Lesson 15: Quartiles and Interquartile Range

Goals

- Calculate the range and interquartile range (IQR) of a data set and interpret (orally and in writing) what they tell us about the situation.
- Comprehend that "interquartile range (IQR)" is another measure of variability that describes the span of the middle half of the data.
- Identify and interpret (in writing) the numbers in the five-number summary for a data set, i.e., the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum.

Learning Targets

- I can use IQR to describe the spread of data.
- I know what quartiles and interquartile range (IQR) measure and what they tell us about the data.
- When given a list of data values or a dot plot, I can find the quartiles and interquartile range (IQR) for data.

Lesson Narrative

Previously, students learned about decomposing a data set into two halves and using the halfway point, the median, as a measure of center of the distribution. In this lesson, they learn that they could further decompose a data set—into quarters—and use the **quartiles** to describe a distribution. They learn that the three quartiles—marking the 25th, 50th, and 75th percentiles—plus the maximum and minimum values of the data set make up a five-number summary.

Students also explore the **range** and **interquartile range (IQR)** of a distribution as two ways to measure its spread. Students reason abstractly and quantitatively (MP2) as they find and interpret the IQR as describing the distribution of the middle half of the data. This lesson prepares students to construct box plots in a future lesson.

Alignments

Addressing

- 6.SP.B.5.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 were gathered.
- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Building Towards

- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

Instructional Routines

- MLR2: Collect and Display
- MLR8: Discussion Supports
- Notice and Wonder
- Think Pair Share

Student Learning Goals

Let's look at other measures for describing distributions.

15.1 Notice and Wonder: Two Parties

Warm Up: 5 minutes

In earlier lessons, students learned that the mean absolute deviation (MAD) is a measure of variability. In this warm-up, they study two distributions that appear very different but turn out to have the same MAD. Students notice that the MAD may not fully tell us about the variability of a data set. The work here motivates the need to have a different way to quantify variability, which is the focus of this lesson. While students may notice and wonder many things about these images, highlight ideas related to the variability of the data sets.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

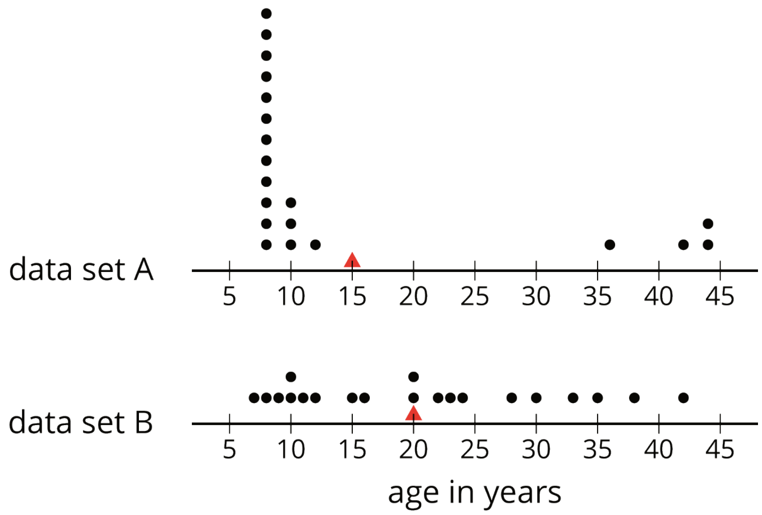
- Notice and Wonder

Launch

Arrange students in groups of 2. Display the dot plots for all to see. Ask students to identify at least one thing they notice and at least one thing they wonder about the dot plots, and to give a signal when they have both. Give students 1 minute of quiet think time, and then 1 minute to discuss their observation and question with their partner. Follow with a whole-class discussion.

Student Task Statement

Here are dot plots that show the ages of people at two different parties. The mean of each distribution is marked with a triangle.



What do you notice and what do you wonder about the distributions in the two dot plots?

Student Response

Answers vary. Sample responses:

Students may notice:

- The mean of the two data sets are different—the mean for the second data set is 5 years higher than that for the first.
- The range in values of the two data sets are about the same.
- Most points in the first data set are clustered around 8 and 10; only a few are much higher.
- The mean for the first data set is located where there are no observations.
- The points in the second data set are not clustered anywhere; they are distributed along the dot plot, between 5 and 42 years.
- The MAD values could be close for the two data sets.

Students may wonder:

- Why the data distributions look so different.
- If the MAD values are the same or close.
- If there could be other distributions that look very different than these two but also have the same MAD.

Activity Synthesis

Invite students to share what they noticed and wondered. Record and display their responses for all to see. If possible, record the relevant reasoning on or near the image. After each response, ask the class if they agree or disagree and to explain alternative ways of thinking, referring back to the dot plots each time. Discuss:

- “Do you think the ages of the people at the first party are alike or different? What about the ages of the people at the second party?”
- “The MAD for both data sets is approximately 10.5 years. What does a MAD of 10.5 years tell us in this context?”
- “Is the MAD a useful description of variability in the first data set? What about in the second data set?”

Two key ideas to uncover here are:

- The MAD is a way to summarize variation from the mean, but the single number does not always tell us how the data are distributed.
- The same MAD could result from very different distributions.

If the key ideas above are not uncovered during discussion, be sure to highlight them.

15.2 The Five-Number Summary

15 minutes
This activity introduces students to the *five-number summary* and the process of identifying the five numbers. Students learn how to partition the data into four sets: using the median to decompose the data into upper and lower halves, and then finding the middle of each half to further decompose it into quarters. They learn that each value that decomposes the data into four parts is called a **quartile**, and the three quartiles are the first quartile (Q1), second quartile (Q2, or the median), and third quartile (Q3). Together with the minimum and maximum values of the data set, the quartiles provide a five-number summary that can be used to describe a data set without listing or showing each data value.

Students reason abstractly and quantitatively (MP2) as they identify and interpret the quartiles in the context of the situation given.

Addressing

- 6.SP.B.5.c

Building Towards

- 6.SP.B.4

Instructional Routines

- MLR2: Collect and Display

Launch

Explain to students that they previously summarized variability by finding the MAD, which involves calculating the distance of each data point from the mean and then finding the average of those distances. Explain that we will now explore another way to describe variability and summarize the distribution of data. Instead of measuring how far away data points are from the mean, we will decompose a data set into four equal parts and use the markers that partition the data into quarters to summarize the spread of data.

Remind students that when there is an even number of values, the median is the average of the middle two values.

Arrange students in groups of 2. Give groups 8–10 minutes to complete the activity. Follow with a whole-class discussion.

Student Task Statement

Here are the ages of the people at one party, listed from least to greatest.

7	8	9	10	10	11	12	15	16
20	20	22	23	24	28	30	33	35
38	42							

1. a. Find the median of the data set and label it "50th percentile." This splits the data into an upper half and a lower half.

   b. Find the middle value of the *lower* half of the data, without including the median. Label this value "25th percentile."

   c. Find the middle value of the *upper* half of the data, without including the median. Label this value "75th percentile."

2. You have split the data set into four pieces. Each of the three values that split the data is called a **quartile**.
   - We call the 25th percentile the *first quartile*. Write "Q1" next to that number.
   - The median can be called the *second quartile*. Write "Q2" next to that number.
   - We call the 75th percentile the *third quartile*. Write "Q3" next to that number.

3. Label the lowest value in the set "minimum" and the greatest value "maximum."

4. The values you have identified make up the *five-number summary* for the data set. Record them here.

   minimum: _____ Q1: _____ Q2: _____ Q3: _____ maximum: _____

5. The median of this data set is 20. This tells us that half of the people at the party were 20 years old or younger, and the other half were 20 or older. What do each of these other values tell us about the ages of the people at the party?

   a. the third quartile

b. the minimum

c. the maximum

Student Response

Q1 25th percentile 10.5 | Q2 50th percentile 20 | Q3 75th percentile 29

min | max

7	8	9	10	10	11	12	15	16	20	20	22	23	24	28	30	33	35	38	42

1.
2. See above.
3. See above.
4. Minimum: 7 years; Q1: 10.5 years; Q2: 20 years; Q3: 29 years; Maximum: 42 years.
5. a. Q3 tells us that a quarter of the party goers are over the age of 29 years old, and the rest are younger.

   b. The minimum tells us that the youngest person at the party is 7 years old.

   c. The maximum tells us that the oldest person at the party is 42 years old.

Are You Ready for More?

There was another party where 21 people attended. Here is the five-number summary of their ages.

minimum: 5 Q1: 6 Q2: 27 Q3: 32 maximum: 60

1. Do you think this party had more children or fewer children than the earlier one? Explain your reasoning.
2. Were there more children or adults at this party? Explain your reasoning.

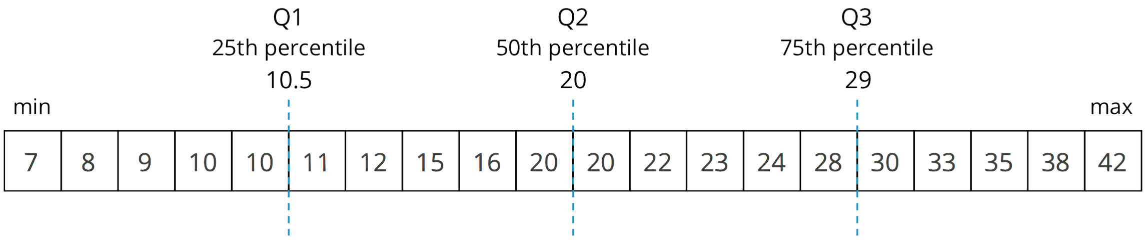
Student Response

Answers vary. Sample responses:

1. There are about the same, or possibly more, kids. Since the first quartile (Q1) is 6 years old, there are at least 6 and as many as 10 children at this party.
2. There are more adults at this party. The median age is 27 years old, an adult. Besides this adult, half of the other guests are adults aged 27 or older.

Activity Synthesis

Ask a student to display the data set they have decomposed and labeled, or display the following image for all to see.



Focus the conversation on students' interpretation of the five numbers. Discuss:

- “In this context, what do the minimum and maximum values tell us?” (The ages of the youngest and oldest partygoers.)
- “Why are Q1 called 25th percentile, Q2 50th percentile, and Q3 75th percentile?” (Each quartile tells us how many quarters of the ordered data values are accounted for up to that point. The first quartile tells us that one quarter, or 25 percent, of data values are less than or equal to that value. The second quartile tells us that two quarters, or 50 percent, of data values are less than or equal to that value, and so on.)
- “In this context, what does Q1 (10.5) tell us?” (That a quarter of the partygoers are 10.5 years old or younger.)
- “What does Q3 (29) tell us?” (That three quarters of the partygoers are 29 years old or younger.)
- “How do the five numbers help us to see the distribution of the data?” (It divides the values in the data into sections containing one fourth of the values each. This gives us an idea about the distribution of the data by looking at how varied each section is.)

Access for Students with Disabilities

Representation: Internalize Comprehension. Use color coding and annotations to identify each value of the five-number summary on a display. For example, label Q1 with the meaning in context (a quarter of the partygoers are 10.5 years old or younger).
Supports accessibility for: Visual-spatial processing

Access for English Language Learners

Representing, Conversing: MLR2 Collect and Display. Use this routine to collect student responses to the question: “How do the five numbers help us see the distribution of the data?” Pay close attention to language that amplifies the idea of variability and spread in each quartile as well as in the entire data set. Display the captured language for the whole class to see. This will help students to reference appropriate mathematical language related to quartiles and the five number summary of data.
Design Principle(s): Support sense-making

15.3 Range and Interquartile Range

15 minutes
In the previous activity, students learned about the five-number summary and how it could be used to summarize a data set. Here, students extend their work to finding the **range** and **interquartile range** (**IQR**) of a data set. They learn that both values provide information about a distribution of data: the range is the difference between the maximum and minimum values in the data, while the IQR is the difference between the third and first quartiles. While the range tells us how spread out (or close together) the overall data values are, the IQR tells us how spread out (or close together) the middle half of the data values are.

Students identify the range and IQR of a data set and analyze distributions with different IQRs. They reason abstractly and quantitatively (MP2) as they use the IQR to describe the variability of data.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

- MLR8: Discussion Supports
- Think Pair Share

Launch

Tell students that they will write the five-number summary of a distribution shown on a dot plot. Give students a moment of quiet time to look at the dot plot in the first question and think about how they might identify the quartiles. Then, ask students to share their ideas. Students might suggest the following strategies.

- List the values of all the data points, put them in order, and then count off the values to find the median and the other two quartiles.

- Count the points by 3's (because the data set is to be decomposed into 4 equal parts and $12 \div 4 = 3$) and mark the end of the first set with Q1, the end of the second set with Q2, etc.
- Divide the points into two halves (by counting 6 points from the left or from the right), and then dividing each half into two halves.

It is not necessary that all of these ideas are brought up at this point, but if no students mentioned the first approach (listing all values), mention it. The concrete process of writing out all the values, in order, is likely to be accessible to most students. The list of values would also be familiar, as it would resemble the one in the preceding activity.

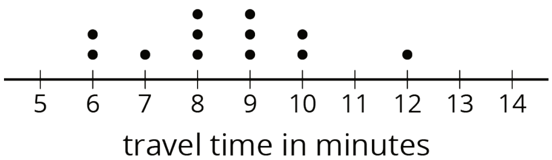
Arrange students in groups of 2. Give students 3–4 minutes of quiet work time for the first question, and 5–7 minutes to discuss their work with their partner and to complete the rest of the activity. Follow with a whole-class discussion.

Anticipated Misconceptions

When finding the IQR of the dot plots in the last question, students might neglect to divide the data set into four parts. Or they might instead divide the distance between the maximum and minimum into four parts (rather than dividing the data points into four parts). Remind students about the conversation at the start of the task about listing all the values or counting off the data points in order to find the quartiles.

Student Task Statement

1. Here is a dot plot that shows the lengths of Elena's bus rides to school, over 12 days.



   Write the five-number summary for this data set. Show your reasoning.

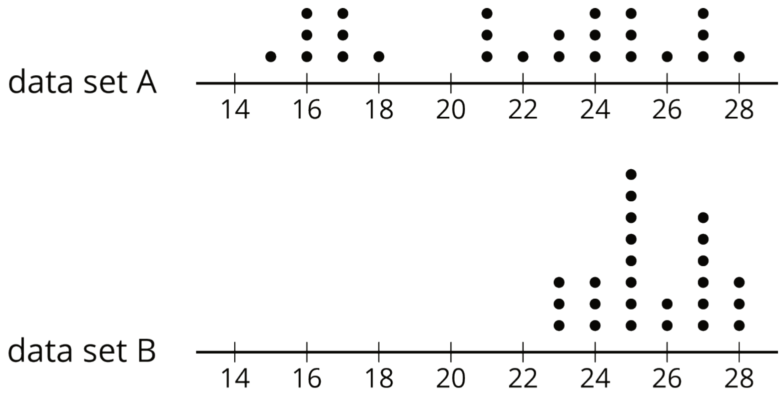
2. The **range** is one way to describe the *spread* of values in a data set. It is the difference between the maximum and minimum. What is the range of Elena's travel times?

3. Another way to describe the spread of values in a data set is the **interquartile range (IQR)**. It is the difference between the upper quartile and the lower quartile.

   a. What is the interquartile range (IQR) of Elena's travel times?

   b. What fraction of the data values are between the lower and upper quartiles?

4. Here are two more dot plots.



Without doing any calculations, predict:

a. Which data set has the smaller range?

b. Which data set has the smaller IQR?

5. Check your predictions by calculating the range and IQR for the data in each dot plot.

Student Response

1. The five-number summary is minimum: 6; Q1: 7.5; Q2: 8.5; Q3: 9.5; maximum: 12. This is found by ordering the travel times in a list. The Q1 is the average of 7 and 8, the 3rd and 4th values. The Q2 is the average of 8 and 9, the 6th and 7th values. The Q3 is the average of 9 and 10, the 9th and 10th values.

2. The range of Elena's data set is 6. The smallest data point is 6, and the largest data point is 12; the difference between these is 6.

3. a. The 1st and 3rd quartiles are 7.5 and 9.5. The difference between these values, or the IQR, is 2.

   b. $\frac{1}{2}$ of the data set is between the upper and lower quartiles.

4. a. Data set B has the smaller range. The minimum and maximum are closer together in data set B than in data set A.

   b. Data set B has the smaller IQR. The data points in data set B are closer together than in data set A, so the distance between Q1 and Q3 should be smaller.

5. The range for data set A is 13 since the maximum value is 28 and the minimum value is 15. The range for data set B is 5 since the maximum value is 28 and the minimum value is 23. The IQR of data set A is 8, since the Q1 is 17, and the Q3 is 25. The IQR of data set B is 2.5, since the Q1 is 24.5, and the Q3 is 27.

Activity Synthesis

Ensure that students know how to find the range and IQR, and then focus the discussion on interpreting these two measures and how they provide information about a distribution.

Select a couple of students to share the range and IQR of Elena's data. Ask:

- “What does a range of 6 minutes tell us about Elena's travel time?” (Elena's travel times vary by 6 minutes at most, or that the difference between the shortest commute and the longest one is 6 minutes.)
- “What does an IQR of 2 minutes tell us about her travel time?” (The middle half of Elena's travel times vary by 2 minutes.)

Then, select a few other students to explain their response to the third question. Discuss:

- “Without calculating, how did you determine which data set had the smaller range?” (The dot plot that has the narrower spread would have the smaller range because the distance between the greatest and least values is smaller.)
- “How did you determine which one had the smaller IQR?” (The dot plot whose middle half of the points seem more clustered together would have the smaller IQR.)
- “In general, what does a larger range tell us?” (A wider spread in the data, more variability in the data set.)
- “What does a larger IQR tell us?” (A wider spread around the center of data, more variability in the middle half of the data set.)
- “Can a data set have a large range and a small IQR?” (It is possible, if the data set has most of its points very close together but there are a few points that are far away from the cluster.)

If not mentioned by students, explain that the IQR plays a similar role as the mean absolute deviation (MAD): it tells us how different and spread out the data values are. But instead of measuring the average distance of data values from the mean, it measures the span of the middle half of the data.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Include the following terms and maintain the display for reference throughout the unit: quartile, range, and interquartile range (IQR). Invite students to suggest language or diagrams to include on the display that will support their understanding of this term.
Supports accessibility for: Memory; Language

Access for English Language Learners

Speaking: MLR8 Discussion Supports. Use this routine to support whole-class discussion. For each response that is shared, ask students to restate and/or revoice what they heard using mathematical language. Consider providing students time to restate what they hear to a partner, before selecting one or two students to share with the class. Ask the original speaker if their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped to clarify the original statement. This will provide more students with an opportunity to produce language and support their understanding of distributions with different IQRs.
Design Principle(s): Support sense-making

Lesson Synthesis

- "What are the **quartiles** for a numerical data set?" (Numbers that show where we split the data up so it is in quarters.)
- "What is the relationship between the quartiles and the median?" (The second quartile is also the median).
- "What is the **Interquartile range (IQR)**? What does it mean?" (The IQR is the difference between the third and first quartile. It is a measure of the variability or spread of the data. It tells us how much "space" the middle half of the data occupies.)
- "Compare MAD and IQR. How are they alike? How are they different?" (They both provide information on the distribution of a set of data. MAD works with the mean while IQR works with the median. MAD considers all the data values and tells us the average distance between each data value and the mean, while IQR focuses on the middle half of the data and tells us how widely distributed it is.)

15.4 How Far Can You Throw?

Cool Down: 5 minutes
Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Student Task Statement

Diego wondered how far sixth-grade students could throw a ball. He decided to collect data to find out. He asked 10 friends to throw a ball as far as they could and measured the distance from the starting line to where the ball landed. The data shows the distances he recorded in feet.

40	76	40	63	47	57	49	55	50	53

1. Find the median and IQR of the data set.
2. On a later day, he asked the same group of 10 friends to throw a ball again and collected another set of data. The median of the second data set is 49 feet, and the IQR is 6 feet.
   a. Did the 10 friends, as a group, perform better (throw farther) in the second round compared to the first round? Explain how you know.
   b. Were the distances in the second data set more variable or less variable compared to those in the first round? Explain how you know.

Student Response

1. The median is 51.5 feet. $(50 + 53) \div 2 = 51.5$. The IQR is 10, because Q1 is 47, Q3 is 57, and $57 - 47 = 10$.
2. a. Worse. Sample reasoning: The median of the second data set is 49 feet, which is 2.5 feet lower than in the first round.
   b. Less variable. Sample reasoning: The IQR of the second data set is smaller, so the values are less spread out.

Student Lesson Summary

Earlier we learned that the mean is a measure of the center of a distribution and the MAD is a measure of the variability (or spread) that goes with the mean. There is also a measure of spread that goes with the median. It is called the interquartile range (IQR).

Finding the IQR involves splitting a data set into fourths. Each of the three values that splits the data into fourths is called a **quartile**.

- The median, or second quartile (Q2), splits the data into two halves.
- The first quartile (Q1) is the middle value of the lower half of the data.
- The third quartile (Q3) is the middle value of the upper half of the data.

For example, here is a data set with 11 values.

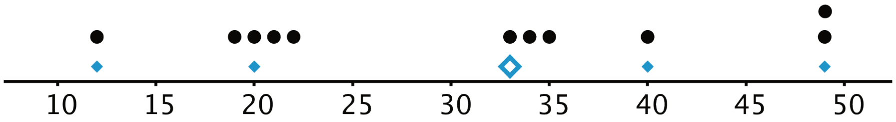
12	19	20	21	22	33	34	35	40	40	49
		Q1			Q2			Q3		

- The median is 33.
- The first quartile is 20. It is the median of the numbers less than 33.
- The third quartile 40. It is the median of the numbers greater than 33.

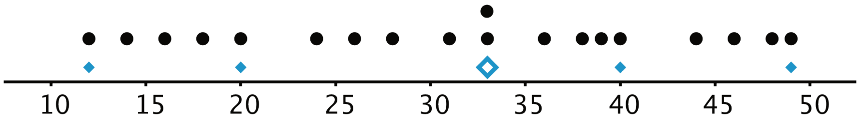
The difference between the maximum and minimum values of a data set is the **range**. The difference between Q3 and Q1 is the **interquartile range (IQR)**. Because the distance between Q1 and Q3 includes the middle two-fourths of the distribution, the values between those two quartiles are sometimes called the *middle half of the data*.

The bigger the IQR, the more spread out the middle half of the data values are. The smaller the IQR, the closer together the middle half of the data values are. This is why we can use the IQR as a measure of spread.

A *five-number summary* can be used to summarize a distribution. It includes the minimum, first quartile, median, third quartile, and maximum of the data set. For the previous example, the five-number summary is 12, 20, 33, 40, and 49. These numbers are marked with diamonds on the dot plot.



Different data sets can have the same five-number summary. For instance, here is another data set with the same minimum, maximum, and quartiles as the previous example.



Glossary

- interquartile range (IQR)
- quartile
- range

Lesson 15 Practice Problems

Problem 1

Statement

Suppose that there are 20 numbers in a data set and that they are all different.

a. How many of the values in this data set are between the first quartile and the third quartile?

b. How many of the values in this data set are between the first quartile and the median?

Solution

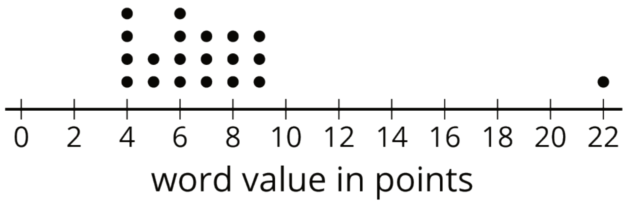
a. 10. There are 5 numbers in each quartile, and there are two quartiles in between the first and third quartiles.

b. 5. The median is the second quartile. The first quartile to the second comprises one quartile.

Problem 2

Statement

In a word game, 1 letter is worth 1 point. This dot plot shows the scores for 20 common words.



a. What is the median score?

b. What is the first quartile (Q1)?

c. What is the third quartile (Q3)?

d. What is the interquartile range (IQR)?

Solution

a. 6.5 points

b. 5

c. 8

d. 3

Problem 3

Statement

Mai and Priya each played 10 games of bowling and recorded the scores. Mai's median score was 120, and her IQR was 5. Priya's median score was 118, and her IQR was 15. Whose scores probably had less variability? Explain how you know.

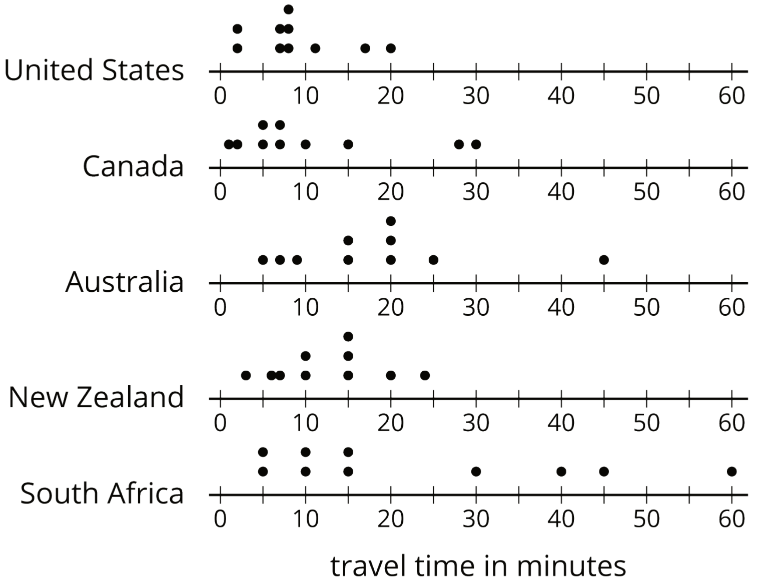
Solution

Answers vary. Sample explanation: Mai's IQR was smaller, so her scores probably varied less than Priya's scores.

Problem 4

Statement

Here are five dot plots that show the amounts of time that ten sixth-grade students in five countries took to get to school. Match each dot plot with the appropriate median and IQR.



a. Median: 17.5, IQR: 11

b. Median: 15, IQR: 30

c. Median: 8, IQR: 4

d. Median: 7, IQR: 10

e. Median: 12.5, IQR: 8

Solution

United States: 3

Canada: 4

Australia: 1

New Zealand: 5

South Africa: 2

Problem 5

Statement

Draw and label an appropriate pair of axes and plot the points. $A = (10, 50)$, $B = (30, 25)$, $C = (0, 30)$, $D = (20, 35)$

Solution

Answers vary. Check student work to ensure they made reasonable choices about axes and scale that allowed them to clearly plot all the points.

(From Unit 7, Lesson 12.)

Problem 6

Statement

There are 20 pennies in a jar. If 16% of the coins in the jar are pennies, how many coins are there in the jar?

Solution

125, because $20 \div 0.16 = 125$.

(From Unit 6, Lesson 7.)