? Remove the outlier and find the new median. How does the outlier affect the median?

$\boldsymbol{Original\ Median = 11\ seconds\quad New\ Median = 10.5\ seconds}$

The new median is very similar to the original median. Medians are the "middle" of the data set, so no matter how extreme an outlier might be, the actual value of the outlier does not affect the median. The median is said to be "resistant" to outliers.

Part 3: Grades

Suppose your math grade for this grading period is to be determined using 10 test, homework, and project scores. All of the scores are equally important. You get to decide which measure of central tendency, the mode, median, or mean, will be used for your grade.

3. Would you ever prefer to use the median rather than the mean? If so, what would have to be true about the scores? If not, explain why you think using the median wouldn't ever help your grade.

Possible Solution
The median could be better if there were several low outliers.

2. Would you ever prefer to use the mean rather than the median? If so, what would have to be true about the scores? If not, explain why you think using the mean wouldn't ever help your grade.

Possible Solution
The mean could be better if there were several high outliers

3. Is it possible that you could prefer the mode rather than the median or mean? If so, what would have to be true about the scores? If not, explain why you think using the mode wouldn't ever help your grade.

Possible Solution
The mode could be better if it happened to be a high score. The mode may not be a good representative if a student who made 100's on the first two activities/exams and failed

everything else with different numeric grades, would get a 100 in the class due to the mode of his/her grades. In this case, this center would not seem fair.

Task: Where's Waldo?

Part 1: Describing and Comparing Side-By-Side Box Plots

Do male and female brains operate at different speeds? Collect data using the steps below to answer this question.

1. Pair up in groups of 2. Boys should be paired with boys, and girls paired with girls.

2. Flip a coin to see who is going to be Student 1 (heads) and Student 2 (tails). Heads wins.

3. The teachers will place a piece of paper on the desk of each pair of students, picture side down. Students are not to touch the paper or turn it over until instructed to do so by the teacher. The picture is a "Where's Waldo" photo.

4. The teacher will show you a picture of Waldo on the board so you know what to look for.

5. When the teachers says, "GO!", Student 1 in each pair is to turn the paper over and find Waldo as quickly as possible. Student 2 is to use a watch/timer on cell phone/clock to time how long it takes for Student 1 to find and point to Waldo. Record your time to the nearest second.

6. Collect data from the class on the board, separating data between boys and girls.

7. Each student must construct two box plots using the steps below to help you. The graph of the 2 box plots must follow. Use graph paper

	Boys	Girls
a. Put the observations in least to greatest order.		
b. Identify the least number of seconds.		
c. Identify the most number of seconds.		
d. Find the median (Q2).		
e. Find the lower quartile (Q1).		
f. Find the upper quartile (Q3).		

g. Create a number line at the bottom of the graph paper that includes all of the above values in steps (a-f) above. Then, draw your box plots on the same number line together on the graph paper.

8. Describe the spread of the data for the boys box plot and the girl box plot. Which gender seems to have the largest range?

9. Which gender, overall, found Waldo the quickest? How did you determine this?

10. Write a conclusion comparing the speed of boys brain function vs the speed of girls brain function. Use numerical data, such as mean, median, and range to support your answer.

11. Do you think if we did the same experiment with 30 other random people, we would come up with the same conclusion? Why or why not?

Part 2: A Closer Look at Box Plots

Below is a data set of the length of times, in seconds, that it took for nine boys to find Waldo:
7 8 8 10 11 12 13 13 29

12. Find the median of the box plot and circle it.

a. About what percent of the values in a data set are below the median? How do you know this?

b. About what percent of values in a data set are above the median? Why?

13. Find the Upper (Q3) and Lower (Q1) Quartiles of the box plot. Draw a small vertical line where Q1 and Q3 are on the list of numbers above.

a. About what percent of the data distribution are in each quartile?

b. About what percent of the values fall ABOVE the lower quartile?

c. About what percent of the values fall below the upper quartile?

d. The Inter-Quartile Range (IQR) is the “box” of the box plot. In other words, the IQR contains all data between Q1 and Q3. What percent of the data fall between the upper quartile and the lower quartile?

e. Find the IQR using the values given above.

f. Why is the IQR important when using it to describe the data?

g. Were there any outliers? Justify your answer

h. Why is it useful to know outliers of a set of data?

i. Find the mean of the data set. Then, remove the outlier and find the new mean. How does the outlier affect the mean?

j. What is the median of the data set? Remove the outlier and find the new median. How does the outlier affect the median?

Part 3: Grades

Suppose your math grade for this grading period is to be determined using 10 test, homework, and project scores. All of the scores are equally important. You get to decide which measure of central tendency, the mode, median, or mean, will be used for your grade.

1. Would you ever prefer to use the median rather than the mean? If so, what would have to be true about the scores? If not, explain why you think using the median wouldn't ever help your grade.

2. Would you ever prefer to use the mean rather than the median? If so, what would have to be true about the scores? If not, explain why you think using the mean wouldn't ever help your grade.

3. Is it possible that you could prefer the mode rather than the median or mean? If so, what would have to be true about the scores? If not, explain why you think using the mode wouldn't ever help your grade.

Task: How Many People Are in Your Family?

ESSENTIAL QUESTIONS

- What is meant by the center of a data set, how is it found and how is it useful when analyzing data?
- How do I choose and create appropriate graphs to represent data?
- How can I describe the spread of a set of data?
- How can I use data to compare different groups?
- What conclusions can be drawn from data?

STANDARDS ADDRESSED:

MCC6.SP.2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
MCC6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
MCC6.SP.4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
MCC6.SP.5. Summarize numerical data sets in relation to their context, such as by:
MCC6.SP.5.a. Reporting the number of observations.
MCC6.SP.5.b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement
MCC6.SP.5.c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
MCC6.SP.5.d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered

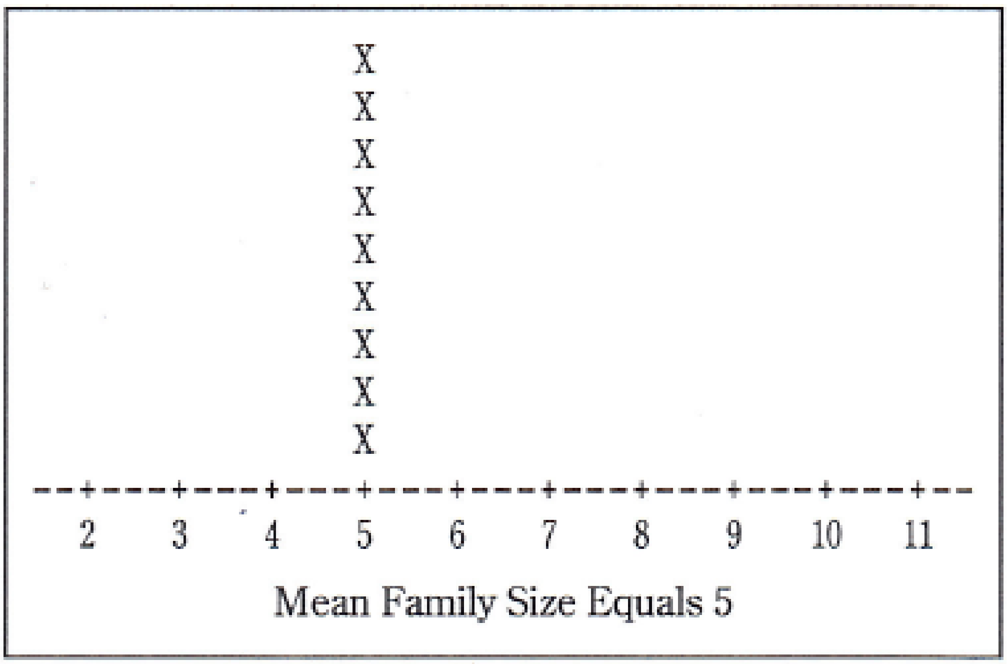
A Deeper look at Mean Absolute Deviation

Adapted from: Kader, Gary D. "Means and MADs." *Mathematics Teaching in the Middle School* 4.6 (1999): 398-403. Print.

In Statistics, data is collected to help make conclusions about people, things, and situations, or whatever you want to collect data on!

Comment: Each group is given a poster board with a number line labeled 2-11. They are also given 9 blank sticky notes.

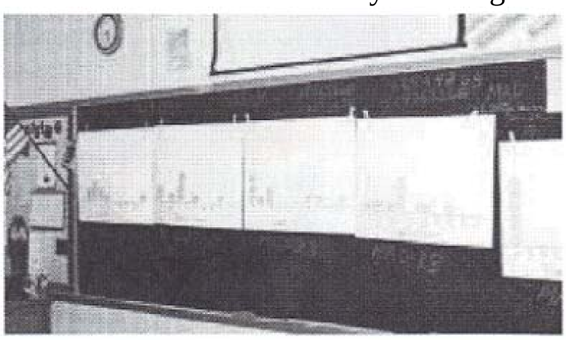
1. Working in groups of 2 or 3, come up with a distribution on your number line that would yield a mean family size of 5. Be creative with your answer and check your answer for accuracy.
Comment: Allow only one or two groups to have symmetric distributions. They will figure out that the total of the family sizes must be 45, so that the average will yield 5.



One example of a mean family size of 5.

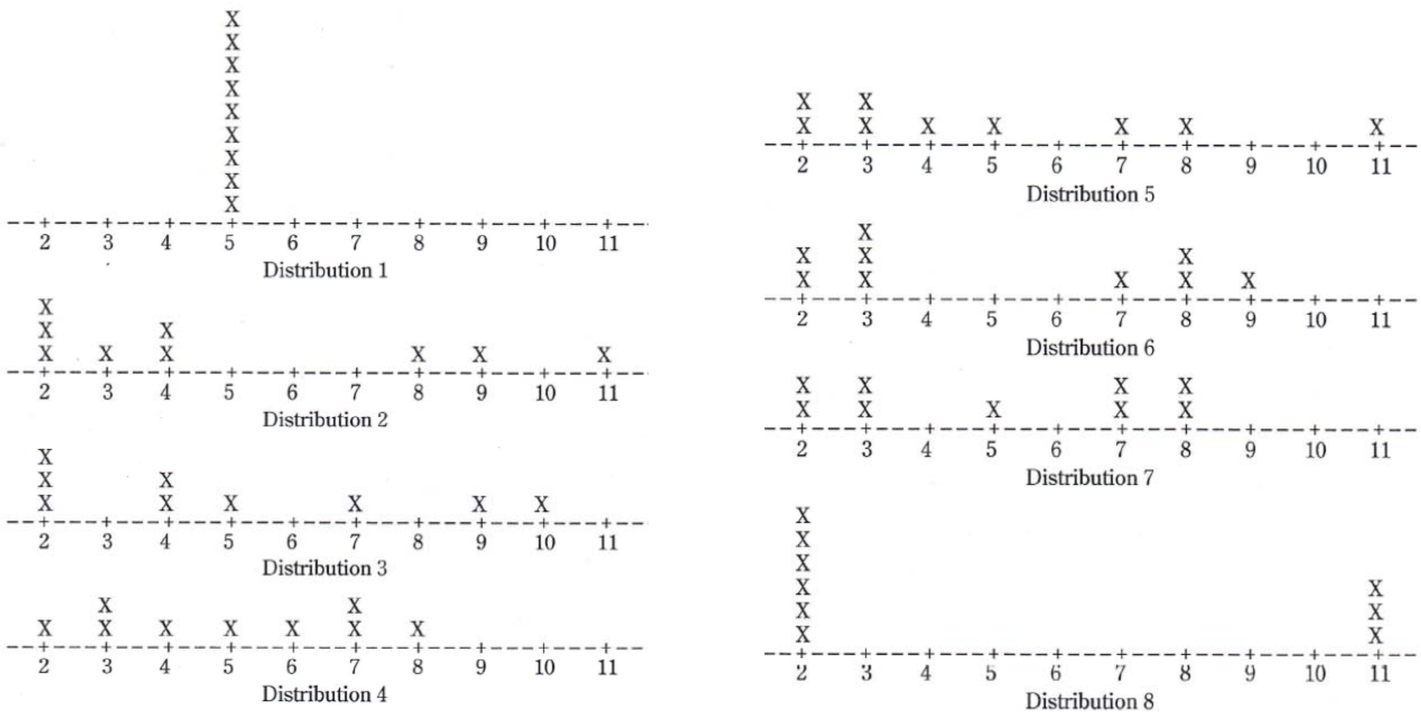
2. What method did you use to come up with 9 numbers that averaged to 5? Explain in detail.
Comment: Example of a groups' strategy: "What our group did was guess and check. We just took a big group of nine numbers and added them up. We kept changing the numbers around until we got 45."

3. Display all of the distributions on the board. Look at the different distributions displayed. Discuss the limitations of only knowing the mean family size, instead of the actual data pieces.



Comment: Example: "If you only have the mean, it limits you to not knowing how many people are in each family."

Only knowing the mean of a data set limits our knowledge on each individual piece of data in the set.
Look at the eight data sets displayed. The mean of ALL of the data sets is 5.



4. Looking at the data sets, which one seems to differ the least from the mean? Explain why you chose this data set.
Students should notice that Distribution 1 has data values that differ the least from the mean of 5.

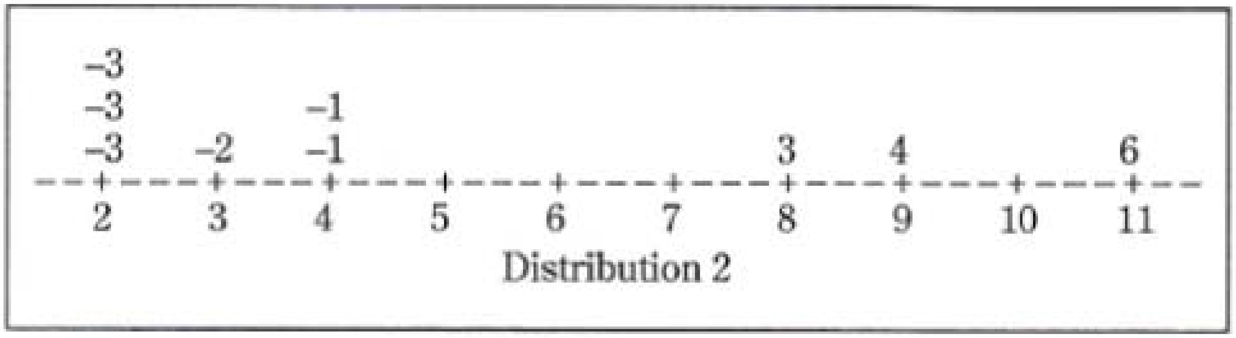
5. Which seems to differ the most from the mean? Explain why you chose this data set.
Students should recognize that distribution 8 has data values that differ the most form the mean of 5
6. With your group, put all of the data sets in order from “Differs Least” from the mean to “Differs Most” from the mean. How did you come up with this list?
Students use visual examination to do this. Most will have difficulty expressing their reasoning, and none may adequately express how they arrive at their list.

7. Share your groups order with the class.

8. As a class, decide on the “best” order for all of the data sets varying least from the mean to greatest from the mean.
Let the class vote on the best order. Many will disagree, but they should recognize the subjectivity of only a visual examination. This leads into finding a quantitative measure (the MAD) that help make decisions about the spread of the data.

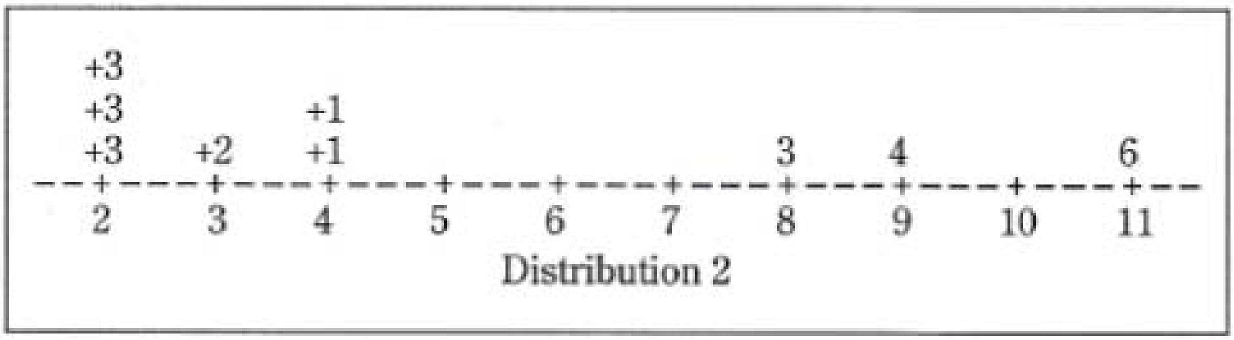
The MAD

9. One way to describe how far a value is from the mean is called the "deviation" from the mean.
Deviation= Value-Mean
Below is a display of the deviations for distribution 2:



10. Since the sum of all deviations from the mean equals zero, let's look at the *distance* each value is away from zero (the absolute value of each individual data piece). Let's look at Distribution 2. Determine each value's *distance* away from zero.

$$Distance\ from\ the\ mean = |deviation\ from\ mean|$$



11. If I were to ask you "on average" how different the data values are away from the mean, you can use the Mean Absolute Deviation to find this. Total up the distances away from the mean, and then find the "average" of these by dividing by the total number of values in the distribution.

$$MAD = \frac{Total\ Distance\ (of\ all\ values\ from\ the\ mean\ value)}{number\ of\ values\ (sample\ size)}$$

12. A small MAD means that the values do not vary much from the mean. Find the MAD of Distribution 4. First find the distances each value is from the mean. Total the distances, then divide by 9 (the number of values).

$$\boldsymbol{MAD\ for\ Distribution\ 4 = \frac{distance\ each\ value\ is\ away\ from\ the\ mean\ of\ 5}{total\ values}}$$

$$\boldsymbol{= \frac{3+2+2+1+0+1+2+2+3}{9} = \frac{16}{9} = 1.78}$$

13. A large MAD means that the values vary greatly from the mean. Find the MAD of Distribution 6.

$$MAD\ for\ Distribution\ 6 = \frac{3+3+2+2+2+2+3+3+4}{9} = \frac{24}{9} = 2.67$$

14. Find the MAD for the rest of the Distributions 1-8.

Distribution	***MAD***
1	***0.00***
2	***2.89***
3	***2.44***
4	***1.78***
5	***2.44***
6	***2.67***
7	***2.22***
8	***4.00***

15. Re-Order your distributions from the smallest MAD to the largest MAD.
Distribution 3 and 5 have the same MAD, so there are two possible answers.
4,7,3,5,6,2 or 4,7,5,3,6,2

16. Remember, the Mean Absolute Deviation tells you "on average" how different the data values are away from the mean. Using the MAD you found for Distribution 2, explain what the MAD means for that distribution.
The MAD is how each data value differs, on average, from the mean value. For distribution 2, each data value varies 2.89 units from mean for Distribution 2, on average.

Task: How Many People Are in Your Family?

A Deeper look at Mean Absolute Deviation

Adapted from: Kader, Gary D. “Means and MADs.” *Mathematics Teaching in the Middle School* 4.6 (1999): 398-403. Print.

Materials Needed: butcher paper or poster board, sticky notes, calculator

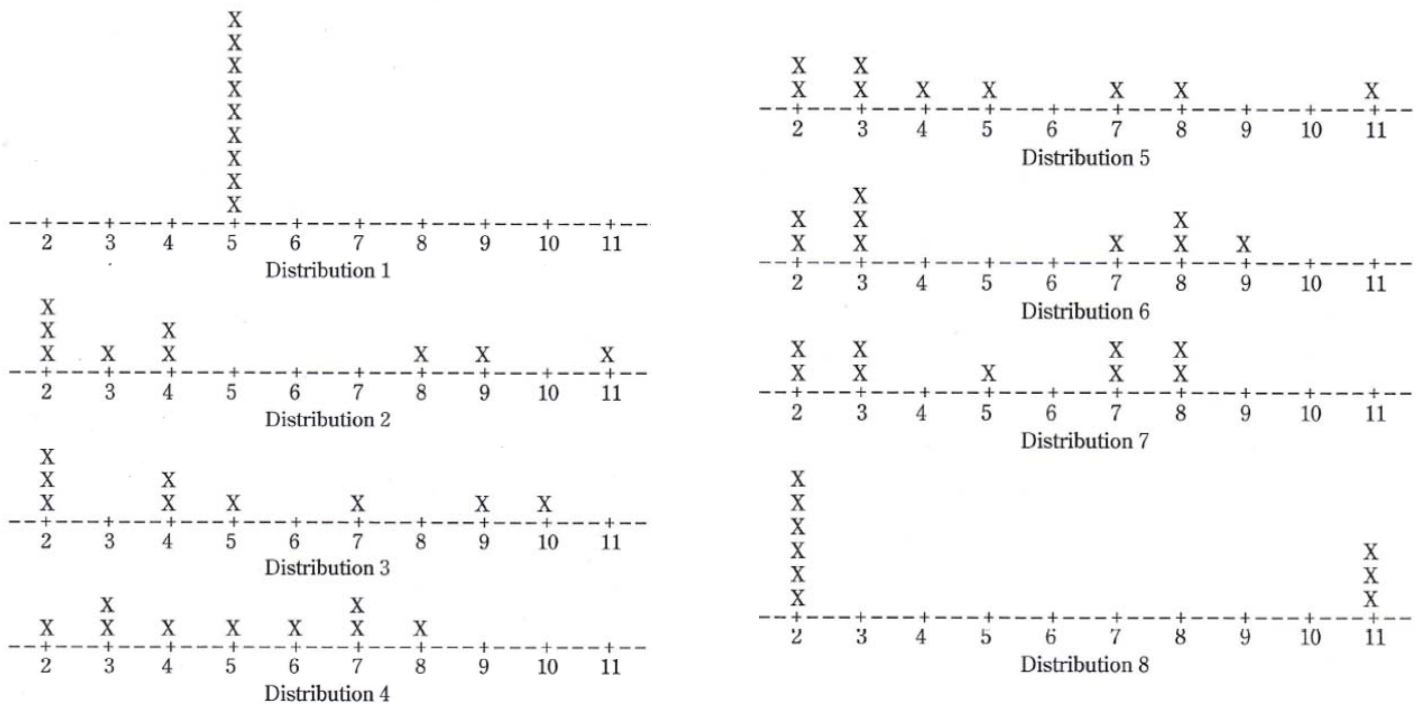
In Statistics, data is collected to help make conclusions about people, things, and situations, or whatever you want to collect data on!

1. Working in groups of 2 or 3, come up with a distribution of nine data pointd on your number line that would yield a mean family size of 5. Be creative with your answer and check your answer for accuracy.

2. What method did you use to come up with 9 numbers that averaged to 5? Explain in detail.

3. Look at the different distributions displayed. Discuss the limitations of only knowing the mean family size, instead of the actual data pieces.

Only knowing the mean of a data set limits our knowledge on each individual piece of data in the set. Look at the nine data sets below. The mean of ALL of the data sets is 5.



4. Looking at the data sets, which one seems to differ the least from the mean? Explain why you chose this data set.

5. Which seems to differ the most from the mean? Explain why you chose this data set.

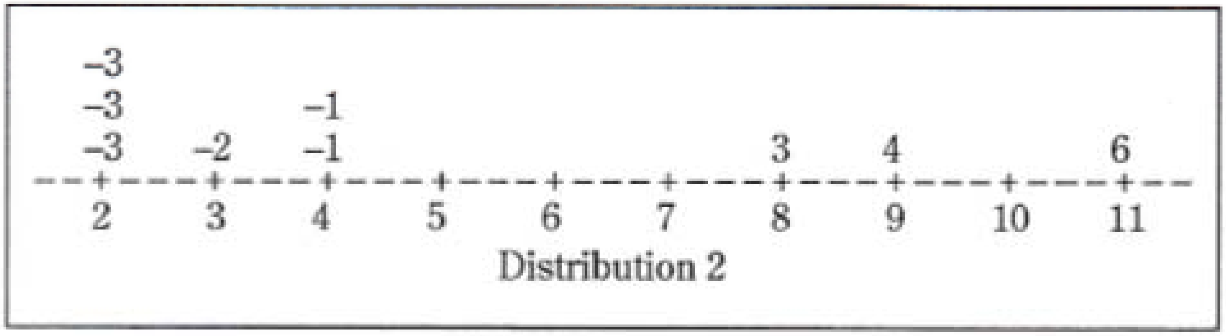
6. With your group, put all of the data sets in order from “Differs Least” from the mean to “Differs Most” from the mean. How did you come up with this list?

7. Share your groups order with the class.

8. As a class, decide on the “best” order for all of the data sets varying least from the mean to greatest from the mean.

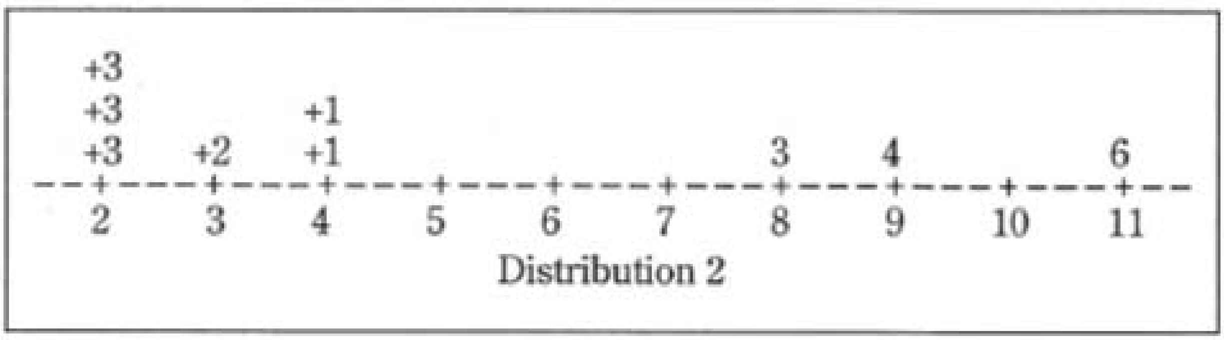
The MAD

9. One way to describe how far a value is from the mean is called the "deviation" from the mean.
Deviation= Value-Mean
Below is a display of the deviations for distribution 2:



10. Since the sum of all deviations from the mean equals zero, let's look at the *distance* each value is away from zero (the absolute value of each individual data piece). Let's look at Distribution 2. Determine each value's *distance* away from zero.

$$Distance\ from\ the\ mean = |deviation\ from\ mean|$$



11. If I were to ask you "on average" how different the data values are away from the mean, you can use the Mean Absolute Deviation to find this. Total up the distances away from the mean, and then find the "average" of these by dividing by the total number of values in the distribution.

$$MAD = \frac{Total\ Distance\ (of\ all\ values\ from\ the\ mean\ value)}{number\ of\ values\ (sample\ size)}$$

12. A small MAD means that the values do not vary much from the mean. Find the MAD of Distribution 4. First find the distances each value is from the mean. Total the distances, then divide by 9 (the number of values).

$$\boldsymbol{MAD\ for\ Distribution\ 4 = \frac{total\ distance\ each\ value\ is\ away\ from\ the\ mean\ of\ 5}{total\ values} =}$$

13. A large MAD means that the values vary greatly from the mean. Find the MAD of Distribution 6.

14. Find the MAD for the rest of the Distributions 1-8.

Distribution	***MAD***
1	
2	
3	
4	
5	
6	
7	
8	

15. Re-Order your distributions from the smallest MAD to the largest MAD.

16. Remember, the Mean Absolute Deviation tells you "on average" how different the data values are away from the mean. Using the MAD you found for Distribution 2, explain what the MAD means for Distribution 2.

Order Up! Fast Food Frenzy

Culminating Task

ESSENTIAL QUESTIONS

- What is meant by the center of a data set, how is it found and how is it useful when analyzing data?
- How do I choose and create appropriate graphs to represent data?
- How can I describe the spread of a set of data?
- How can I use data to compare different groups?
- What conclusions can be drawn from data?

STANDARDS ADDRESSED:

MCC6.SP.1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.

MCC6.SP.2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

MCC6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

MCC6.SP.4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

MCC6.SP.5. Summarize numerical data sets in relation to their context, such as by:

MCC6.SP.5.a. Reporting the number of observations.

MCC6.SP.5.b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement

MCC6.SP.5.c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

MCC6.SP.5.d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Every day in the United States, millions of customers eat fast food. There are so many different types of numerical data and information that we can gather about fast food, and then make decisions based on the evidence found.

Part 1: Burgers~ I'm Lovin' It!

Sandwiches	Total Fat (g)
Hamburger	9
Cheeseburger	12
Double Cheeseburger	23
McDouble	19
Quarter Pounder®	19
Quarter Pounder® with Cheese	26
Double Quarter Pounder® with Cheese	42
Big Mac®	29
Big N' Tasty®	24
Big N' Tasty® with Cheese	28
Angus Bacon & Cheese	39
Angus Deluxe	39
Angus Mushroom & Swiss	40
Filet-O-Fish®	18
McChicken ®	16
McRib ®†	26
Premium Grilled Chicken Classic Sandwich	10
Premium Crispy Chicken Classic Sandwich	20
Premium Grilled Chicken Club Sandwich	17
Premium Crispy Chicken Club Sandwich	28

Source: McDonald's USA

1. Create a stem and leaf plot of the data above. Which stem do majority of the sandwiches fall in?

0 : 9
1 : 0 2
1 : 6 7 8 9 9
2 : 0 3 4
2 : 6 6 8 8 9
3 :
3 : 9 9
4 : 0 2

a) What is the total number of observation (pieces of data) that we are analyzing? n=_____

n=20

Measures of Center

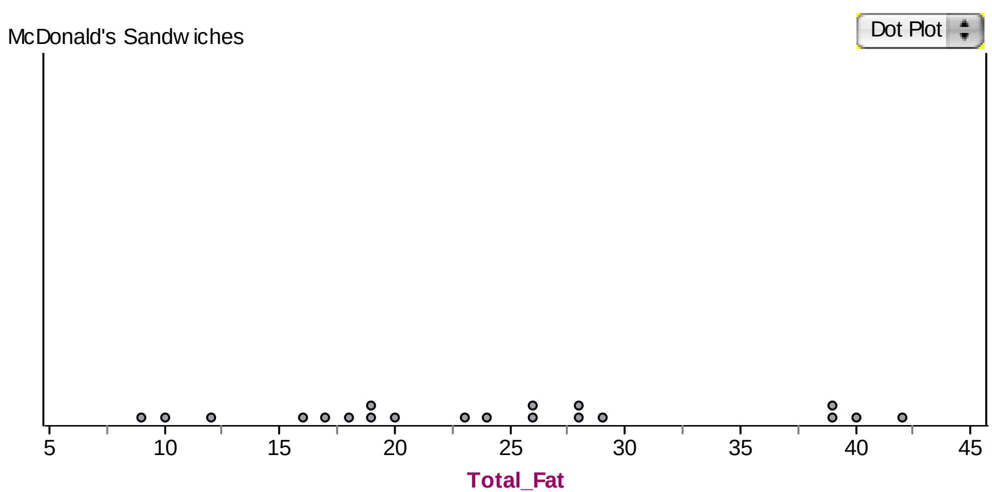
2. What is the mean (average) number of calories in a sandwich from McDonalds?

The mean of fat grams is 24.2.

3. In context of the data given, describe the meaning of the mean:

The mean number of fat grams that can be found in a McDonalds Sandwich is 24.2 grams.

4. Create a dot plot of the data. Be sure to make your tic marks evenly spaced on the axis below:



Number of Fat Grams

Measures of Spread

5. What is highest amount of fat grams? _____ What is the least amount of fat grams?________
The sandwich with the highest amount of fat grams is the Double Quarter Pounder with Cheese with 42 fat grams. The sandwich with the least amount of fat grams is the Hamburger with 9 fat grams.

6. What is the range of the data?___________
The range of the data is the highest value minus the lowest value. $42 - 9 = 33$

Explain what the range of the data means in terms of fat grams.
The range of fat grams that can be found in McDonald's sandwiches is 33 grams. This helps us conclude that there is a wide variety in the number of fat grams found in McDonald's sandwiches.

7. Find the **Mean Absolute Deviation (MAD)**, which is the average distance of each data value from the mean:
The mean is 24.2. So, you must find the DISTANCE that each sandwich is from the mean. Then total the distances, and divide by the number of data values.

Sandwiches	*Total Fat (g)*	*Distance from the Mean of 24.2*
Hamburger	*9*	*15.2*
Cheeseburger	*12*	*12.2*
Double Cheeseburger	*23*	*1.2*
McDouble	*19*	*5.2*
Quarter Pounder®	*19*	*5.2*
Quarter Pounder® with Cheese	*26*	*1.8*
Double Quarter Pounder® with Cheese	*42*	*17.8*
Big Mac®	*29*	*4.8*
Big N' Tasty®	*24*	*0.2*
Big N' Tasty® with Cheese	*28*	*3.8*
Angus Bacon & Cheese	*39*	*14.8*
Angus Deluxe	*39*	*14.8*
Angus Mushroom and Swiss	*40*	*15.8*
Filet-O-Fish ®	*18*	*6.2*
McChicken ®	*16*	*8.2*
McRib ®	*26*	*1.8*
Premium Grilled Chicken Classic Sandwich	*10*	*14.2*
Premium Crispy Chicken Classic Sandwich	*20*	*4.2*
Premium Grilled Chicken Club Sandwich	*17*	*7.2*

Premium Crispy Chicken Club Sandwich	***28***	***3.8***

Sum of Distances from the Mean=	***158.4***

$$MAD = \frac{\text{sum of distances from mean}}{\text{total number of observations}} = \frac{158.4}{20} = 17.6$$

a. A small MAD indicates that the data values are similar to the mean. A large MAD indicates that the data values are quite different from the mean. Interpret what type of MAD you have found for this problem.
There is a large MAD for the fat grams in McDonald's sandwiches. This means that there is a wide range of fat grams in sandwiches, and these values can vary significantly from the mean.

b. Explain what the MAD means in terms of fat grams.
The MAD states that, on average, the number of fat grams in a McDonald's sandwiches varies from the mean by 17.6 grams.

Shape

8. Are there about the same amount of data on both sides of the mean? Look at your dot plot to inspect this (this is called a "symmetric" shape). If not, then your dot plot might be skewed (the bulk of the data on one side, while the other side tails off).
Looking at both sides of the mean on the dot plot, there appears to be approximately the same number of observations on each side of the mean. So this data set is approximately symmetric

Outliers

9. Are there any points that are clearly apart from the body of the data (outliers)? Y/N.
No, there does not seem to be any clear outliers for this data set. There is a break in the data set between 29 grams and 39 grams, but since there are 4 sandwiches that have fat grams of 39 or higher, they should not be considered outliers.

10. Critical Thinking: *According the USDA, the 30% of your daily calorie intake should come from fat. For a person that consumes 2,000 calories/day, this is 67 grams of fat.*
Based on this information, which type of sandwich would you NOT suggest an individual to eat and why?
Sample Answer:
The double quarter pounder with cheese is clearly the highest in fat, with 42 fat grams. I wouldn't eat any of the sandwiches above 29 fat grams. Since this is only one meal of the day, it would be difficult to stay within the USDA recommendation when eating almost half of your fat grams in one meal.

Part 2: Biggie Size those Fries…. Or Maybe Not!!!

	Small	Medium	Large
McDonalds	230	380	500
Wendy's	330	410	540
Burger King	340	440	540
Dairy Queen		310	500
Sonic	204	326	448
Steak and Shake	240	440	640
Chick-fil-A	290	380	430
Arby's	410	540	640

1. Find the **5 Number Summary** for each size of fries:

5 Number Summary	Small Fry	Medium Fry	Large Fry
Minimum	*204*	*310*	*430*
Quartile 1	*230*	*353*	*474*
Median (Quartile 2)	*290*	*395*	*520*
Quartile 3	*340*	*440*	*590*
Maximum	*410*	*540*	*640*

2. Make stacked box plots of each size of fry below, using the information from above. Be sure to mark the horizontal axis with calories.

Small

Medium

Large

Amount of Calories

Spread

3. What is the range of calories for small, medium, and large fries?

Small=__________ Medium=____________ Large=____________

Comment
It is useful for students to know how the data is spread out.

Solution
The range of calories is 136 calories for small, 230 calories for medium, and 210 calories for a large fry.

4. Using the five-number summary, find the Inter-Quartile Range (IQR) for small, medium, and large fries.

Small Fry IQR=________ Medium Fry IQR=_________ Large Fry IQR=________

Solution: Small Fry IQR: 340-230=110. Medium Fry IQR: 440-353=87. Large Fry IQR: 590-474=116

5. Which size fry has the greatest range? ________ Which size fry has the greatest IQR? _________.

Solution: Medium fries have the greatest range and large fries have the greatest IQR.

6. In trying to determine which size fry has the more consistent amount of calories, would you compare the ranges or the IQR? Explain why you chose the measure you did.

Solution: The IQR is the middle 50% of the data, so it may be more useful to look at instead of the range. The IQR eliminates any potential outliers, so it is a better "snapshot" of the data.

Shape

7. Do the box plots appear to have symmetrical? Justify your response.
Solution: No, none of these boxes have symmetry about the median.

8. Are there any noticeable outliers in any size of fries? If so, state the size of the fry and the restaurant.
Solution: The medium fry has a long whisker going towards the maximum, so 540 appears to be an outlier in the medium fries.

9. Overall, which restaurant had fries that were very high in calories?_______________
Solution: Arby's

Which restaurant had fries that were overall lowest in calories? ___________________
Solution: Sonic

Variability

10. Name some factors that might cause one restaurant's large fry order be much higher in calories than another restaurant's large fry. These factors are called *uncontrolled variables.*
Sample Solution: Each restaurants' Small, Medium, and Large fry may vary in size. The type of oil that the restaurant uses to fry can make calories higher or lower.

Part 3: I'm Thirsty!

Below is a list of drinks to choose from at McDonalds.

	Calories in 21 fluid ounces	Grams of Sugar in 21 fluid ounces
Coca-Cola Classic®	210	40
Dasani Water®	0	0
Diet Coke®	0	0
Sprite®	210	56
Hi-C® Orange	240	64
Orange Juice (22 oz)	280	58
Powerade ®	150	31
Sweet Tea	180	45
Iced Mocha (22 oz)	340	41
Hot Chocolate (20 oz)	460	54
Frappe Mocha (22 oz)	680	87
Strawberry Bananna Smoothie (22 oz)	330	70
Chocolate McCafe Shake (22 fl oz)	720	98

1. Create a histogram displaying the number of calories found in the popular drinks at McDonalds.

Comment: Histograms can look very different, if different class widths are chosen in the frequency table. (Example counting from 0-10 for one histogram bar versus counting from 0-20). So, histograms can look different. You can suggest a class width size to the class

2. Find the mean and the median of the calorie data. Which measure of center would best describe the caloric data and why?

Solution: The mean is 292.31 and the median is 240 calories. The median is a better choice for describing data because it is the center of the data. The mean is strongly affected by outlier (such as 0 calories and 720 calories in this case).

3. Find the Mean Absolute Deviation for the number of calories found in a drink at McDonalds. Then, interpret the Mean Absolute Deviation in terms of caloric intake.

Solution: The MAD for the drinks at McDonalds is about 164.38. There is a wide variance in the amount of calories in the drinks at McDonalds.

4. Create a histogram in the graphing calculator using the grams of sugar data. Sketch a graph of your histogram below.

Comment: Histograms can look very different, if different class widths are chosen in the frequency table. (Example counting from 0-10 for one histogram bar versus counting from 0-20). So, histograms can look different. You can suggest a class width size to the class.

5. Find your mean and median using the 1-VAR STATS tool under the STATS menu. Which would be best in describing grams of sugar in drink at McDonalds?
Solution: Mean is 49.54 grams and the median is 54 grams. The median is the best description since it is not influenced by outliers.

6. Were there any outliers present in the grams of sugar data? How did the outlier(s) affect the mean of the data? How did the outlier(s) affect the median of the data?
Sample Answer: 0 grams of sugar is certainly an outlier, (and 87 and 98 grams of sugar also seem to not fit with the bulk of the data). The drinks with 0 grams of sugar would pull the mean lower than it would be without this outlier being present.
These outliers do not greatly affect the median of the data.

7. Is Range or Mean Absolute Deviation best when describing any distribution? Why?
Comment: some students may not choose to use MAD because we just determined the mean in the last problem was affected by outliers. The MAD is also affected by outliers, but gives a more detailed description of the spread of the data than the range of the data.
Sample Answer: MAD gives a better picture about how many grams of sugar is present in drinks, ON AVERAGE. The MAD also takes every data value into consideration, while the range only uses the first and last numbers.

Order Up! Fast Food Frenzy
Culminating Task

Every day in the United States, millions of customers eat fast food. There are so many different types of numerical data and information that we can gather about fast food, and then make decisions based on the evidence found.

Part 1: Burgers~ I'm Lovin' It!

Sandwiches	Total Fat (g)
Hamburger	9
Cheeseburger	12
Double Cheeseburger	23
McDouble	19
Quarter Pounder®	19
Quarter Pounder® with Cheese	26
Double Quarter Pounder® with Cheese	42
Big Mac®	29
Big N' Tasty®	24
Big N' Tasty® with Cheese	28
Angus Bacon & Cheese	39
Angus Deluxe	39
Angus Mushroom & Swiss	40
Filet-O-Fish®	18
McChicken ®	16
McRib ®†	26
Premium Grilled Chicken Classic Sandwich	10
Premium Crispy Chicken Classic Sandwich	20
Premium Grilled Chicken Club Sandwich	17
Premium Crispy Chicken Club Sandwich	28

Source: McDonald's USA

1. Create a stem and leaf plot of the data above. Which stem do majority of the sandwiches fall in?

a) What is the total number of observation (pieces of data) that we are analyzing? n=_____

Measures of Center

2. What is the mean (average) number of calories in a sandwich from McDonalds?

3. In context of the data given, describe the meaning of the mean:

4. Create a dot plot of the data. Be sure to make your tic marks evenly spaced on the axis below:

Number of Fat Grams

Measures of Spread

5. What is highest amount of fat grams? _____ What is the least amount of fat grams?________

6. What is the range of the data?__________

Explain what the range of the data means in terms of fat grams.

7. Find the **Mean Absolute Deviation (MAD)**, which is the average distance of each data value from the mean:

The mean is 24.2

Sandwiches	***Total Fat (g)***	***Distance from the Mean of 24.2***
Hamburger	***9***	
Cheeseburger	***12***	
Double Cheeseburger	***23***	
McDouble	***19***	
Quarter Pounder®	***19***	
Quarter Pounder® with Cheese	***26***	
Double Quarter Pounder® with Cheese	***42***	
Big Mac®	***29***	
Big N' Tasty®	***24***	
Big N' Tasty® with Cheese	***28***	
Angus Bacon & Cheese	***39***	
Angus Deluxe	***39***	
Angus Mushroom and Swiss	***40***	
Filet-O-Fish ®	***18***	
McChicken ®	***16***	
McRib ®	***26***	
Premium Grilled Chicken Classic Sandwich	***10***	
Premium Crispy Chicken Classic Sandwich	***20***	
Premium Grilled Chicken Club Sandwich	***17***	
Premium Crispy Chicken Club Sandwich	***28***	

Sum of Distances from the Mean=	

$$MAD = \frac{sum\ of\ distances\ from\ mean}{total\ number\ of\ observations} = \frac{\quad}{\quad} =$$

a. A small MAD indicates that the data values are similar to the mean. A large MAD indicates that the data values are quite different from the mean. Interpret what type of MAD you have found for this problem.

b. Explain what the MAD means in terms of fat grams.

Shape

8. Are there about the same amount of data on both sides of the mean? Look at your dot plot to inspect this (this is called a “symmetric” shape). If not, then your dot plot might be skewed (the bulk of the data on one side, while the other side tails off).

Outliers

9. Are there any points that are clearly apart from the body of the data (outliers)? Y/N.

10. Critical Thinking: *According the USDA, the 30% of your daily calorie intake should come from fat. For a person that consumes 2,000 calories/day, this is 67 grams of fat.*
Based on this information, which type of sandwich would you NOT suggest an individual to eat and why?

Part 2: Biggie Size those Fries…. Or Maybe Not!!!

	Small	Medium	Large
McDonalds	230	380	500
Wendy's	330	410	540
Burger King	340	440	540
Dairy Queen		310	500
Sonic	204	326	448
Steak and Shake	240	440	640
Chick-fil-A	290	380	430
Arby's	410	540	640

1. Find the <u>5 Number Summary</u> for each size of fries:

5 Number Summary	Small Fry	Medium Fry	Large Fry
Minimum			
Quartile 1			
Median (Quartile 2)			
Quartile 3			
Maximum			

2. Make stacked box plots of each size of fry below, using the information from above. Be sure to mark the horizontal axis with calories.

Small

Medium

Large

____

Amount of Calories

Spread

3. What is the range of calories for small, medium, and large fries?

Small=__________ Medium=____________ Large=____________

4. Using the five-number summary, find the Inter-Quartile Range (IQR) for small, medium, and large fries.

Small Fry IQR=________ Medium Fry IQR=_________ Large Fry IQR=________

5. Which size fry has the greatest range? ________ Which size fry has the greatest IQR? _________.

6. In trying to determine which size fry has the more consistent amount of calories, would you compare the ranges or the IQR? Explain why you chose the measure you did.

Shape

7. Do the box plots appear to be symmetrical? Justify your response

8. Are there any noticeable outliers in any size of fries? If so, state the size of the fry and the restaurant.

9. Overall, which restaurant had fries that were very high in calories?_______________

Which restaurant had fries that were overall lowest in calories? ___________________

Variability

10. Name some factors that might cause one restaurant's large fry order be much higher in calories than another restaurant's large fry. These factors are called *uncontrolled variables.*

Part 3: I'm Thirsty!

Below is a list of drinks to choose from at McDonalds.

	Calories in 21 fluid ounces	Grams of Sugar in 21 fluid ounces
Coca-Cola Classic®	210	40
Dasani Water®	0	0
Diet Coke®	0	0
Sprite®	210	56
Hi-C® Orange	240	64
Orange Juice (22 oz)	280	58
Powerade ®	150	31
Sweet Tea	180	45
Iced Mocha (22 oz)	340	41
Hot Chocolate (20 oz)	460	54
Frappe Mocha (22 oz)	680	87
Strawberry Bananna Smoothie (22 oz)	330	70
Chocolate McCafe Shake (22 fl oz)	720	98

1. Create a histogram displaying the number of calories found in the popular drinks at McDonalds.

2. Find the mean and the median of the calorie data. Which measure of center would best describe the caloric data and why?

3. Find the Mean Absolute Deviation for the number of calories found in a drink at McDonalds. Then, interpret the Mean Absolute Deviation in terms of caloric intake.

4. Create a histogram in the graphing calculator using the grams of sugar data. Sketch a graph of your histogram below.

5. Find your mean and median using the 1-VAR STATS tool under the STATS menu. Which would be best in describing grams of sugar in drink at McDonalds?

6. Were there any outliers present in the grams of sugar data? How did the outlier(s) affect the mean of the data? How did the outlier(s) affect the median of the data?

7. Is Range or Mean Absolute Deviation best when describing any distribution? Why?