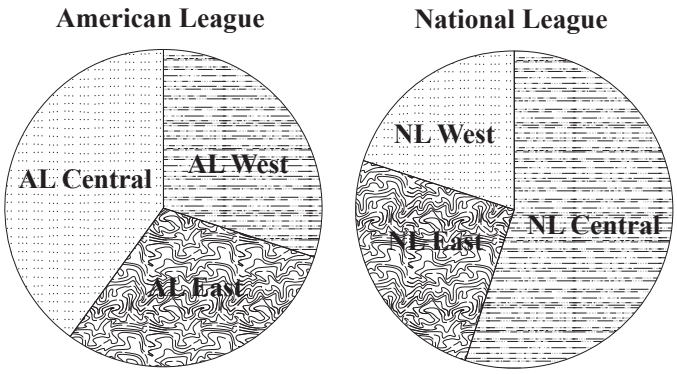


Home Runs, Statistics, and Probability

LEVEL

7th grade in a unit on graphical displays

CONNECTION TO AP*

Graphical Display and Distributions

**Advanced Placement and AP are registered trademarks of the College Entrance Examination Board. The College Board was not involved in the production of this product.*

MODALITY

NMSI emphasizes using multiple representations to connect various approaches to a situation in order to increase student understanding. The lesson provides multiple strategies and models for using those representations indicated by the darkened points of the star to introduce, explore, and reinforce mathematical concepts and to enhance conceptual understanding.

P – Physical
V – Verbal
A – Analytical
N – Numerical
G – Graphical

ABOUT THIS LESSON

In this lesson, students display two related sets of data using a circle graph, a dotplot, a stemplot, a split stemplot, and a boxplot. They use both the data and the graphical displays to determine measures of central tendency. Students also use the data and displays to confirm or refute given statements, determine probabilities, and make and defend their own conjectures.

OBJECTIVES

Students will

- determine measures of center and range.
- create graphical displays.
- interpret graphical displays.
- determine probability for simple events.

TEACHER PAGES

COMMON CORE STATE STANDARDS FOR MATHEMATICAL CONTENT

This lesson addresses the following Common Core State Standards for Mathematical Content. The lesson requires that students recall and apply each of these standards rather than providing the initial introduction to the specific skill.

Targeted Standards

7.SP.4 Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book.*
See questions 4, 9, 10

7.SP.5 Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.
See questions 2d, 2f-g, 3c, 3e, 5-7

6.SP.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
See questions 2, 8a

6.SP.5c Summarize numerical data sets in relation to their context, such as by:
(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 was gathered.
See questions 2, 4b-e, 8b-c

Reinforced/Applied Standards

6.SP.5a Summarize numerical data sets in relation to their context, such as by:
(a) reporting the number of observations.
See questions 2, 4-7

6.SP.5d Summarize numerical data sets in relation to their context, such as by:
(d) relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data was gathered.
See questions 4b

6.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.
See questions 2, 9-10

6.RP.3c Use ratio and rate reasoning to solve real-world and mathematical problems, e.g., by reasoning about tables of equivalent ratios, tape diagrams, double number line diagrams, or equations.
(c) Find a percent of a quantity as a rate per 100 (e.g., 30% of a quantity means 30/100 times the quantity); solve problems involving finding the whole, given a part and the percent.
See questions 1, 2d-h, 4c, 4e, 5-7

COMMON CORE STATE STANDARDS FOR MATHEMATICAL PRACTICE

These standards describe a variety of instructional practices based on processes and proficiencies that are critical for mathematics instruction. NMSI incorporates these important processes and proficiencies to help students develop knowledge and understanding and to assist them in making important connections across grade levels. This lesson allows teachers to address the following Common Core State Standards for Mathematical Practice.

MP.1: Make sense of problems and persevere in solving them.
Students must decide if they need to use the data for the American League, National League, or both Leagues combined, read and interpret data from different graphical displays, and then choose a graphical display from the lesson that they feel best represents the data in order to write a paragraph to represent the data for a newspaper article.

MP.3: Construct viable arguments and critique the reasoning of others.
Students read and interpret multiple types of graphical displays to draw conclusions about the home run leaders that can and cannot be supported by the display.

FOUNDATIONAL SKILLS

The following skills lay the foundation for concepts included in this lesson:

- Determine mean, median, mode, and range for a data set
- Create a dotplot, stemplot, and boxplot

ASSESSMENTS

The following type of formative assessment is embedded in this lesson:

- Students engage in independent practice.

The following assessments are located on our website:

- Graphical Displays – 6th Grade Free Response Questions
- Graphical Displays – 7th Grade Free Response Questions
- Graphical Displays – 6th Grade Multiple Choice Questions
- Graphical Displays – 7th Grade Multiple Choice Questions

MATERIALS AND RESOURCES

- Student Activity pages

TEACHING SUGGESTIONS

The intent of this lesson is for students to learn to use graphical displays to determine measures of center and to analyze the data by confirming or refuting given statements. Students are also asked to make their own conjectures based on graphical displays.

This lesson gives students two related sets of data based on number of home runs hit by the leading batters in each of the American and National Leagues at a specific point in time. Students use this data to construct a line plot, a stem-and-leaf plot, and a box-and-whisker plot and then use the graphical displays to interpret the given data. Note: In AP* Statistics these graphical displays are called dotplot, stemplot, and boxplot, respectively.

To help students who might struggle with an explanation for each part of question 1, work through part (a) together. Students will base their answers on either the table or the graphs.

If students need clarification for the definitions of measures of central tendency, work question 2 as an entire class activity. Model writing an explanation of how to determine median by first asking students: "What would you do to determine the median of these data?" Until you have an exact process, continue to refine this concept with student input.

In question 3, students are asked to create a split stemplot. Since this graphical display is new for most students, take the time to explain how the data is divided between the two stems. For the first stem, record the leaves from 0 – 4. The second stem is used for leaves 5 – 9. This type of graphical display is used when each stem contains a large number of leaves. The split stemplot allows for a better visual representation of the distribution of the data.

Question 4 asks the students to create a back-to-back stemplot. If students have not mastered the creation of this graphical display, questions 4a can be completed as an entire class activity. Spend time working with students on creating their own conjectures for parts b – d. To modify this lesson, ask students leading questions to help them determine information that can and cannot be determined by the stemplot. For example, a possible answer for question 4b could be that the American League had one player that hit more homeruns than anyone else. For question 8b, students are asked the same types of questions for a boxplot. Encourage students to use percentages based on the 5-number summary to make their statements. For example, 75% of the home run leaders hit more than 26 home runs.

As students complete question 9, ask them which graphical display they used to refute or support their answers. To scaffold the lesson, tell the students which graphical display will help them decide if the statement is true or false.

Suggested modifications for additional scaffolding include the following:

1 Transfer the data from the table to the circle graph.

2 Provide the dot plot with the American League data and have students complete the National League.

2f Provide the value of the mean.

8c Provide the method for calculating the IQR, $Q_3 - Q_1$.

8d Provide the method for calculating the outliers, $Q_1 - 1.5(IQR)$ and $Q_3 + 1.5(IQR)$

NMSI CONTENT PROGRESSION CHART

In the spirit of NMSI's goal to connect mathematics across grade levels, a Content Progression Chart for each module demonstrates how specific skills build and develop from sixth grade through pre-calculus in an accelerated program that enables students to take college-level courses in high school, using a faster pace to compress content. In this sequence, Grades 6, 7, 8, and Algebra 1 are compacted into three courses. Grade 6 includes all of the Grade 6 content and some of the content from Grade 7, Grade 7 contains the remainder of the Grade 7 content and some of the content from Grade 8, and Algebra 1 includes the remainder of the content from Grade 8 and all of the Algebra 1 content.

The complete Content Progression Chart for this module is provided on our website and at the beginning of the training manual. This portion of the chart illustrates how the skills included in this particular lesson develop as students advance through this accelerated course sequence.

6th Grade Skills/Objectives	7th Grade Skills/Objectives	Algebra 1 Skills/Objectives	Geometry Skills/Objectives	Algebra 2 Skills/Objectives	Pre-Calculus Skills/Objectives
Create, interpret, and compare dotplots (line plots), stemplots, and bar graphs.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.
Calculate the mean, median, mode, and range from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, and range from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, and range from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, and range from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, and range from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, and range from tabular or graphical data or data presented in paragraph form.

Home Runs, Statistics, and Probability

Answers

1. a. No, this data is just the top 20 home run hitters in each league.

   b. Yes, the National League Central division has 11 of the top 20 home run hitters in the National League which is 55%.

   c. No, there are 8 players from the AL east division while there are 6 players from the west and central divisions.

   d. Yes, 15 of the 20 which is 75% of the home run hitters are in the west and central divisions.

   e. Cannot be determined from the data given. We only have information on the number of home run hitters not the number of home runs hit.

2. **Home Runs Hit**

   a. Mode: For the top home run hitters in both leagues, the most common number of home runs hit was 24 and 27. On the dotplot the two numbers of home runs with the most dots are 24 and 27.

   b. Median: At least 50% of the home run leaders hit greater than or equal to 28.5 home runs. On the dotplot the tenth value is 28 and the eleventh is 29 which means that the median is $\frac{28+29}{2} = 28\frac{1}{2}$.

   c. Range: The range, the difference of the greatest and least amount of home runs hit, of the top home run hitters is 14 home runs. The smallest value shown on the dotplot is 24, and the largest is 38, so the range is $38-24=14$.

   d. $\frac{7}{20} = 35\%$ of leading home run hitters in both leagues hit more than 30 home runs.

   e. The ratio of the number of leading hitters who hit 36 home runs to the player who hit 27 home runs is $2:3$.

   f. $\frac{8}{20} = 40\%$ of the players are above the mean and $\frac{12}{20} = 60\%$ of players are below the mean.

   g. $\frac{3}{20} = 15\%$ are leading home run hitters in both leagues that hit exactly 27 home runs.

h. The ratio of the number of leading hitters who hit less than 27 home runs to the number of leading hitters who hit more than 32 home runs but less than 38 home runs is $6:5$.

3.

Home Runs Hit

2	4 4 4
2	5 6 6 7 7 7 8 9 9
3	0 2 3 4 4
3	6 6 8

Key: 3|2 means 32 Home Runs

4. a.

Home Runs Hit

American League		National League
	2	4 4 4
9 7 7	2	5 6 6 7 8 9
4 4 3 2 0	3	
8 6	3	6

Key: 8|3|6 represents 38 Home Runs in the American League and 36 in the National League

b. Since the graph is skewed right, the mean would be greater than the median.

c. $\frac{7}{10} = 70\%$ of the American League home run hitters hit more home runs than all but one of the National League home run leaders.

d. No; Even though the National League top home run hitters appear to hit less than the American League, there is some overlap of the number of homeruns which means that Joey cannot say that all of the American league leaders are better batters than the National League top home run hitters.

e. The mean of the National League home run leaders is 26.9. $\frac{4}{10} = 40\%$ of National league home run leaders hit more home runs than the mean.

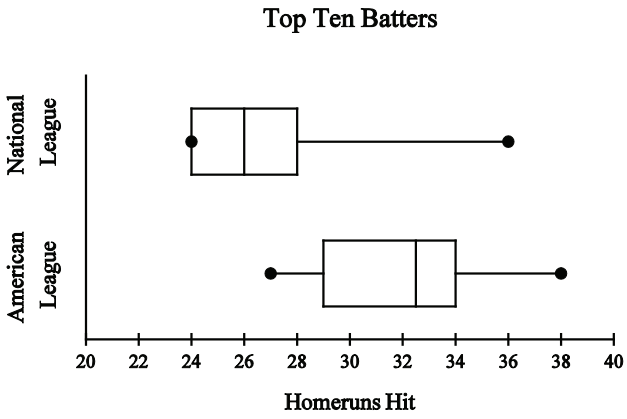
5. $\frac{3}{20}$, 0.15, 15%

6. $\frac{16}{20}$, 0.80, or 80%

7. $\frac{4}{10}$, 0.40, or 40%

8. a. American League: Minimum Value is 27; Lower Quartile is 29; Median is 32; Upper Quartile is 34; Maximum Value is 38.

   National League: Minimum Value is 24 Lower Quartile is 24; Median is 26; Upper Quartile is 28; Maximum Value is 36.



   b. Yes, because 75% (from Min to Q_3) of the boxplot for the National League lies to the left of Q_1 of the American League.

   c. National League IQR: $28-24=4$
   American League IQR: $34-29=5$
   The American League has an inner quartile range of 5 which is greater than the National League inner quartile range.

   d. Yes; The National League had a player hit 36 home runs which is more than 1.5 times the IQR away from the upper quartile for the National League data. $Q_3+1.5(IQR)\rightarrow 28+1.5(4)=34$

TEACHER PAGES

9. a. False. For example, some National League home run leaders hit more home runs than some of the American League home run leaders.

   b. True. For example, 80% of American League home run leaders hit more home runs than 80% of the National League home run leaders.

   c. False. Only one American League player had a higher number of home run hits than all 10 National League leaders.

   d. True. One American League player hit 38 home runs, more than any of the National League players.

   e. True. The National League player who hit the most home runs hit 36 home runs, which is the same as one player in the American League and less than the player with the most in the American League.

10. Example: The majority of American League home run leaders hit more than 30 home runs while the majority of the National League home run leaders hit less than 30 home runs.

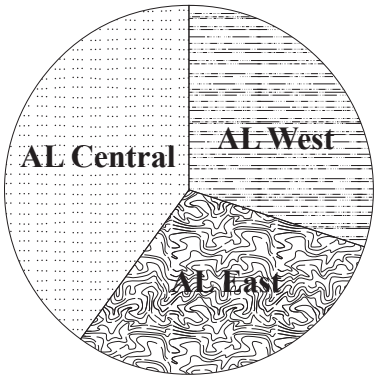
11. Answers will vary based upon students choice of graphical display.

Home Runs, Statistics, and Probability

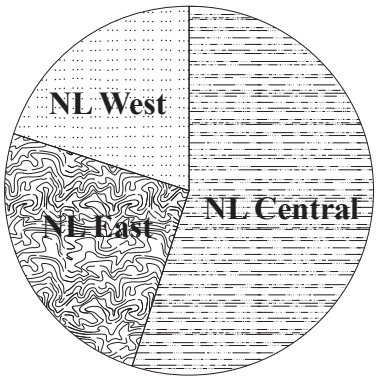
1. This data represents the 20 leading home run hitters in each of the American and National Leagues for 2012 as of September 12, 2012.

League	Division	Number of HR Hitters
American	Central	6
	East	6
	West	8
National	Central	11
	East	5
	West	4

American League



National League



Using the table and circle graphs, confirm or refute each statement or explain why neither can be justified.

a. The data given is representative of all of the players who hit home runs in both the National League and the American League.

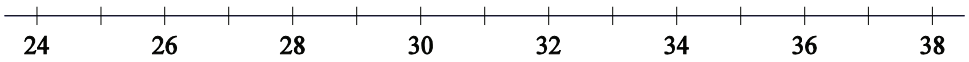
b. National League Central division has more than 50% of the top 20 home run hitters in the National League.

c. The American League top 20 home run hitters are evenly distributed among its three divisions.

d. Seventy-five percent of leading home run hitters in the National League play in the west or central division.

e. Seventy-five percent of home runs in the National League are hit by players in the west or central divisions.

2. As of September 12, 2012, the top 10 home run leaders in the American League had hit 38, 36, 34, 34, 33, 32, 30, 29, 27, and 27 home runs. The top 10 home run leaders in the National League had hit 36, 29, 28, 27, 26, 26, 25, 24, 24, and 24 home runs. Make one dotplot (line plot) from this information.



Use the dotplot to answer (a) – (h). Interpret each of the values in the context of the situation and how to determine each value from the dotplot for parts (a) – (c).

a. Mode

b. Median

c. Range

d. What percent of the leading home run hitters in both leagues hit more than 30 home runs?

e. What is the ratio of the number of leading hitters who hit 36 home runs to the players who hit 27 home runs?

f. Draw a vertical line on the dotplot to represent the value of the mean. What percent of the players are above the mean? What percent of the players are below the mean?

g. What is the percent of leading home run hitters in both leagues that hit exactly 27 home runs?

h. What is the ratio of the number of leading hitters who hit less than 27 home runs to the number of leading hitters who hit more than 32 home runs but less than 38 home runs?

3. Combine the data given in question 2 to complete the given split stemplot.-

Home Runs Hit

2	4 4 4
2	
3	
3	6

Key: 3|2 means 32 Home Runs

4. a. Use the data given in question 2 to complete the back-to-back stemplot.

Home Runs Hit

American League		National League
	2	4 4 4
	2	
	3	
8 6	3	

Key: 8|3|6 represents 38 Home Runs in the American League and 36 in the National League

b. Based on the shape of the stemplot for the National League, what is the relationship between the mean and the median? Explain your answer using only the graph.

c. What percent of the American League home run leaders hit more home runs than all but the National League home run leader who hit the most home runs? Show the work that leads to your answer.

d. Your friend Joey comments that all of the American League home run leaders are better batters than all of the National League home run leaders. Based on the stemplot, do you agree with his statement? Explain your reasoning using the graph.

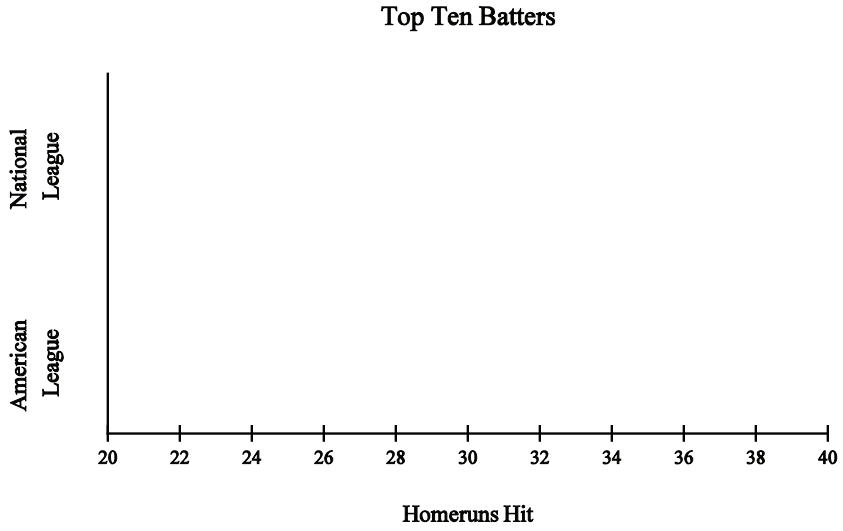
e. What percent of National League home run leaders hit more home runs than the mean of National League home run hits?

5. If a player is selected at random from either the National league or the American League, what is the probability that he hit exactly 27 home runs as of September 12, 2012?

6. If a player is selected at random from the National League or the American League, what is the probability that he hit more than 25 home runs as of September 12, 2012?

7. Given that a player is from the National League, what is the probability that he hit at least 27 home runs?

8. a. Determine the five-number summary for the data given in question 2 for each league and construct parallel boxplots. Label each boxplot with the five-number summary.



	AL	NL
Min		
Q_1		
Med		
Q_3		
Max		

b. Can one say that approximately 75% American League home run leaders hit more home runs than almost all the National League home run leaders? Explain using the boxplots.

c. Which league has the greatest IQR? Explain.

d. Joey states that one of the home run leaders from the National League is an outlier for the National League. Do you agree with this statement? Justify your response.

9. Use the graphical displays you created and the data given in question 2 to refute or support the following statements:
   a. When considering the top ten home run leaders in the American League and the National League, all of the American League leaders hit more home runs than all of the National League leaders.

   b. When considering the top ten home run hitters in the American League and the National League, in general, the American League home run leaders tend to hit more home runs than the National League home run leaders.

   c. All American League top home run hitters hit more home runs than all of the National League top ten home run hitters.

   d. At least one American League player hit more home runs than all National League players.

   e. The National League player who hit the most home runs hit more home runs than all but two batters in the American League.

10. Create one additional statement about the American League home run hitters and the National League home run hitters that you can support from the given data.

11. The local newspaper has asked you to write a paragraph on the leading home run hitters in the American League versus the National League. Choose one of the graphical displays from the lesson that you would use to publish in the newspaper and write a paragraph to compare the two leagues.