Lesson 15: More Practice with Box Plots

Student Outcomes

- Given a box plot, students summarize the data by the 5-number summary (Min, Q1, Median, Q3, Max.)
- Students describe a set of data using the 5-number summary and the interquartile range.
- Students construct a box plot from a 5-number summary.

Lesson Overview

In this lesson, students are expected to summarize and describe distributions. They consider data displayed in dot plots and box plots and summarize data sets in relation to their context, describing center and spread using 5-number summaries (Minimum, Q1, Median, Q3, and Maximum) and the interquartile range. The questions in this lesson focus students on how box plots provide information about variability in a distribution.

Students begin by looking at a box plot, finding the 5-number summary, and using the 5-number summary to describe the data. They then consider the variability in two different data sets, the maximum speeds of selected birds and of land animals, with very different spreads. They create box plots using the 5-number summary for each set. In the last example, students interpret the IQR for different data sets.

To help students make sense of box plots and to confront typical misconceptions they may have, it would be very valuable to engage students with an interactive dynamic file that allows them to explore the relation between box plots and dot plots. One such example is the activity *Introduction to Boxplots* available at http://education.ti.com/calculators/timathnspired/US/Activities/?sa=5026&t=1190.

The file can be played using TI™-Nspire handhelds, TI™-Nspire software or on the TI™-Nspire Player, which is free and can be downloaded from http://education.ti.com/calculators/products/US/document-player/

Classwork

You reach into a jar of Tootsie Pops. How many Tootsie Pops do you think you could hold in one hand? Do you think the number you could hold is greater than or less than what other students can hold? Is the number you could hold a typical number of Tootsie Pops? This lesson examines these questions.

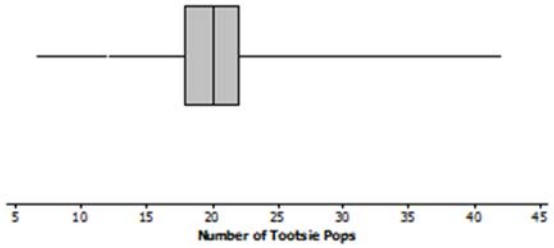
Example 1 (2–3 minutes): Tootsie Pops

You may want to actually do this experiment with students. Have them see how many Tootsie Pops they can grab and then replace the data in the example with the class data. (The data in the example were not collected from sixth grade students, so results might be different as hand sizes might be larger as students get older.) In later grades, data from the size of hand spans could be used to see whether any correlation exists between hand span and the number of Tootsie Pops someone can hold.

Example 1: Tootsie Pops

As you learned earlier, the five numbers that you need to make a box plot are the minimum, the lower quartile, the median, the upper quartile, and the maximum. These numbers are called the 5-number summary of the data.

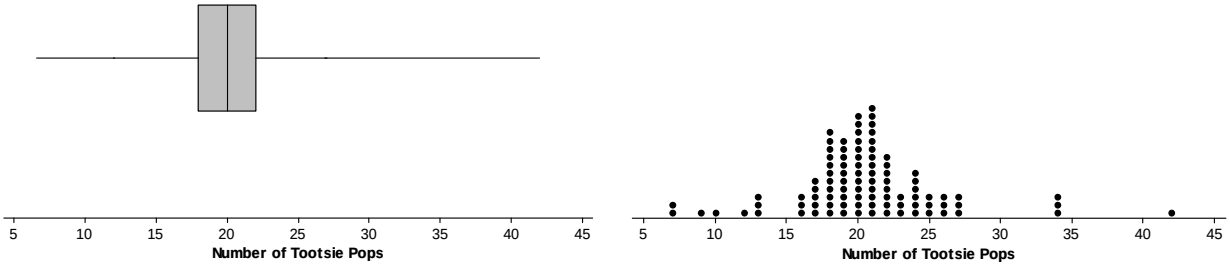
Ninety-four people were asked to grab as many Tootsie Pops as they could hold. Here is a box plot for these data. Are you surprised?



Exercises 1–5 (6–8 minutes)

As students work with the exercises, they should recognize that at least one person was able to hold a lot of Tootsie Pops because the upper segment extends to 42. They should also note that the typical number was pretty close to the median; the box that contains about half of the values only spans about four numbers. The second plot of the data informally introduces the concept of outliers as data values that seem far away from all of the others; you may want to use the term without giving a formal definition.

Technology can make the transition very visible from a dot plot to a box plot.



Note: An example on software that provides this visual transition is TI™ Nspire software. With this software, the dot plot is changed to a box plot as the dots "climb" into their place in the box plot. Going back and forth between plots several times will give students a very visual impression of what a box plot represents. Being able to simultaneously look at the two plots allows teachers to ask, "Why is the segment on the right so long? What do you think would happen if the point at 42 were removed or had been a 35?" and other questions that probe at student understanding. Interactive dynamic technology allows students to make conjectures and actually test them out by moving points and observing the consequences (See the *Mathematical Education of Teachers, Edition 2* from the Conference Board of Mathematical Sciences).

As students work through the exercises in small groups, ask them the following questions:

- How many Tootsie Pops do you think people can hold in one hand? Make a prediction.
  - Record students' estimates for this question. If possible, demonstrate for students.

- How do you find the upper and lower quartiles?
  - *Summarize with students the process they addressed previously as they order the data, find the median of the ordered data, and then find the middle of the upper half and the lower half as the upper and lower quartiles.*
- What do we mean by a 5-number summary?
  - *The 5-number summary refers to the following: the lowest or minimum data value, the lower quartile (Q1), the median, the upper quartile (Q3), and the maximum data value.*
- About what fraction of the data values should be in each section of the box plot?
  - *Approximately $\frac{1}{4}$ or 0.25 or 25% should be found in each section.*

Exercises 1–5

1. **What might explain the variability in how many Tootsie Pops those 94 people were able to hold?**

   Answers will vary: Size of people's hands, hand span, whether they are flexible in moving their fingers.

2. **Estimate the values in the 5-number summary from the box plot.**

   Min = 7, Q1 = 18, Median = 20, Q3 = 22, Max = 42

MP.6

3. **Describe how the box plot can help you understand the difference in the number of Tootsie Pops people could hold.**

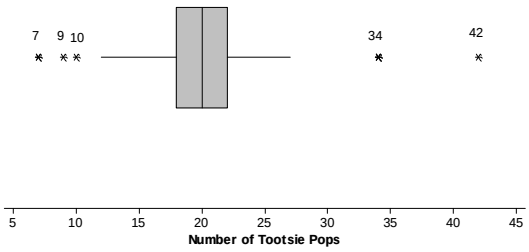
   The maximum of about 42 and minimum of about 7 show you the range of 35 Tootsie Pops. The box shows that about half of the people can hold about 2 more or 2 fewer Tootsie Pops than the median, which was 20 Tootsie Pops. The box plot shows the overall spread, the bottom half, and the middle half of the number of Tootsie Pops people can hold.

MP.3

4. **Here is Jayne's description of what she sees in the plot. Do you agree or disagree with her description? Explain your reasoning.**
   "One person could hold as many as 42 Tootsie Pops. The number of Tootsie Pops people could hold was really different and spread about equally from 7 to 42. About one half of the people could hold more than 20 Tootsie Pops."

   You cannot really tell that they are evenly spread – the box contains about half of the numbers of Tootsie Pops. However, the box is only four units long. That means half of the people were bunched over those four numbers.

5. **Here is a different plot of the same data on the number of Tootsie Pops 94 people could hold.**



   a. **Why do you suppose the five values are separate points and are labeled?**

      Maybe because they are far away from most of the other values. It shows that more than half of the data is from about 12 to 27 Tootsie Pops.

b. Does knowing these data values change anything about your responses to Exercises 1 to 4 above?

Not really, except maybe to say that only two of the people could hold a 34 and 42 Tootsie Pops; the rest held less than that.

Exercises 6–10 (15 minutes): Maximum Speeds

The intention in these exercises is not to compare the two data sets but rather to think about how the variability is different for birds and land animals. Note that two of the speeds are accurate to the hundredths place; the speed for the horse could have been clocked at a race track, but it is not clear how researchers were able to record such an accurate speed for the hummingbird. If you prefer, you could round the values for your students.

In working with any data set, a good habit is to start by looking over the values to see what might be unusual, different, or in some way interesting, which is the reason for the first question. When students describe the plots, encourage them to use fractions or percentages to talk about each of the four sections rather than "most" or "lots", i.e., $\frac{1}{4}$ or 25% of the speeds were less than 76 mph.

Exercises 6–10: Maximum Speeds

The maximum speeds of selected birds and land animals are given in the tables below.

Bird	Speed (mph)
Peregrine falcon	242
Swift bird	120
Spine-tailed swift	106
White-throated needletail	105
Eurasian hobby	100
Pigeon	100
Frigate bird	95
Spur-winged goose	88
Red-breasted merganser	80
Canvasback duck	72
Anna's Hummingbird	61.06
Ostrich	60

Land Animal	Speed (mph)
Cheetah	75
Free-tailed bat (in flight)	60
Pronghorn antelope	55
Lion	50
Wildebeest	50
Jackrabbit	44
African wild dog	44
Kangaroo	45
Horse	43.97
Thomson's gazelle	43
Greyhound	43
Coyote	40
Mule deer	35
Grizzly bear	30
Cat	30
Elephant	25
Pig	9

Data Source: Natural History Magazine, March 1974, copyright 1974; The American Museum of Natural History; and James G. Doherty, general curator, The Wildlife Conservation Society; http://www.thetravelalmanac.com/lists/animals-speed.htm; http://en.wikipedia.org/wiki/Fastest_animals

As students answer the exercises, ask the following questions individually or in small groups to help students connect their work to the outcomes:

- The top recorded speed for a human is 27.79 mph for Usain Bolt during a 100-meter sprint in 2009. How does the human compare to the other land animals?
  - *One of the fastest human speeds would be similar to the fastest speeds of elephants and wild cats.*

6. As you look at the speeds, what strikes you as interesting?

Some might suggest birds are really fast, especially the falcon. Others may notice that only two of the speeds have decimals. The speeds of specific animals might strike students as interesting.

7. Do birds or land animals seem to have the greatest variability in speeds? Explain your reasoning.

It looks like the speeds of the birds vary a lot as they go from 60 mph for some birds to 242 mph for others. The speeds of the land animals vary, but not as much; they go from 9 mph to 75 mph.

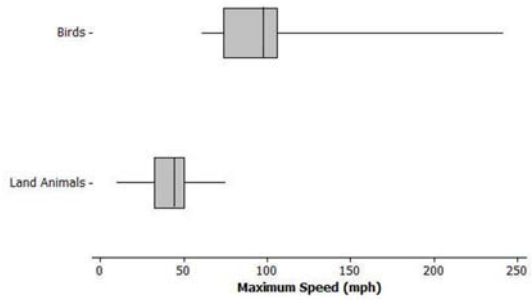
8. Find the 5-number summary for the speeds in each data set. What do the 5-number summaries tell you about the distribution of speeds for each data set?

Land Animal 5-number summary – Min = 9, Q1 = 32.5, Median = 43.97, Q3 = 50, Max = 75

Bird 5-number summary – Min = 60, Q1 = 76, Median = 97.5, Q3 = 105.5, Max = 242

The summaries give me a sense of the range or span of the speeds (Maximum – minimum speed) and how the speeds are grouped around the median.

9. Use the 5-number summaries to make a box plot for each of the two data sets.



10. Write several sentences to tell someone about the speeds of birds and land animals.

At least one bird flies really fast, the falcon at 242 mph. Three fourths of the birds fly less than 106 mph, and the slowest bird flies at 60 mph. The land animals' running speeds are slower going from 9 mph to 75 mph. The middle half of the speeds for land animals is between 32.5 mph and 50 mph.

Exercises 11–15: What is the Same and What is Different? (10 minutes)

The focus in thinking about the three box plots should be on the IQR for each, noting that the minimum, median, and maximum for each plot are the same. The spread of the middle half of the data is across the entire range (minimum to maximum) for the second plot. This could happen because the distribution is bimodal with the lower quartile and the minimum the same value and the upper quartile and the maximum the same value. The spread of the middle half of the data is much more tightly packed around the median in the third plot. Students estimate the quartiles from the plots; if their answers vary a bit that is okay because the emphasis is on the concept.

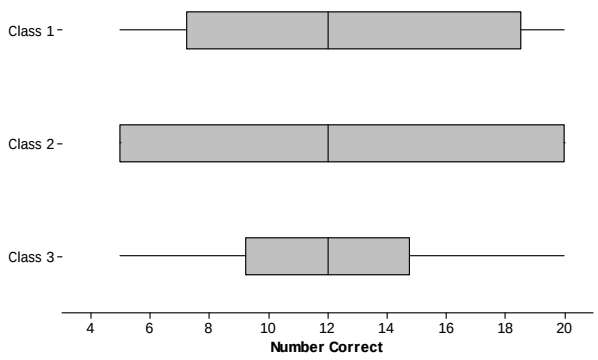
As students answer the questions for this exercise, ask the following questions individually or in small groups to help students connect their work to the outcomes:

- Are the IQR and the range enough to make a box plot? If not, what else do you need to know?
  - *No. We need to also know the median.*

- How many data values will be between the quartiles? Will there always be the same number in each quartile? Why or why not?
  - *Four data values are within each quartile. The number of data values within the quartiles will generally be the same as boundaries may include a data value. Consider discussing with students how many animals would be within each quartile if there were 18 animals or 19 animals or 20 animals. If one of the data values is the median, or two of the data values are the first quartile (Q1) and the third quartile (Q3), the values within the quartiles are the same. There are exceptions that students begin to see with the following examples.*

Exercises 11–15: What is the Same and What is Different?

Consider the following box plots, which show the number of questions different students in three different classes got correct on a 20-question quiz.



11. Describe the variability in the scores of the three classes.

The range (Max–Min) is the same for all three classes and so is the median, but the boxes that contain the middle half of the scores are spread very differently about the median. The third class has a small box so the scores are close together. In Class 2 the minimum and lower quartile are the same score, and the maximum and upper quartile are also the same score, so lots of scores are piled at the ends of the range. The middle half of the scores in Class 1 are spread out more than Class 3 but not as much as Class 2.

12.

a. Estimate the interquartile range for each of the three sets of scores.

Class 1 IQR = 10; Class 2 IQR = 15; Class 3 IQR = 5

b. What fraction of students does the interquartile range represent?

About one half.

c. What does the value of the IQR tell you about how the scores are distributed?

For Class 1, half of the scores are spread over an interval of width 10; for Class 2, half of the scores are spread out over an interval of width 15; and for Class 3, half of the scores are bunched together over an interval of width 5.

13. The teacher asked students to draw a box plot with a minimum value at 34 and a maximum value at 64 that had an interquartile range of 10. Jeremy said he could not draw just one because he did not know where to put the box on the number line. Do you agree with Jeremy? Why or why not?

The box could go anywhere from 34 to 44 all the way to from 54 to 64 and any width of 10 in between so Jeremy is correct.

14. Which class do you believe performed the best? Be sure to use the data from the box plots to back up your answer.

Class 3 as it has the smallest IQR. About half of the students scored close to the median score. Scores were more consistent for this class. Students may select Class 3 based on the smallest variability. Students might also make a case that although the variability is greater, approximately 25% of the students in Class 1 scored 18 or higher compared to 25% of the students in Class 3 scored 15 or higher. In Class 2, several students must have scored near the top for the Q3 and maximum to be the same. (Allow students to select the box plot they think answers the question and to describe why they selected the box plot.)

15.

a. Find the IQR for the three data sets in the first two examples: maximum speed of birds, maximum speed of land animals, and number of Tootsie Pops.

Land Animals: $50 - 32.5$ for an IQR of 17.5

Birds: $105.5 - 76$ for an IQR of 29.5

Tootsie Pops: $22 - 18$ for an IQR of 4

b. Which data set had the highest percentage of data values between the lower quartile and the upper quartile? Explain your thinking.

All of the data sets should have about half of the data values between the quartiles.

Closing (2 minutes)

Lesson Summary

In this lesson, you learned about the 5-number summary for a set of data: minimum, lower quartile, median, upper quartile, and maximum. You made box plots after finding the 5-number summary for two sets of data (speeds of birds and land animals), and you estimated the 5-number summary from box plots (number of Tootsie Pops people can hold, class scores). You also found the interquartile range (IQR), which is the difference between the upper quartile and lower quartile. The IQR, the length of the box in the box plot, indicates how closely the middle half of the data is bunched around the median. (Note that because sometimes data values repeat and the same numerical value may fall in two sections of the plot, it is not always exactly half. This happened with the two speeds of 50 mph – one went into the top quarter of the data and the other into the third quarter – the upper quartile was 50.)

You also practiced describing a set of data using the 5-number summary, making sure to be as precise as possible-avoiding words like "a lot" and "most" and instead saying about one half or three fourths.

Exit Ticket (3 minutes)

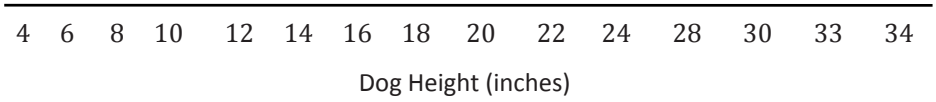
Name _____ Date_____

Lesson 15: More Practice with Box Plots

Exit Ticket

Given the following information, create a box plot and find the IQR.

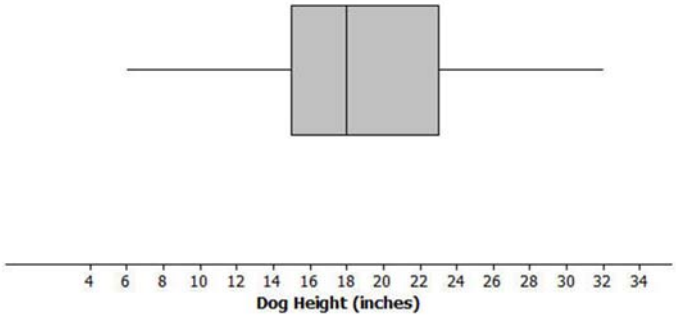
For a large group of dogs, the shortest dog was 6 inches, and the tallest was 32 inches. One half of the dogs were taller than 18 inches. One fourth of the dogs were shorter than 15 inches. The upper quartile of the dog heights was 23 inches.



Exit Ticket Sample Solutions

Given the following information, create a box plot and find the IQR.

For a large group of dogs, the shortest dog was 6 inches, and the tallest was 32 inches. One half of the dogs were taller than 18 inches. One fourth of the dogs were shorter than 15 inches. The upper quartile of the dog heights was 23 inches.

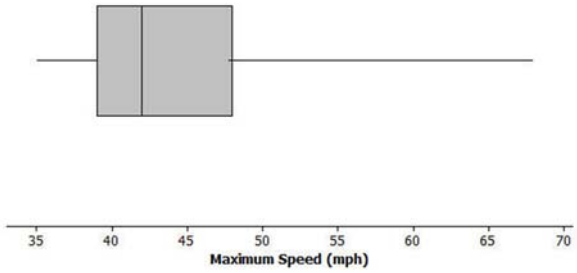


The IQR is $23 - 15 = 8$.

Problem Set Sample Solutions

All students should be encouraged to do problems 1 and 2 to be sure they understand the concepts developed in the lesson. Problem 4b should be discussed in some way as a whole class to raise awareness that medians are about counts and relative position of ordered data and not about distance or location.

1. The box plot below summarizes the maximum speeds of certain kinds of fish.



a. Estimate the 5-number summary from the box plot.

Answers will vary: Min – 35 mph; Q1 – 39 mph; Median – 42 mph; Q3 – 48 mph; Max – 68 mph.

b. The fastest fish is the sailfish at 68 mph followed by the marlin at 50 mph. What does this tell you about the spread of the fish speeds in the top quarter of the plot?

The Q3 is about at 48, so all but one of the top quarters are bunched between 48 and 50 mph.

c. Use the 5-number summary and the IQR to describe the speeds of the fish.

The speeds of fish vary from 35 mph to 68 mph. The IQR is 9 mph; the middle half of the speeds is between 39 mph and 48 mph. Half of the speeds are less than 42 mph.

Note: Data for box plot is provided below.

Fish	Maximum speed (mph)
Sailfish	68
Marlin	50
Wahoo	48
Tunny	46
Bluefin tuna	44
Great blue shark	43
Bonefish	40
Swordfish	40
Bonito	40
Four-winged flying fish	35
Tarpon	35

Data Source: http://www.thetravelalmanac.com/lists/fish-speed.htm

2. **Suppose you knew that the interquartile range for the number of hours students spent playing video games during the school week was 10. What do you think about each of the following statements? Explain your reasoning.**

   a. **About half of the students played video games for 10 hours during a school week.**

   This may not be correct as you know the width of the interval was 10, but you do not know where it starts or stops. You do not know the lower or upper quartile.

   b. **All of the students played at least 10 hours of video games during the school week.**

   This may not be correct for the same reason as in part (a).

   c. **About half of the class could have played video games from 10 to 20 hours a week or from 15 to 25 hours.**

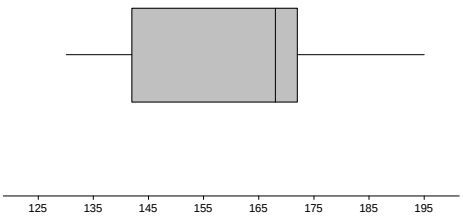
   Either could be correct as the only information you have is the width of 10, and the statement says "could" not "is".

3. **Suppose you know the following for a data set: minimum value is 130, the lower quartile is 142, the IQR is 30, half of the data are less than 168, and the maximum value is 195.**

   a. **Think of a context for which these numbers might make sense.**

   Answers will vary: The number of calories in a serving of fruit.

   b. **Sketch a box plot.**



   c. **Are there more data values above or below the median? Explain your reasoning.**

   The number of data values on either side of the median should be about the same, one half of all of the data.

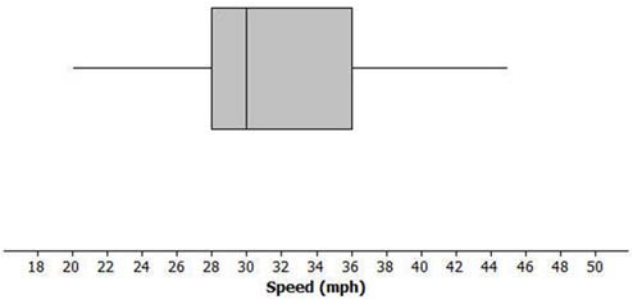
4. The speeds for the fastest dogs are given in the table below.

Breed	Speed (mph)
Greyhound	45
African Wild Dog	44
Saluki	43
Whippet	36
Basanji	35
German Shepherd	32
Vizsla	32
Doberman Pinscher	30

Breed	Speed (mph)
Irish Wolfhound	30
Dalmatian	30
Border Collie	30
Alaskan Husky	28
Giant Schnauzer	28
Jack Russell Terrier	25
Australian Cattle Dog	20

Data Source: http://www.vetstreet.com/our-pet-experts/meet-eight-of-the-fastest-dogs-on-the-planet; http://canidaepetfood.blogspot.com/2012/08/which-dog-breeds-are-fastest.html

a. Find the 5-number summary for this data set and use it to create a box plot of the speeds.



Min = 20, Q1 = 28, Median = 30, Q3 = 36, Max = 45

b. Why is the median not in the center of the box?

The median is not in the center of the box because about $\frac{1}{4}$ of the speeds are between 30 and 36, and another $\frac{1}{4}$ are closer together, between 28 and 30. The data are skewed with lots of them at the lower speeds.

c. Write a few sentences telling your brother or sister about the speed of the fastest dogs.

Half of the dogs run faster than 30 mph; the fastest dog in the list is the greyhound with a speed of 45 mph. The slowest dog in the list is the Australian cattle dog. The middle 50% of the speeds are between 28 and 36 mph.