Lesson 14: Summarizing a Distribution Using a Box Plot

Student Outcome

- Students construct a box plot from a given set of data.

Lesson Overview

In this lesson, students transition from using dot plots to display data to using box plots. The lesson begins with exercises that lay the foundation for the development of a box plot. Students inspect dot plots of several sets of data and think about how to group or section the plots to get a sense of the span of data values in each of the sections. When individual students determine how to make the sections, the results differ and the process seems arbitrary and inconsistent. Thus, there is a need for a standard procedure for making a box plot. Using the median and the quartiles introduced in the previous lesson seems like a good strategy.

The lesson begins and ends with interactive activities. If time allows, students will create a "human box plot" of the time it took them to come to school. Supplies will be needed for this exercise and are listed in the teacher notes.

Classwork

A box plot is a graph that is used to summarize a data distribution. What does the box plot tell us about the data distribution? How does the box plot indicate the variability of the data distribution?

Example 1 (5–7 minutes): Time to Get to School

MP.3

The questions are designed to help students begin to think about grouping data in order to get a sense of the spread of the data values within sections of the data. Let students write an estimate of the time it took them to come to school on a post-it note. Teachers may want the individual students to place their post-it note on a dot plot that is displayed on the classroom board at the beginning of class or graph the dot plot as a class.

Example 1: Time to Get to School

What is the typical amount of time it takes for a person in your class to get to school? The amount of time it takes to get to school in the morning varies for each person in your class. Take a minute to answer the following questions. Your class will use this information to create a dot plot.

Write your name and an estimate of the number of minutes it took you to get to school today on a post-it note.

Answers will vary.

What were some of the things you had to think about when you made your estimate?

Answers will vary: Does it count when you have to wait in the car for your sister? Usually I walk, but today I got a ride. Does it matter that we had to go a different way because the road was closed? The bus was late.

As students discuss and complete the questions in this example; additionally, ask the following questions:

- What does a dot on the dot plot represent?
- What is an estimate of the median? What is the typical amount of time it takes someone to get to school?
- What are the minimum and maximum values?

Exercises 1–4 (7–10 minutes)

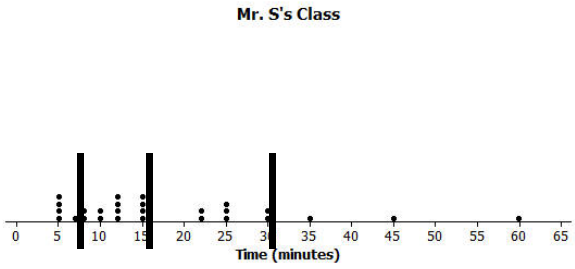
Let students work individually on the exercises and compare their plots with their neighbors. Students should recognize that their divisions of the data into four sections are close but not the same.

If time permits, bring them together for a discussion of their answers and stress the idea that it is useful to see how the data values group together in intervals across the entire distribution. Ask the students the following questions:

- Are there a lot of values in the middle or at one end?
- What was the shortest time a student in Mr. S's class got to school? The longest?
- Looking at the plot, how long does it seem to take a typical student in Mr. S's class to get to school?

Exercises 1–4

Here is a dot plot of the estimates of the times it took students in Mr. S's class to get to school one morning.



1. **Put a line in the dot plot that seems to separate the shortest times and the longest times.**

   Some might put the dividing line between 15 and 20.

2. **Put another line in the plot that separates those who seem to live really close to school and one that marks off those who took a long time to get to school.**

   Responses will be different. Some might put a line at 30 and a line at 10.

3. **Your plot should be divided into four sections. Record the number of times in each of the four sections.**

   Answers will vary: Depending on the divisions, 7 or 8 in the lower one, 9 in the next, 5 in the next, and 5 in the upper section.

4. **Share your marked up dot plot with some of your classmates. Compare how each of you divided the plot into four sections.**

   Different responses; students should recognize that the divisions might be close but are different.

Exercises 5–7 (10 minutes): Time to Get to School

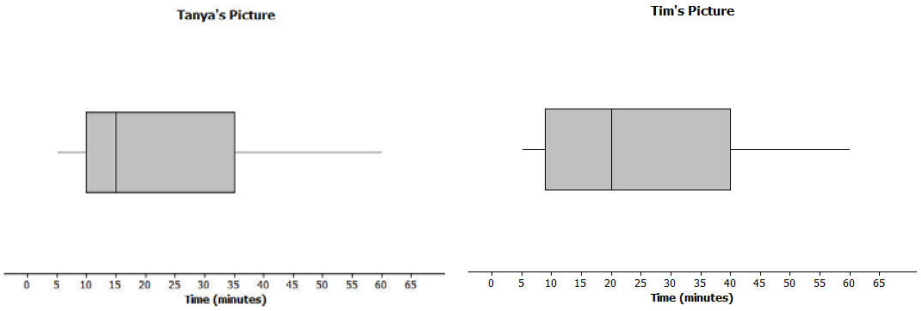
Let students work individually on the questions. Then discuss and confirm answers. Start to get students thinking about using the 5-*number summary* (or the *minimum*, lower quartile or *Q1*, *median*, upper quartile or *Q3*, and the *maximum*) and how it relates to the construction of the box plot. Ask students the following questions:

- How could the quartiles and the median be helpful in making a plot that summarized the data?
- Where do you think the median is on Tanya's and Tim's plots?
- Where are the lower and upper quartiles? (Q1 and Q3)?
- Why are their values different?

Exercises 5–7: Time to Get to School

The teacher asked the class to make a representation that would summarize the times it took students in Mr. S's class to get to school and how they are spread out. Tim decided to get rid of the dots and just use a picture of the divisions he made of the shortest times and the longest times. He put a box around the two middle sections.

Tanya thought that was a good idea and made a picture of the way she had divided the times. Here are their pictures.



5. **What do the pictures tell you about the length of time it takes the students to get to school?**

   Answers will vary: The fastest times for getting to school were from 3 or 4 minutes to 10 minutes, and the longest time was around 62 minutes. A bunch of students took from 10 to 35 or 40 minutes to get to school. It kind of looks like most students live closer to school – or at least got to school in a short time.

6. **What don't the pictures tell you about the length of time it takes the students to get to school?**

   The pictures don't tell how many students were in the class or even how many students were in any of the sections.

7. **How do the two pictures compare?**

   The pictures are pretty close, but Tim had a longer box with a group from 10 to 20, and Tanya had a shorter box with a group from 10 to 15. Their next groups were different too.

Example 2 (7–10 minutes): Making a Box Plot

This example defines the procedure for finding a box plot, referring back to Lesson 13 on quartiles and the IQR. You may want students to read through the process themselves and then ask them what the directions mean, or have the whole class work through the process together.

Ask the students the following when the box plot is complete:

- Why is it important to have a standard way to make a box plot?
- What does a box plot tell you about the *story* in the data? What does it not tell you?
- What proportion (percent) of the data is in each of the sections of the box plot? How do you know?

Example 2: Making a Box Plot

Mr. S suggested that to be sure everyone had the same picture, statisticians developed a standard procedure for making the cut marks for the sections.

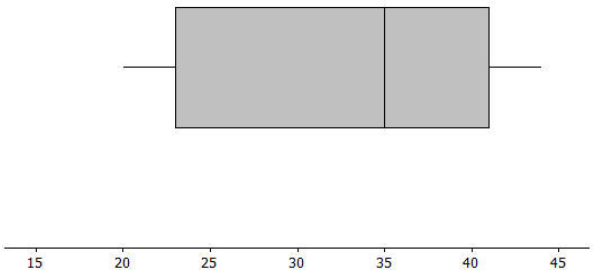
Mr. S. wrote the following on the board:

To make a box plot

- **Find the median of all of the data**
- **Find Q1, the median of the bottom half of the data, and Q3, the median of the top half of the data.**
- **Draw a box that goes from Q1 to Q3, the two middle sections.**
- **Draw a line segment connecting the minimum value to the box and one that connects the maximum value to the box.**

Now use the given number line to make a box plot of the data below.

$20, 21, 25, 31, 35, 38, 40, 42, 44$



The 5-number summary is as follows:

Min = 20
Q1 = 23
Median = 35
Q3 = 41
Max = 44

Exercises 8–11 (10–12 minutes): A Human Box Plot

Depending on the amount of time left in class, this exercise can either be completed on the board as a class or by using selected students to create a Human Box Plot. If possible, try to involve approximately 12 students. This would put 3 students in each quartile. If time is running short and the problem is completed as a class, use the focus questions listed below.

Preparation for Human Box Plot:

The data is already ordered from the post-it notes on the dot plot on the classroom board. Call out students' names to have them form an ordered line of data. Find a place in the classroom (or hall) that will allow all of the students to line up and that can accommodate a number line that will go from 0 to as large as 60 minutes. Have some props that can locate a number line – large cards with a scale in 5-minute intervals can work. Do not start with the number line, however. A ball of yarn or rope can be used to mark off the *box* part of the box plot.

Once students are in line, have them identify the median and the two quartiles. Give the signs for the five summary values to the appropriate students and ask them to step out with their signs. Ask each of the quartiles to hold one end of the rope marking off the *box* that extends from Q1 to Q3. Students may not recognize at first that the plot has no scale and thus does not really tell the story. Try to get them to see how important the scale is by asking questions such as the following:

- You all are a human box plot of the times it took you to come to school. Did it take most of you a short time or a longer time?
- Did it take anyone a really long time?
- Can you tell from our plot?

When students realize that it looks like the times were all evenly spaced because of how they are standing, bring out the props for the number line (be sure to make the intervals wide enough to accommodate several students). Then have students rearrange themselves using the scale and recreate the box plot with the five number summary values and the rope to represent the box. Then, ask the following:

- How many people are in each of the sections? Which section has the most people? The fewest?
- Were there any sections where the people were all crowded together? How did this show up in the box plot?
- Why do we need a scale to make a box plot?

Exercises 8–11: A Human Box Plot

Consider again your post-it note that you used to write down the number of minutes it takes you to get to school. If possible, you and your classmates will form a human box plot of the number of minutes it takes your class to get to school.

8. Find the median of the group. Does someone represent the median? If not, who is the closest to the median?

Answers will vary.

9. Find the maximum and minimum of the group. Who are they?

Answers will vary.

10. Find Q1 and Q3 of the group. Does anyone represent Q1 or Q3? If not, who is the closest to Q1? Who is closest to Q3?

Answers will vary.

11. Sketch the box plot for this data set below.

Answers will vary.

Closing (2–3 minutes)

Lesson Summary

The focus of this lesson is moving from a plot that shows all of the data values (dot plot) to one that summarizes the data with five points (box plot).

You learned how to make a box plot by doing the following:

- **Finding the median of all of the data**
- **Finding Q1, the median of the bottom half of the data, and Q3, the median of the top half of the data.**
- **Drawing a box that goes from Q1 to Q3, the two middle sections.**
- **Drawing a line segment connecting the minimum value to the box and one that connects the maximum value to the box.**

You also learned important characteristics of box plots:

- **$\frac{1}{4}$ of the data are in each of the sections of the plot.**
- **The length of the interval for a section does not indicate either how the data are grouped in that interval or how many values are in the interval.**

Exit Ticket (5 minutes)

Name _____ Date_____

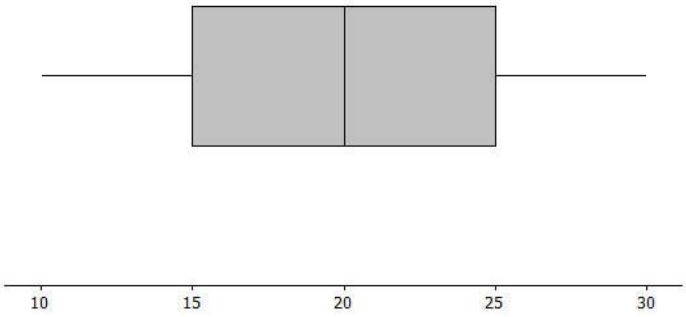
Lesson 14: Summarizing a Distribution Using a Box Plot

Exit Ticket

Sulee explained how to make a box plot to her sister as follows:

"First you find the smallest and largest values and put a mark halfway between them, and then put a mark halfway between that mark and each end. So, if 10 is the smallest value and 30 is the largest value, you would put a mark at 20. Then another mark belongs half way between 20 and 10, which would be at 15. And then one more mark belongs half way between 20 and 30, which would be at 25. Now, you put a box around the three middle marks and draw lines from the box to the smallest and largest values."

Here is her box plot. What would you say to Sulee?

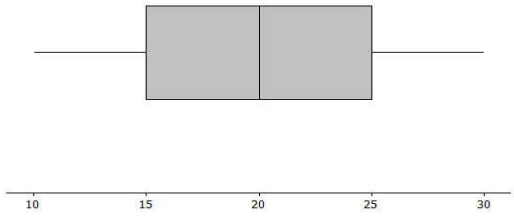


Exit Ticket Sample Solutions

Sulee explained how to make a box plot to her sister as follows:

"First you find the smallest and largest values and put a mark halfway between them, and then put a mark halfway between that mark and each end. So, if 10 is the smallest value and 30 is the largest value, you would put a mark at 20. Then another mark belongs half way between 20 and 10, which would be at 15. And then one more mark belongs half way between 20 and 30, which would be at 25. Now, you put a box around the three middle marks and draw lines from the box to the smallest and largest values."

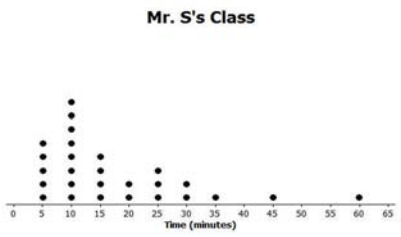
Here is her box plot. What would you say to Sulee?

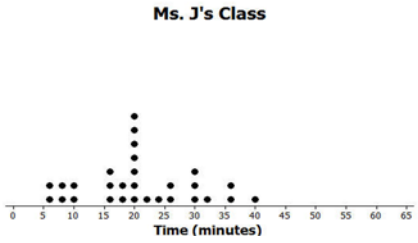


Sulee is wrong. This is not the correct way to create a box plot. Sulee did not find the median or the quartiles using the data values; she just divided up the length between the smallest and largest numbers.

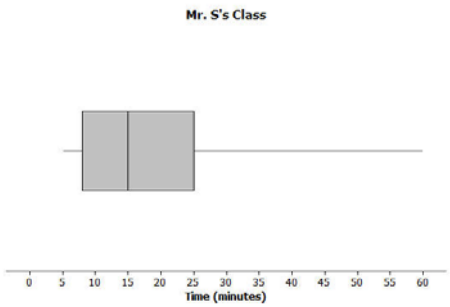
Problem Set Sample Solutions

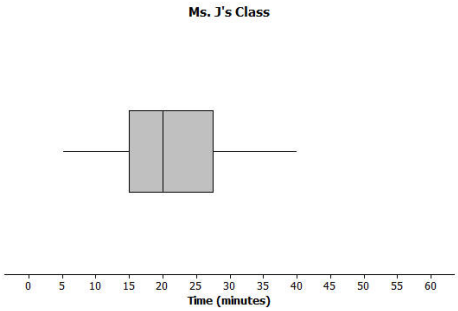
1. **Dot plots for the amount of time it took students in Mr. S's and Ms. J's classes to get to school are below.**





a. **Make a box plot of the times for each class.**



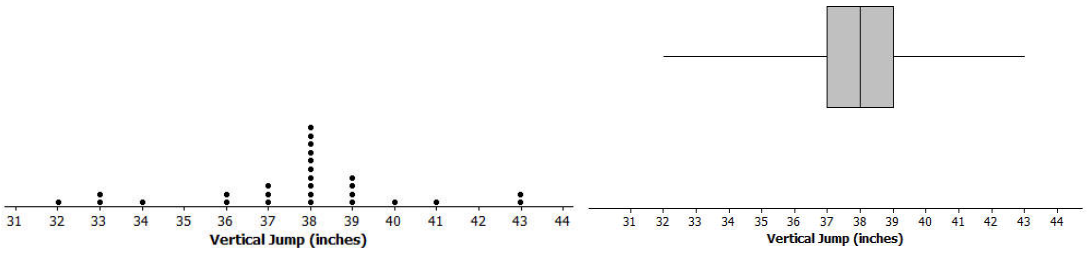


Mr. S five summary values: $5, 10, 15, 25, 60$ *and Ms. J five summary values:* $5, 15, 20, 27.5, 40$

b. What is one thing you can see in the dot plot that you cannot see in the box plot? What is something that is easier to see in the box plot than in the dot plot?

The dot plot shows individual times which you cannot see in the box plot. The box plot shows the location of the median and of the lower and upper quartiles.

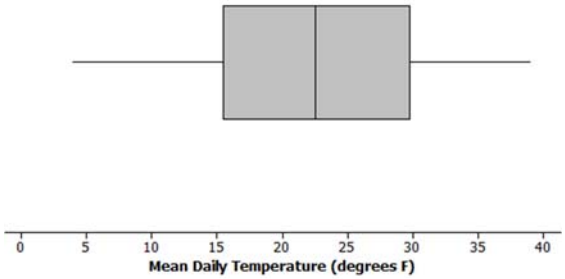
2. The dot plot below shows the vertical jump of some NBA players. A vertical jump is how high a player can jump from a standstill. Draw a box plot of the heights for the vertical jumps of the NBA players above the dot plot.



Five summary values: $32, 37, 38, 39, 43$

3. The mean daily temperatures in °F for the month of February for a certain city are as follows:
$4, 11, 14, 15, 17, 20, 30, 23, 20, 35, 35, 31, 34, 23, 15, 19, 39, 22, 15, 15, 19, 39, 22, 23, 29, 26, 29, 29$

a. Make a box plot of the temperatures.



5-summary values: $4, 16, 22.5, 29.5, 39$

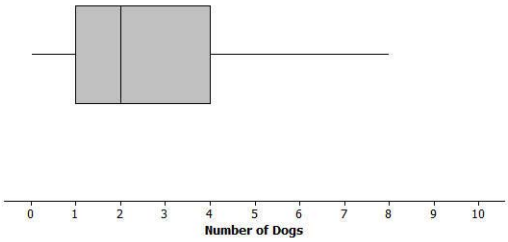
b. Make a prediction about the part of the United States you think the city might be located. Explain your reasoning.

Answers will vary: The city was probably somewhere in the northern states, either Midwest or northeast or maybe Montana or Wyoming, because the temperatures are pretty cold.

c. Describe the data distribution of temperature. Include a description of the center and spread.

The IQR is $29.5° - 16°$, *or* $13.5°$. *Half of temperatures were near the middle between* $16°$ *and* $29.5°$. *The median is* $22.5°$. *A quarter of the temperatures are less than* 16 *but greater than or equal to* $4°$. *A quarter of the temperatures are greater than* $29.5°$ *and less than or equal to* $39°$.

4. The plot below shows the results of a survey of households about the number of dogs they have. Identify the following statements as true or false. Explain your reasoning in each case.



a. The maximum number of dogs per house is 8.

True because the line segment at the top goes to 8.

b. At least $\frac{1}{2}$ of the houses have 2 or more dogs.

True because 2 is the median.

c. All of the houses have dogs.

False because the lower line segment starts at 0 so at least one household does not have a dog as a pet.

d. Half of the houses surveyed have between 2 and 4 dogs.

False because only about 25% of the houses would have between 2 and 4 dogs.

e. Most of the houses surveyed have no dogs.

False because at least $\frac{3}{4}$ of those surveyed had 1 or more dogs.