Math Objectives

- Students will describe the overall pattern of a distribution by its shape, center, and spread.
- Students will recognize and identify properties of distributions such as symmetry, skewness, and bimodality from boxplots.
- Students will identify outliers.
- Students will recognize that moving data past a quartile can affect that quartile's value.

Vocabulary

- boxplot
- skewness
- symmetry
- quartile
- outlier
- interval
- median
- interquartile range, IQR
- range
- distribution

About the Lesson

- This lesson involves representing distributions of data using boxplots.
- As a result, students will:
  - Develop the reasoning skills to recognize properties of a distribution from its boxplot.
  - Understand the relationship between individual data values and the five-number summary.
  - Move data within a dot plot and observe the changes within the corresponding boxplot to identify particular characteristics that most directly affect the five-number summary.

TI-Nspire™ Navigator™ System

- Send out the *Boxplots_Introduction.tns* file.
- Monitor student progress using Class Capture.
- Use Live Presenter to spotlight student answers.

Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software
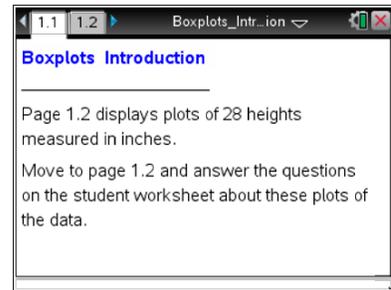
Tech Tips:

- This activity includes class captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/ calculators/pd/US/Online-Learning/Tutorials

Lesson Files:
Student Activity

- Boxplots_Introduction_ Student.pdf
- Boxplots_Introduction_ Student.doc

TI-Nspire document

- Boxplots_Introduction.tns

Discussion Points and Possible Answers

> **Tech Tip:** If students experience difficulty dragging the point, check to make sure that they have moved the arrow until it becomes a hand (✋) getting ready to grab the point. Also, be sure that the word *point* appears. Then press `ctrl` 🖱 to grab the point and close the hand (✊).

> **Tech Tip:** If students experience difficulty grabbing and dragging a point, have them tap and hold the desired point for a few seconds and then drag the point to the desired location.

> **Teacher Tip:** Students will reason abstractly and quantitatively. Students should reason from the graphical representations, both the dot plot and the bar graphs, about the population from which the samples are drawn. Multiple representations encourage a greater depth of student understanding.

In this activity, students will use boxplots to represent distributions of data of 28 heights measured in inches. They will also learn about the relationship between individual data values and the five-number summary.

Move to page 1.2.

1. In the top work area, move the cursor over the boxplot from left to right, and identify the five summary percentiles of the boxplot. These five important statistical values are called the five-number summary.
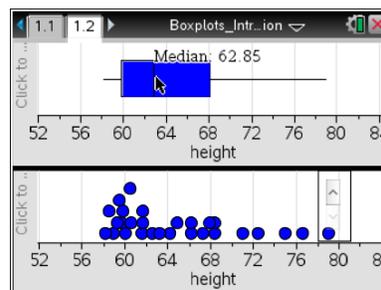
Answer:

Min = 58.12 inches

Q1 = 59.9 inches

Median = 62.85 inches

Q3 = 68.1 inches

Max = 79 inches

> **Tech Tip:** This document has an up arrow in the lower right work area of page 1.2 that can be used to return the data to its original state. If you change the data by mistake, select `ctrl` `Z` to return the data to the original state or to undo an action; select `ctrl` `Y` to redo an action.

Tech Tip: If you change the data by mistake, select the undo/redo icon ⟳ to return the data to the original state or to undo an action; select and hold on the icon and select **Redo** to redo an action. Or use the up arrow in the dot plot area of page 1.2.

Tech Tip: To display each of the five-number summary values on the boxplot, the students should tap the line or region on the boxplot.

Teacher Tip: Caution students not to grab and drag the data points until instructed to do so or the answers given below will be incorrect. Be sure students understand how to move points. When a point has been selected, it will be darkened. To deselect a point, move the cursor to an open region and select or tap. To move a single point, select a point in the dot plot and move the darkened point by grabbing and moving the point from either plot. To move several points at once, select each one, grab and move just one selected point, and all of the selected points will move as a unit. To move the set of points in a quartile of the boxplot, select that interval in the boxplot, and the points will drag as a group.

2. a. Determine the longest interval in the box. How long is it?

   Answer: The longest interval is between Q3 and the maximum.
   $$79 - 68.1 = 10.9$$

   b. Which interval do you think has the most data points? Explain your reasoning, and then check your answer by selecting each interval.

   Sample Answers: They all contain the same number because each interval should contain 25% of the data. If there are 28 points, you would expect to see seven in each interval. In order to check your answer, you have to go back and forth between the boxplot and the dot plot because the points are so close together. For example, from the median to Q3 in the boxplot, it appears there are only six points, but the dot plot shows seven; in the upper quartile on the boxplot, there seem to be only six as well, but the dot plot shows all seven of the points.

   Teacher Tip: Exceptions can occur in other distributions. Examples include 1) the number of data is not divisible by four or 2) duplicate data values near a quartile.

3. Consider the dot plot in the lower work area on the same page.

   a. Describe the center, shape, and spread of the distribution of heights.

      Answer: This is a right-skewed distribution with a center of approximately 63, with the data
      more spread out as the heights increase. Typical values fall between 60 and 68.

      > **Teacher Tip:** Median is an acceptable measure of center. Interquartile
      > range or range is an acceptable student measure of spread.

   b. How is the shape of the distribution visible in the boxplot?

      Sample Answers: The skewness in the distribution is visible because the interval lengths
      generally increase from left to right.

   > **TI-Nspire Navigator Opportunity:** *Quick Poll*
   > **See Note 1 at the end of this lesson.**

      > **Teacher Tip:** From the boxplot, you do not know how the other heights in
      > the interval are distributed, so the most you can say is that the given
      > height is above Q1. Be sure this is clear to students.

4. a. Find the person with a height of 62.5 inches. Use the boxplot to describe how this person's
      height relates to the distribution as a whole.

   b. Compare the conclusions you can draw about the distribution of heights by looking at just the
      boxplot and the conclusions you can draw by looking at just the dot plot.

      Sample Answers: Any statement about the quartiles or median can be made from the boxplot;
      for example, about half of the heights are between 60 and 68 inches or half of the people were
      taller than about 63 inches. Just by inspection you cannot make such statements from the dot
      plot. You can make statements about individual points from the dot plot but not from the boxplot.

      > **Teacher Tip:** You might ask students to make a statement about the
      > distribution of heights from looking at the boxplot that you cannot make
      > from looking at the dot plot and vice versa.

5. You can move the points in the plots.
   a. Select on the point in the dot plot representing the largest height. Note that when a point is selected, it will be shaded. Move that point, and describe what happens.

      Answer: The corresponding point in the boxplot appears and is shaded. When you move the point in the dot plot, the point in the boxplot moves as well.

      > **Teacher Tip:** Be sure students undo the move by selecting the reset arrow to return to the original data and plots.

   b. Grab and drag the maximum point to the right, creating an outlier. At what value does it become an outlier?

      Sample Answers: At approximately 80, the value breaks away from the whisker.

   c. Explain how you can check mathematically to see if it is really an outlier.

      Sample Answers: The definition for an outlier in a boxplot is any value that is greater than the upper quartile plus 1.5 times the interquartile range or that is less than the lower quartile minus 1.5 times the interquartile range. Here, 68.1 + 1.5(68.1 – 59.9) = 68.1 + 12.3 = 80.4. So any value above 80.4 will be an outlier.

      > **Teacher Tip:** Probe student thinking about outliers by having them keep the minimum and maximum unchanged and figure out how to drag points in the dot plot so that the maximum becomes an outlier. If a data point is greater than 1.5 times the IQR, then that value is an outlier. It's easy to "make an outlier" as in Question 5, but students need to THINK about the 1.5 IQR rule to see that changing the IQR is another way to produce an outlier. This should be discussed with the class.

6. Return to the original plot by selecting the up arrow.
   a. Select the points in the dot plot representing the three largest heights. Grab one point, and move it. Describe what happens in the two plots.

      Sample Answers: The corresponding points in the boxplot appear and are shaded. When you move one of the points, the entire set of selected points moves as a group in both plots**.**

   b. Select the upper whisker in the boxplot. Move one of the points in that interval, and describe what happens in the two plots.

      Sample Answers: All of the points in the upper whisker appear and are shaded. When one of the shaded points in either of the plots is moved, all of the points in both plots move as a group.

7. Return to the original plot. Keep the median unchanged. Grab and drag points in the dot plot to create a symmetric boxplot.

   a. Describe your boxplot.

      Sample Answers: The median line should be more centrally located within the box. Whiskers should be approximately the same length.

   b. What did you do to get a symmetric distribution, and why?

      Sample Answers: Move data points from the right tail closer to the median line. Some points between the median and Q3 might need to be moved as well. Since each interval represents the same number of points, the goal was to equalize the spread within corresponding intervals.

   TI-Nspire Navigator Opportunity: *Class Capture*
   See Note 2 at the end of this lesson.

   > **Teacher Tip:** Students might want to move all of the points to the left, but since the median is required to stay fixed, that strategy will not work in this problem.

8. Keep the same minimum and maximum values. Grab and drag points in the dot plot to create a left-skewed distribution.

   a. Sketch your boxplot below, and describe how your process was different from the process you used in Question 7. If all you had were the boxplot, how would you describe the heights of the class?
      Sample Answers: More points are moved from left to right, and most of the points were moved from the middle regions.

   b. How many data values would you expect to find in the interval covered by the box? Explain your reasoning, and then check your answer.

      Sample Answers: There are 28 data values all together, and half of them, or 14, should be in the box part of the plot. The actual number of values might be different if some of the data values are equal.

      > **Teacher Tip:** Students might want to move all of the points to the right, but since the minimum is required to stay fixed, that strategy will not work in this problem.

9. Keep the minimum, Q1, Q3, and maximum unchanged. Describe how you would grab and drag points so the median line is closer to Q3 than to Q1. Check your answer using the two plots.

   Sample Answers: Strategies will probably differ. One might be to move two or three points from the 2nd interval (between the lower quartile and the median) into the 3rd interval. The median is the average of the 14th and 15th data points, so the strategy should be based on moving those points to the right.

   > **Teacher Tip:** Since the median is the only member of the five-number summary that is permitted to change, students will discover that points near the other quartiles should <u>not</u> be moved.

10. a. Describe how you would move points to change the graph so that the left whisker gets absorbed into the box.

   Sample Answers: One strategy is to stack the bottom seven data points at quartile one. Answers will vary, but the whisker is defined by the lower 25% of the data points. Therefore, eliminate the distance between the 1st and 8th point.

   b. Write a sentence about the distribution that you can surmise from the boxplot but not from the dot plot.

   Sample Answers: Answers will vary depending on how the data points are arranged. The answers, however, should only refer to some description of the intervals and the percents they represent.

   c. Write a sentence about the distribution that you can surmise from the dot plot but not from the boxplot.

   Sample Answers: Answers will vary depending on how the data points are arranged but could mention gaps or clusters that cannot be seen from just looking at intervals represented by the quartiles.

11. If you observe an interval in a boxplot that is noticeably shorter than the other intervals, what does this indicate about the corresponding region in a dot plot? Explain why your dot plots are reasonable.

   Sample Answers: There is less variability in this interval, so the dot plot would be taller. About the same number of points (depending on multiple points, etc.) are packed more closely together, so the interval that contains the points is shorter.

12. Suppose you had additional pages in your document with the following boxplots. Sketch a possible dot plot for each and describe the shape, center, and spread. Identify a set of data you think is represented by the boxplot or dot plot.

a.
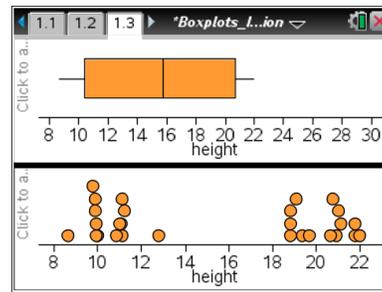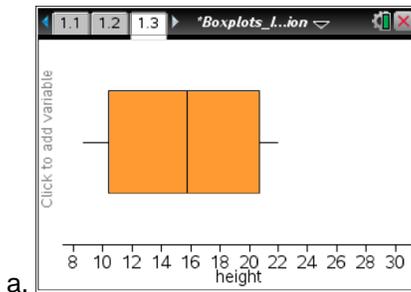
Sample Answers: Center around 16, symmetric, bimodal, range appears to be 14. Identification of data will vary. Possible answers might be heights of dogs in inches, or lengths of picture frames, or any collection that might fit the descriptions summarized above.

> **Teacher Tip:** Students' dot plots might not have gaps but should have a center lower than the tails.

b.
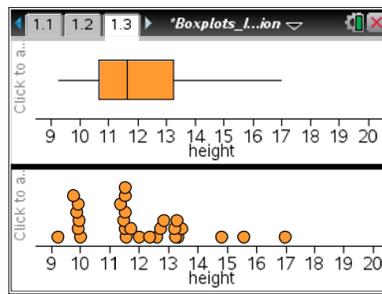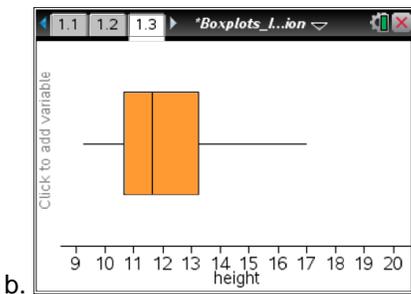
Sample Answers: Center around 11, right skewed, unimodal, IQR about 3. Identification of data will vary.

> **Teacher Tip:** IQR is a more appropriate measure of spread, since this distribution is unimodal.

c.
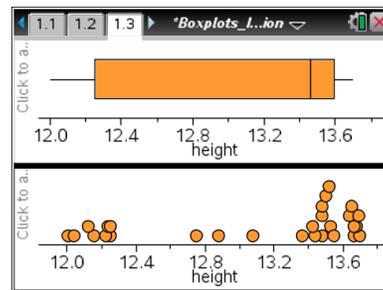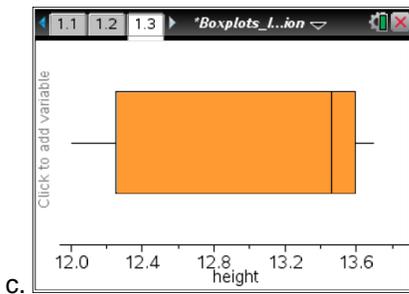
Sample Answers: Median is around 13.5, bimodal, non-symmetric. Since the median is so close to the maximum, 50% of the data are near 13.5. Another 25% are around 12.25. Identification of data will vary.

> **Teacher Tip:** You might also have students describe possible histograms for each, getting students to notice that "tall" parts of histograms correspond to "short" intervals in boxplots.

Wrap Up

Upon completion of the discussion, the teacher should ensure that students are able to understand:
- Moving values across quartiles can change the five-number summary.
- Each interval of a boxplot contains approximately 25% of all the data.
- Lengths of intervals in boxplots are inversely related to the density of the data in those intervals.

TI-Nspire Navigator

Note 1

Question 3a, *Quick Poll*

This is a good point to send a Quick Poll to students about the different measures.
- Which is an acceptable measure of spread of the data: mean, median, or range?
- Which is a better measure of center in this case: mean, median, mode, or range?

Note 2

Question 7, *Class Capture*

You might want to collect some class captures to ensure students are following directions and have correctly found a symmetric distribution. You might also want to have students show their work and discuss their answers to question 7b.