Lesson 16: Box plots

Goals

- Compare and contrast (orally) a dot plot and a box plot that represent the same data set.

- Create a box plot to represent a data set.

- Describe (orally) the parts of a box plot that correspond with each number in the five-number summary, the range, and the IQR of a data set.

Learning Targets

- I can use the five-number summary to draw a box plot.

- I know what information a box plot shows and how it is constructed.

Lesson Narrative

In this lesson, students use the five-number summary to construct a new type of data display: a **box plot**. Similar to their first encounter with the median, students are introduced to the structure of a box plot through a practical activity. Using the class data set that contains the numbers of letters in their names (from an earlier lesson), they first identify the numbers that make up the five-number summary. Then, they use their numbers to position themselves on a number line on the ground, and are guided through how a box plot would be constructed with them as the data points.

Later, students draw and make sense of the structure of a box plot on paper. They notice that, unlike the dot plot, it is not possible to know all the data points from a box plot. They understand that the box plot summarises a data set by showing the range of the data, where the middle half of the data set is located, and how the values are divided into quarters by the quartiles.

Addressing

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

- Giving quantitative measures of centre (median and/or mean) and variability (interquartile range and/or range), 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 centre and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- Collect and Display

- Clarify, Critique, Correct

- Notice and Wonder

Required Materials
Index cards
Markers
Scissors
String
Tape

Required Preparation

For the Human Box Plot activity:

- Each student will need the index card that shows their name and the number of letters in their name (used for the Finding the Middle activity), as well as a class data set.

- Compile the numbers on the cards into a single list or table. Prepare one copy of the data set for each student.

- Have some extra index cards available for students who might have been absent in that earlier lesson.

- Prepare five index cards that are labelled with "minimum," "maximum," "Q1," "Q2," and "Q3."

- Make a number line on the ground using thin masking tape (0.5 inch). It should show whole number intervals and span at least from the lowest data value to the highest. The intervals should be at least a student's shoulder's width.

- Prepare a roll of wide masking tape (2- or 3-inch wide) to create a box and two whiskers on the ground.

Student Learning Goals

Let's explore how box plots can help us summarise distributions.

16.1 Notice and Wonder: Puppy Weights

Warm Up: 5 minutes

This warm-up encourages students to analyse a data set carefully and reason about its distribution. It also gives them a chance to use the language they have acquired in this unit in an open-ended setting.

As students discuss their observations and questions with their partners, identify a few who notice or wonder about the features, centre, and spread of the distribution.

Instructional Routines

- Notice and Wonder

Launch

Arrange students in groups of 2. Display the table for all to see. Give students 1 minute of quiet time to look at the data set and identify at least one thing they notice and at least one thing they wonder about the distribution of the data. Ask students to give a signal when they have noticed or wondered about something. When the minute is up, give students 1 minute to discuss their observations and questions with their partner. Follow with a whole-class discussion.

Student Task Statement

Here are the birth weights, in ounces, of all the puppies born at a kennel in the past month.

```
13  14  15  15  16  16  16  16
17  17  17  17  17  17  17  18
18  18  18  18  18  18  18  19
20
```

What do you notice and wonder about the distribution of the puppy weights?

Student Response

Answers vary. Sample responses:

Students may notice:

- The median birth weight is 17 ounces.

- The median looks like it's the same as the mean.

- Almost a third of the puppies weigh 18 ounces.

- There are no gaps in the data or values that are very far from the centre.

- There are 25 puppies.

Students may wonder:

- If the IQR is less than 1.

- If a box plot could be used to represent the data.

- If a dot plot could be used to represent this data.

Activity Synthesis

While students may notice and wonder about many aspects of the data set, focus the discussion on observations and questions about the features, centre, and spread of the distribution. Select previously identified students to share their responses. After each person shares, ask if others noticed or wondered about the same thing.

If students commented on or wondered about the shape of the distribution, consider displaying the following dot plot for all to see and discussing questions such as:

- "Can you use the dot plot to answer what you wondered about?"

- "Did you anticipate the distribution to have this shape?"

- "What do you notice about the distribution here that you did not notice in the table?"

- "Which representation of data—the table or the dot plot—allows you see the centre and spread of the distribution more easily? Why?"

- "Which representation allows you to find the mean more easily? What about the median?"

.
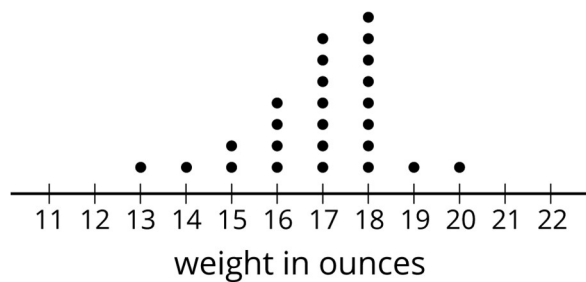
16.2 Human Box Plot

15 minutes

Previously, students learned to identify the median, quartiles, and five-number summary of data sets. They also calculated the range and interquartile range of distributions. In this activity, students rely on those experiences to make sense of **box plots**. They explore this new representation of data kinaesthetically: by creating a human box plot to represent class data on the lengths of student names, which they collected in the Finding the Middle activity in an earlier lesson.

Instructional Routines
- Collect and Display

Before the lesson, use thin painter's tape to make a number line on the ground. If the floor is tiled with equal-sized tiles, consider using the tiles for the intervals of the number line. Otherwise, mark off equal intervals on the tape. The number line should cover at least the distance between least data value (the fewest number of letters in a student's name) to the greatest (the most number of letters).

Provide each student with a copy of the data on the lengths of students' names from the Finding the Middle activity. If any students were absent then, add their names and numbers of letters to the data set.

Give students 4–5 minutes to find the quartiles and write the five-number summary of the data. Then, invite several students to share their findings and come to an agreement on the five numbers. Record and display the summary for all to see.

Explain to students that the five-number summary can be used to make another visual representation of a data set called a box plot. Tell students that they will create a human box plot in a similar fashion as when they were finding the median.

- Return to students the index cards from the lesson on finding the median. If any students were absent when the cards were made, give them each an index card and ask them to record on the card their full name and the number of letters in their name. If any student who made a card is absent, have another student with the same number of letters in their name hold the card of the absent student.

- Ask students to stand up, holding their index card in front of them, and place themselves on the point on the number line that corresponds to their number. (Consider asking students to do so without speaking at all.) Students who have the same number of letters should stand one in front of the other.

- Hold up the index cards that have been labelled with "minimum." Ask students who should claim the card, then hand the card to the appropriate student. Do the same for the other labels of a five-number summary. If any of the quartiles falls between two students' numbers, write that number of the index card and have both students hold that card together.

Now that the five numbers are identified and each associated with one or more students, use wide painter's tape to construct a box plot.

- Form a rectangle on the ground by affixing the tape around the group of students between Q1 and Q3. If a quartile is between two people, put the tape down between them. If a quartile has the value of a student's number, put the tape down at that value and have the student stand on it.

- Put a tape segment at Q2, from the top side of the rectangle to the bottom side, to subdivide the rectangle into two smaller rectangles. If Q2 is a student's number, have the student stand on the tape.

- For the left whisker, affix the end of tape to the Q1 end of the rectangle; extend it to where the student holding the "minimum" card is standing. Do the same for the right whisker, from Q3 to the maximum.

- Tape the five-number summary cards and students' cards that correspond to them in the right locations.

This image shows an example of a completed human box plot.
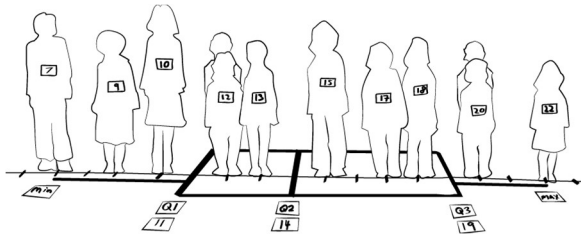
Explain to students that they have made a human box plot. Consider taking a picture of the box plot for reference and discussion later.

Action and Expression: Internalise Executive Functions. Provide students with a graphic organiser to collect the class data.
Supports accessibility for: Language; Organisation

Student Task Statement

Your teacher will give you the data on the lengths of names of students in your class. Write the five-number summary by finding the data set's minimum, Q1, Q2, Q3, and the maximum.

Pause for additional instructions from your teacher.

Student Response

Answers vary.

Activity Synthesis

Tell students that a box plot is a representation of a data set that shows the five-number summary. Discuss:

- "Where can the median be seen in the box plot?" (It is the line inside the box.) What about the first and third quartiles? (The left and right sides of the box.)

- "Where can the IQR be seen in the box plot?" (It is the length of the box.)

- "The two segments of tape on the two ends are called 'whiskers.' What do they represent?" (The lower one-fourth of the data and the upper one-fourth of data.)

- "How many people are part of the box, between Q1 and Q3? Approximately what fraction of the data set is that number?" (About half. Note that the number of

people that are part of the box may not be exactly one half of the total number of people, depending on whether the number of data points is odd or even, and depending on how the values are distributed.)

- "Why might it be helpful to summarise a data set with a box plot?" (It could help us see how close together or spread out the values are, and where they are concentrated.)

Explain to students that we will draw and analyse box plots in upcoming activities and further explore why they might be useful.

Representing, Speaking: Collect and Display. Use this to collect student language for the question: "Why might it be helpful to summarise a data set with a box plot?" Create a visual reference that highlights words such as median, quartiles, IQR, and distribution of data. This will help students use mathematical language and make sense of using a box plot to represent data.
Design Principle(s): Support sense-making; Maximise meta-awareness

16.3 Studying Blinks

15 minutes

In the last activity, students constructed a box plot based on the five-number summary of their name length data. In this activity, they learn to draw a box plot and they explore the connections between a dot plot and a box plot of the same data set.

Instructional Routines
- Clarify, Critique, Correct

Launch

Tell students that they will now draw a box plot to represent another set of data. For their background information, explain that scientists believe people blink their eyes to keep the surface of the eye moist and also to give the brain a brief rest. On average, people blink between 15 and 20 times a minute; some blink less and others blink much more.

Arrange students in groups of 2. Give 4–5 minutes of quiet work time for the first set of questions and a minute to discuss their work with their partner. Ask students to pause afterwards.

Display the box plot for all to see. Reiterate that a box plot is a way to represent the five-number summary and the overall distribution. Explain:

- "The left and right sides of the box are drawn at the first and third quartiles (Q1 and Q3)."

- "A vertical line inside the box is drawn at the median (Q2)."

- "The two horizontal lines (or 'whiskers') extend from the first quartile to the minimum and from the third quartile to the maximum."

- "The height of the box does not give additional information about the data, but should be tall enough to distinguish the box from the whiskers."
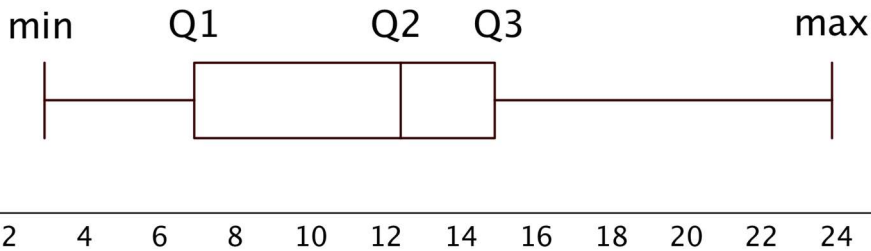
Ask students to now draw a box plot on the same grid, above their dot plot. Give students 4–5 minutes to complete the questions. Follow with a whole-class discussion.

Action and Expression: Internalise Executive Functions. Chunk this task into more manageable parts to support students who benefit from support with organisational skills in problem solving. Check in with students within the first 3-5 minutes of work time to see if they were able to find Q1, Q2, and Q3. If necessary, remind students how to find each quartile.
Supports accessibility for: Organisation; Attention

Student Task Statement

Twenty people participated in a study about blinking. The number of times each person blinked while watching a video for one minute was recorded. The data values are shown here, in order from smallest to largest.

3   6   8   11   11   13   14   14
14   14   16   18   20   20   20   22
24   32   36   51

   a. Use the grid and axis to make a dot plot of this data set.
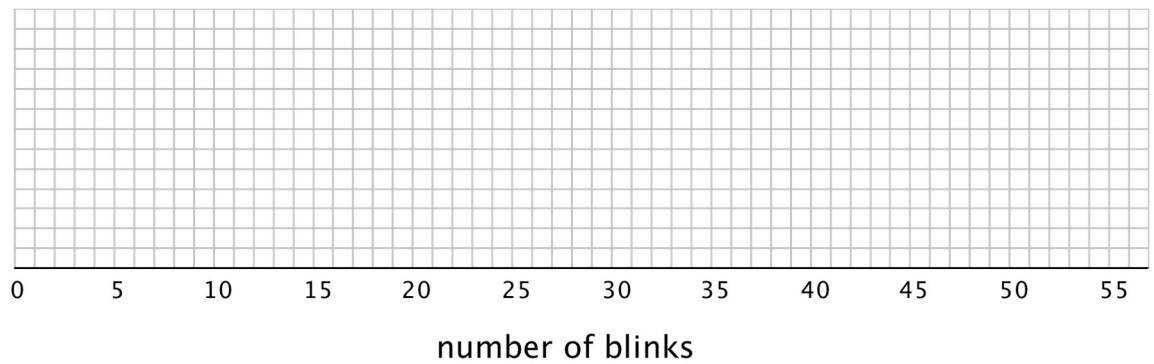
b. Find the median (Q2) and mark its location on the dot plot.

c. Find the first quartile (Q1) and the third quartile (Q3). Mark their locations on the dot plot.

d. What are the minimum and maximum values?

1. A **box plot** can be used to represent the five-number summary graphically. Let's draw a box plot for the number-of-blinks data. On the grid, *above* the dot plot:

a. Draw a box that extends from the first quartile (Q1) to the third quartile (Q3). Label the quartiles.

b. At the median (Q2), draw a vertical line from the top of the box to the bottom of the box. Label the median.

c. From the left side of the box (Q1), draw a horizontal line (a whisker) that extends to the minimum of the data set. On the right side of the box (Q3), draw a similar line that extends to the maximum of the data set.

2. You have now created a box plot to represent the number of blinks data. What fraction of the data values are represented by each of these elements of the box plot?

a. The left whisker

b. The box

c. The right whisker

Student Response
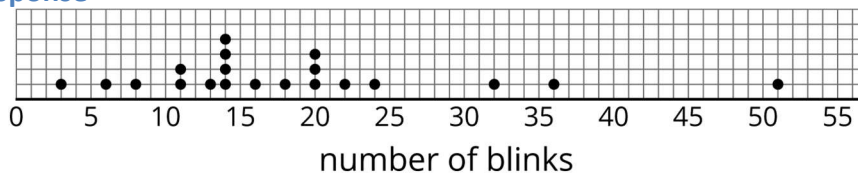
a.

b. 15. The median is 15 blinks, which is the average of the 10th and 11th values.

c. 12. The first quartile is 12 blinks, which is the average of the 5th and 6th values. The third quartile is 21 blinks, which is the average of the 25th and 26th values.

d. 3, 51. The minimum data point is 3 blinks, and the maximum data point is 51 blinks.
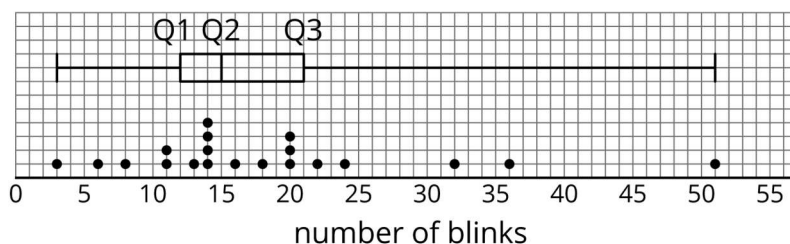
1.

2.

    a. $\frac{1}{4}$. The left whisker contains all the data values from the minimum to Q1. There are 5 points, which represents $\frac{1}{4}$ of the data.

    b. $\frac{1}{2}$. The box contains the values in between the 1st and 3rd quartile. There are 10 values, which comprises $\frac{1}{2}$ of the data.

    c. $\frac{1}{4}$. The right whisker contains all the data values from Q3 to the maximum. There are 5 points, which represents $\frac{1}{4}$ of the data.

Are You Ready for More?

Suppose there were some errors in the data set: the smallest value should have been 6 instead of 3, and the largest value should have been 41 instead of 51. Determine if any part of the five-number summary would change. If you think so, describe how it would change. If not, explain how you know.

Student Response

Reasonings vary. Sample reasonings:

- Minimum: Yes. The minimum would change from 3 to 6.

- First quartile (Q1): No. The lower quartile would still be the average of 11 and 13, because the number of data points in the set has not changed.

- Median (Q2): No. The median would still be the average of 14 and 16, because the number of data points in the set has not changed.

- Third Quartile (Q3): No. The upper quartile would still be the average of 20 and 22, because the number of data points in the set has not changed.

- Maximum: Yes. The maximum would change from 51 to 41.

Activity Synthesis

Display the dot plot and the box plot for all to see.
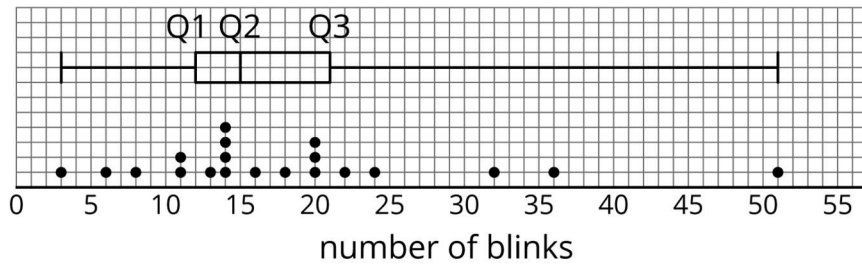
Discuss:

- "How many data values are included in each part of the box plot?" (5 data values in each part.)

- "If you just look at the box plot, can you tell what any of the data values are?" (Only the minimum and the maximum values.)

- "If you just look at the dot plot, can you tell where the median is? Can you tell which values of the data make up the middle half of the data? Can you tell where each quarter of the data values begin and end?" (It is possible to tell, but it is not straightforward; it requires some counting.)

The focus of this activity is on constructing a box plot and understanding its parts, rather than on interpreting it in context. If students seem to have a good grasp of the drawing process and what the parts entail and mean, consider asking them to interpret the plots in the context of the research study. Ask: "Suppose you are the scientist who conducted the research and are writing an article about it. Write 2–3 sentences that summarise your findings, based on your analyses of the dot plot and the box plot."

Representing, Conversing: Clarify, Critique, Correct. Use this routine to help students clarify a common misconception when creating a box plot from a dot plot. Display a box plot with the correct median of 15, but use 11 for Q1 and 22 for Q2, which are the values students might calculate if they did not average the two middle numbers. Invite students to work with a partner to identify the error and write a correct response. Look for and amplify mathematical uses of language involving the shape, centre and spread of the distribution. This will help students affirm their understanding of how to use box plots and dot plots to graphically represent a numerical data set.
Design Principle(s): Maximise meta-awareness; Cultivate conversation

Lesson Synthesis

In the lesson we see another way to graphically represent a numerical data set.
Review with students:

- "How is a **box plot** made?" (The box is a rectangle with the left side at Q1 and the right side at Q3. The line inside the box is the median. The "whiskers" on the sides extend to the minimum and maximum values of the data set.)

- "What does a box plot tell you about the shape, centre, and spread of a distribution?" (The median is the line in the middle, which tells you about the centre. The IQR is the width of the box in the middle, which tells you about the spread. You can also tell if the distribution is roughly symmetrical.)

- "Why is it useful to use a dot plot and a box plot together?" (The dot plot shows the actual data values while the box plot tells the story of the data in fourths.)

- "How can the box plot be helpful in comparing two data sets?" (You could compare the minimum and maximum values, where the median falls, and how the data is distributed among the four quarters.)

![MEI Mathematics Education Innovation logo]

16.4 Boxes and Dots

Cool Down: 5 minutes

1. Here are two box plots that summarise two data sets. Do you agree with each of the following statements?
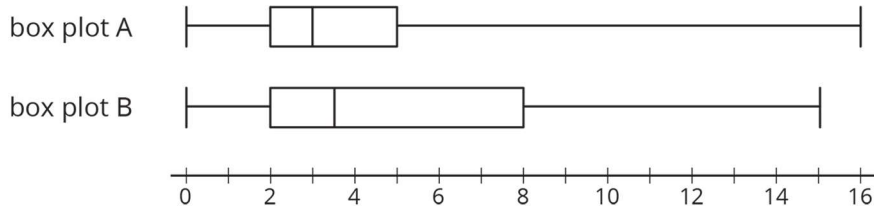
   A. Both data sets have the same range.

   B. Both data sets have the same minimum value.

   C. The IQR shown in box plot B is twice the IQR shown in box plot A.

   D. Box plot A shows a data set that has a quarter of its values between 2 and 5.

2. These dot plots show the same data sets as those represented by the box plots. Decide which box plot goes with each dot plot. Explain your reasoning.
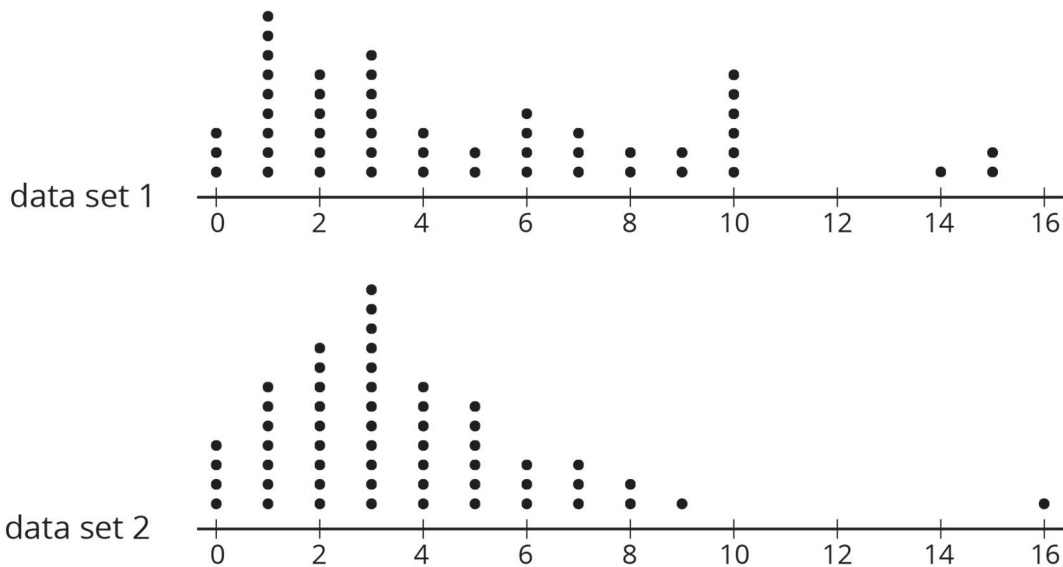
Student Response

   A. Disagree

   B. Agree

   C. Agree.

D.   Disagree

1.   Box plot A goes with data set 2. Box plot B goes with data set 1. Sample reasonings:

–   The maximum values tell which box plot goes with which dot plot.

–   The middle half of the points in data set 1 are more spread out compared to those in data set 2, so box plot B, which has a longer box, goes with data set 1.

–   Three quarters of the points in data set 2 are between 0 and 5, which matches box and left whisker in box plot A.

Student Lesson Summary

A **box plot** represents the five-number summary of a data set.

It shows the first quartile (Q1) and the third quartile (Q3) as the left and right sides of a rectangle or a box. The median (Q2) is shown as a vertical segment inside the box. On the left side, a horizontal line segment—a "whisker"—extends from Q1 to the minimum value. On the right, a whisker extends from Q3 to the maximum value.

The rectangle in the middle represents the middle half of the data. Its width is the IQR. The whiskers represent the bottom quarter and top quarter of the data set.

Earlier we saw dot plots representing the weights of pugs and beagles. The box plots for these data sets are shown above the corresponding dot plots.

pug weights in kilograms   beagle weights in kilograms
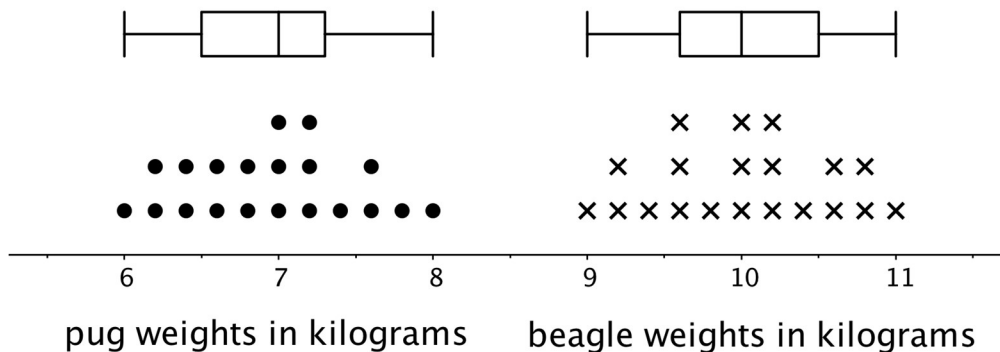
We can tell from the box plots that, in general, the pugs in the group are lighter than the beagles: the median weight of pugs is 7 kilograms and the median weight of beagles is 10 kilograms. Because the two box plots are on the same scale and the rectangles have similar widths, we can also tell that the IQRs for the two breeds are very similar. This suggests that the variability in the beagle weights is very similar to the variability in the pug weights.
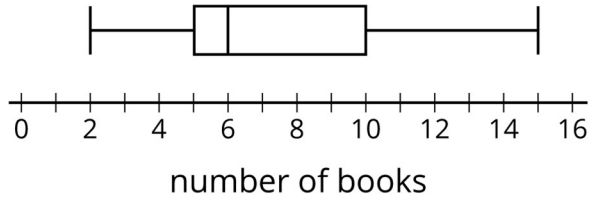
Glossary
•   box plot

Lesson 16 Practice Problems

Problem 1 Statement

Each student in a class recorded how many books they read during the summer. Here is a box plot that summarises their data.

number of books

a. What is the greatest number of books read by a student in this group?

b. What is the median number of books read by the students?

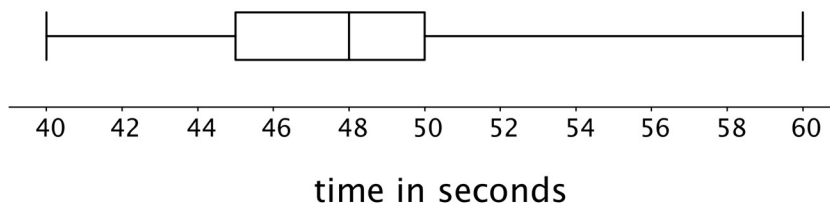c. What is the interquartile range (IQR)?

Solution

a. 15

b. 6

c. 5

Problem 2 Statement

Use this five-number summary to draw a box plot. All values are in seconds.

- Minimum: 40

- Lower quartile (Q1): 45

- Median: 48

- Upper quartile (Q3): 50

- Maximum: 60

Solution

time in seconds

Problem 3 Statement

The data shows the number of hours per week that each of 13 year 8 students spent doing homework. Create a box plot to summarise the data.
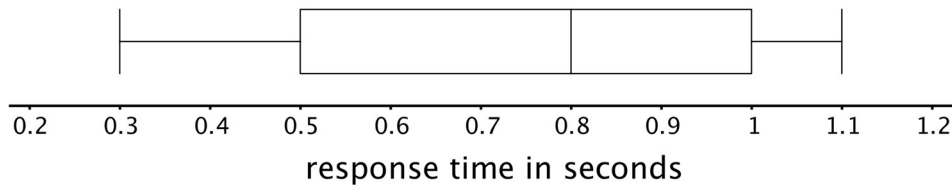
3	10	12	4	7	9	5	5
11	11	5	12	11			

Solution

0   1   2   3   4   5   6   7   8   9   10   11   12   13

time in hours
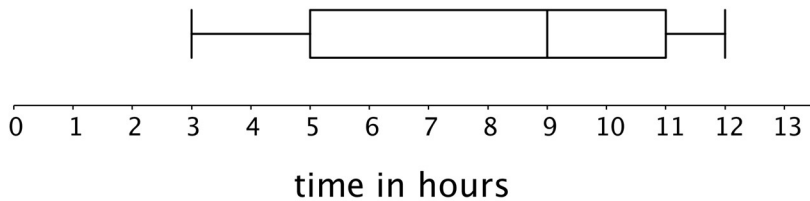
Problem 4 Statement

The box plot displays the data on the response times of 100 mice to seeing a flash of light. How many mice are represented by the rectangle between 0.5 and 1 second?
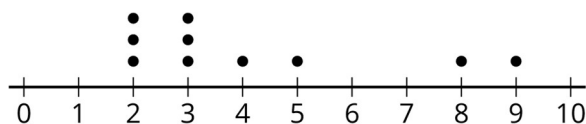

0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9   1   1.1   1.2

response time in seconds

Solution

50

Problem 5 Statement

Here is a dot plot that represents a data set. Explain why the mean of the data set is greater than its median.

0   1   2   3   4   5   6   7   8   9   10

Solution

Explanations vary. Sample explanation: The dot plot has a peak on the left and the median is near that peak. The values 8 and 9 are much larger than the other values. These large values make the mean larger than the median.

Problem 6 Statement

Jada earns money from babysitting, walking her neighbour's dogs, and running errands for her aunt. Every four weeks, she combines her earnings and divides them into three equal parts—one for spending, one for saving, and one for donating to a charity. Jada donated £26.00 of her earnings from the past four weeks to charity.

How much could she have earned from each job? Make two lists of how much she could have earned from the three jobs during the past four weeks.

Solution

Answers vary. Sample response:

- – £15, £16, £19, £28

- – £17, £23, £12.50, £25.50

(Any correct response will have a total of £78.)