Lesson 5: Using Dot Plots to Answer Statistical Questions

Goals

- Compare and contrast (orally and in writing) dot plots that represent two different data sets measuring the same quantity, paying attention to the “center” and “spread” of each distribution.
- Critique or justify (orally and in writing) claims about the center of a distribution represented on a dot plot.

Learning Targets

- I can use a dot plot to represent the distribution of a data set and answer questions about the real-world situation.
- I can use center and spread to describe data sets, including what is typical in a data set.

Lesson Narrative

In this lesson, students continue to use dot plots to develop their understanding of center and spread—by identifying values of center, describing spread, comparing centers and spreads of different distributions, and making use of the structure of the distributions (MP7) to understand them in the context of situations (MP2). In future lessons, they will make their descriptions and analyses more precise, as they learn about measures of center and spread.

Alignments

Building On

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

Addressing

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- 6.SP.B: Summarize and describe distributions.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Building Towards

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- 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.

Instructional Routines

- MLR5: Co-Craft Questions
- MLR8: Discussion Supports
- Think Pair Share

> **Student Learning Goals**
>
> Let's use dot plots to describe distributions and answer questions.

5.1 Packs on Backs

Warm Up: 5 minutes
In a previous lesson, students were exposed to the ideas of center and spread. Here, they begin connecting that idea informally to the word “typical” and a value that could be considered typical or characteristic of a data set by thinking about two good options and reasonings. They continue to interpret a dot plot in the context of a situation (MP2).

During the partner discussion, identify two students—one who agrees with Clare and another who agrees with Tyler—to share during the whole-class discussion (MP3).

Addressing

- 6.SP.B.4

Building Towards

- 6.SP.A.3

Instructional Routines

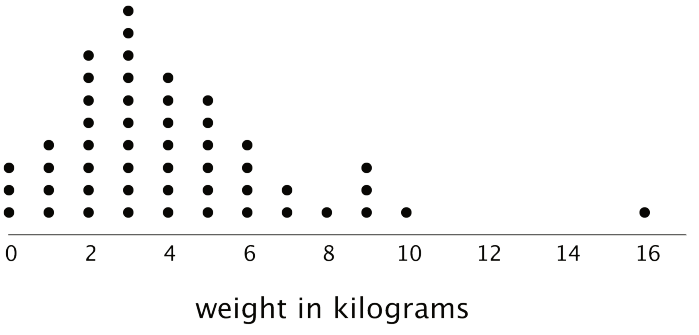
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 2 minutes of quiet work time and a minute to share their responses with a partner. Follow with a whole-class discussion.

Student Task Statement

This dot plot shows the weights of backpacks, in kilograms, of 50 sixth-grade students at a school in New Zealand.



1. The dot plot shows several dots at 0 kilograms. What could a value of 0 mean in this context?

2. Clare and Tyler studied the dot plot.

   - Clare said, "I think we can use 3 kilograms to describe a typical backpack weight of the group because it represents 20%—or the largest portion—of the data."
   - Tyler disagreed and said, "I think 3 kilograms is too low to describe a typical weight. Half of the dots are for backpacks that are heavier than 3 kilograms, so I would use a larger value."

   Do you agree with either of them? Explain your reasoning.

Student Response

1. Answers vary. Sample response: A value of 0 could represents students who don't use backpacks.

2. Answers vary. Sample response:
   - I agree with Clare. There are more backpacks that are 3 kilograms than any other weights, and half of the dots are around 3 kilograms (between 2 and 4 kilograms).
   - I agree with Tyler. The middle or half-way point of the data set is between 3 and 4, around 3.5, with the same number of dots above and below, so that value would be a reasonable value to use to describe what is typical for the group.

Activity Synthesis

Ask students to share their response to the first question about data points. Record and display their responses for all to see. Ask the selected students—one who agrees with Clare and another who agrees with Tyler—to share their reasoning. Ask if anyone disagrees with both students, and if so, what value they would consider a better description of the center of the data.

Students should have a reasonable explanation for each argument they favor, but it is not necessary to confirm one way or another at this point. Tell students that we will look more closely at different ways to determine a value that is characteristic of a data set in upcoming activities.

5.2 On the Phone

15 minutes
Earlier, in the backpack example, students saw a distribution described in terms of where data points are clustered on a dot plot and which values have a large number of occurrences. The shape of that distribution was approximately symmetric. In this activity, they continue to analyze distributions in those terms, and try to identify and interpret the **center** and **spread** of a distribution that is not symmetric. The two distributions used here allow students to contrast a narrow spread and a wide spread and develop a deeper understanding of variability.

As students work, notice how students identify a general location for the center of a data set and the descriptions they use to talk about the spread (e.g., "wide," "narrow," or "something in between"). Identify students who connect the size of a spread to how different or alike the data points are; ask them to share later. Additionally, identify students who measure spread as the range of the entire data as well as those who use the distance to the center.

Building On

- 6.SP.B.4

Addressing

- 6.SP.A.2
- 6.SP.B

Building Towards

- 6.SP.B.5.c

Instructional Routines

- MLR5: Co-Craft Questions
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 5–6 minutes of quiet work time for the first three sets of questions, and another 4–5 minutes to share their responses and then discuss the last two questions with a partner.

Students are asked to find a percentage. If necessary, briefly review how to find a percentage.

Access for Students with Disabilities

Representation: Internalize Comprehension. Activate or supply background knowledge about calculating percentages. Allow students to use calculators to ensure inclusive participation in the activity.
Supports accessibility for: Memory; Conceptual processing

Access for English Language Learners

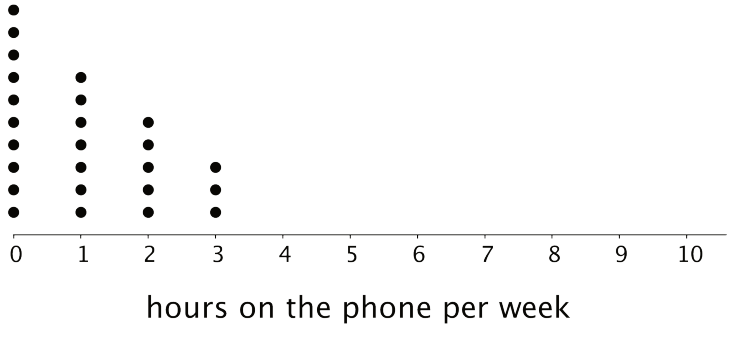
Representing, Writing: MLR5 Co-Craft Questions. Invite pairs of students to create one or two mathematical questions that could be answered by the data displayed in the dot plot. Note questions that vary in complexity, making sure to have students share examples that ask about percent, center, or spread, if available. Allow students to ask clarifying questions of their peers regarding how the dot plot could be used to answer the questions, if needed. This will help students to make sense of the data using informal language prior to connecting their existing understanding to formal language.
Design Principle(s): Maximize meta-awareness; Support sense-making

Anticipated Misconceptions

Students may neglect to change the rate given (from minutes per week to hours per week, or to minutes per day) and may draw incorrect conclusions as a result. Ask them to think about the unit they are using in their responses.

Student Task Statement

Twenty-five sixth-grade students were asked to estimate how many hours a week they spend talking on the phone. This dot plot represents their reported number of hours of phone usage per week.



1. a. How many of the students reported not talking on the phone during the week? Explain how you know.

b. What percentage of the students reported not talking on the phone?

2. a. What is the largest number of hours a student spent talking on the phone per week?

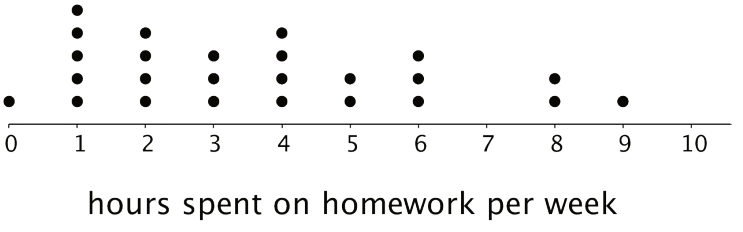
   b. What percentage of the group reported talking on the phone for this amount of time?

3. a. How many hours would you say that these students typically spend talking on the phone?

   b. How many minutes per day would that be?

4. a. How would you describe the **spread** of the data? Would you consider these students' amounts of time on the phone to be alike or different? Explain your reasoning.

   b. Here is the dot plot from an earlier activity. It shows the number of hours per week the same group of 25 sixth-grade students reported spending on homework.



hours spent on homework per week

   Overall, are these students more alike in the amount of time they spend talking on the phone or in the amount of time they spend on homework? Explain your reasoning.

5. Suppose someone claimed that these sixth-grade students spend too much time on the phone. Do you agree? Use your analysis of the dot plot to support your answer.

Student Response

1. a. Ten students. I could tell from the number of dots showing a value of 0.

   b. 40% of the group reported not using the phone.

2. a. The highest usage is 3 hours per week.

   b. 12% of the students reported talking on the phone for this amount of time.

3. Answers vary. Sample response:
   a. I would say that these students typically spend about 1 hour on the phone per week.

   b. About 9 minutes per day, because $60 \div 7$ is about 9.

4. Answers vary. Sample response:

a. The spread is pretty small. The hours reported span from 0 to 3, with a little more than half of the values being either 0 or 1, and the rest being 2 or 3. I think the phone usage is fairly different between those who don't talk on the phone at all (40% of the group) and those who talk for 2 or 3 hours a week (40% of the group).

b. The students are more alike in their phone usage than in the time they spend on homework. The spread of the data is much larger on the homework time dot plot, which means there is much more variability in the time they spend doing homework than the time they spend on the phone.

5. Answers vary. Sample response: I disagree. If 8.5 minute a day is typical, it is not a huge amount of time.

Activity Synthesis

The purpose of the discussion is to help students find good ways to describe a distribution based on center and spread.

Select a few students to answer the questions about the first dot plot. Tell students that the "typical" value for the data is generally considered the center. Ask, "What would you consider the center for the two dot plots shown in this activity?"

Ask students how they thought about the spread of the data. If possible, select students who thought of spread as the range of the entire data and those who thought of it as an interval around the center. Ask students to share their interpretation of the what the spread means in the context of using the phone. Make sure to include previously identified students who connect spread to how alike or different the data points are.

Tell students that distributions are generally described using the center and spread. Select a few students to describe the distributions of the two data sets shown in this activity.

5.3 Click-Clack

15 minutes

Previously students analyzed distributions to identify center and spread. In this activity, they continue to practice finding reasonable values for centers of data and describing variability. The focus, however, is on making use of the structure of distributions (MP7) to compare groups in those terms and interpreting their analyses in the context of a situation (MP2).

By comparing distributions, seeing how center and spread for the same population could change, and making sense of what these changes mean, students deepen their understanding of these concepts before learning about more formal measures of center and variability.

Building On

- 6.SP.B.4

Addressing

- 6.SP.A.2
- 6.SP.B

Building Towards

- 6.SP.A.3

Instructional Routines

- MLR8: Discussion Supports
- Think Pair Share

Launch

Give students a brief overview on keyboarding courses. Explain that these are classes designed to help people improve their typing speed and accuracy, which they may need for their jobs. Typing proficiency is usually measured in terms of number of words typed per minute; the more words typed correctly per minute, the faster or more proficient one's typing is.

Keep students in groups of 2. Give them 5–6 minutes of quiet time to work on the first two questions, and then 2–3 minutes to discuss their responses and complete the last question together.

Access for English Language Learners

Representing, Writing, Conversing: MLR8 Discussion Supports. When explaining the context of keyboarding courses, ask students to complete the sentence frame, "A person's typing speed increases when they type (more / fewer) words per minute because…." This will help students to connect the language of the questions to the meaning of keyboarding improvement. *Design Principle(s): Support sense-making; Cultivate conversation*

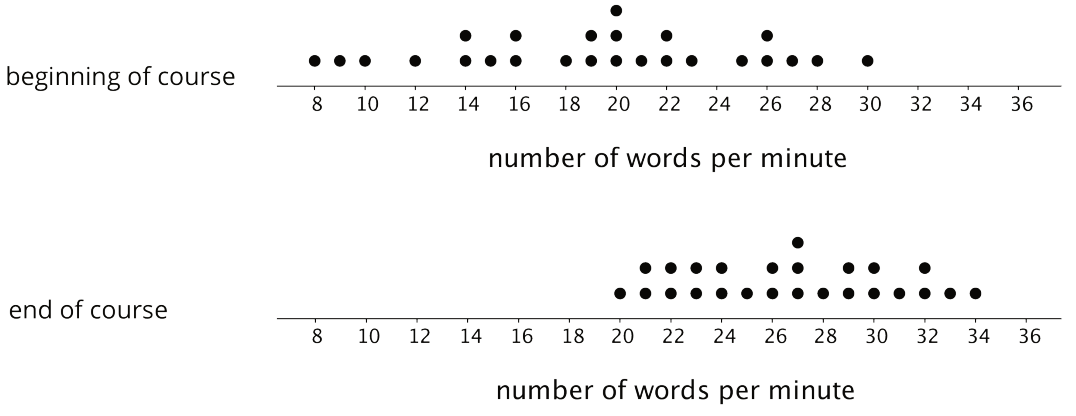
Anticipated Misconceptions

Some students might find it challenging to tell where the center of a distribution could be just by looking at a single dot plot. The idea of center might be more apparent when presented in comparative terms. For example, ask them to describe in their own words how the distribution of the first dot plot differs from that of the second dot plot. Students are likely able to say that, compared to the first dot plot, the aggregation of dots in the second dot plot is overall higher on the number line. Ask them if there is a location on each dot plot around which data points seem to congregate.

Student Task Statement

1. A keyboarding teacher wondered: "Do typing speeds of students improve after taking a keyboarding course?" Explain why her question is a statistical question.

2. The teacher recorded the number of words that her students could type per minute at the beginning of a course and again at the end. The two dot plots show the two data sets.



Based on the dot plots, do you agree with each of the following statements about this group of students? Be prepared to explain your reasoning.

a. Overall, the students' typing speed did not improve. They typed at the same speed at the end of the course as they did at the beginning.

b. 20 words per minute is a good estimate for how fast, in general, the students typed at the beginning of the course.

c. 20 words per minute is a good description of the **center** of the data set at the end of the course.

d. There was more variability in the typing speeds at the beginning of the course than at the end, so the students' typing speeds were more alike at the end.

3. Overall, how fast would you say that the students typed after completing the course? What would you consider the center of the end-of-course data?

Student Response

1. The question is a statistical question because data are needed to answer it, and we can expect it to have variability.

2. Answers vary. Sample response:
   a. Disagree. The set of dots were, as a group, placed lower on the number line at the beginning of the course and higher at the end, which means that, as a class, they were typing more words per minute or typing faster.

   b. Agree. At the beginning, 20 is more or less in the middle of the set of dots. It is reasonable to use this value to describe the set in general. (Or: Disagree. There are quite a few data points that are much lower or higher than 20. It doesn't seem that 20 is a good estimate of the set.)

c. Disagree. Almost all students were typing more than 20 words a minute, so 20 would not be a good description of the group's typing speed at the end of the course.

d. Agree. The spread in first dot plot is wider, which means the speeds of the students were quite different or more variable. The dots on the second dot plot are closer together, or the spread is narrower, which means at the end of the course the typing speeds were less variable.

3. Answers vary. Sample response: I would consider the center of the data to be about 26. In general, the group of students were typing at about 26 words per minute after taking the course.

Are You Ready for More?

Use one of these suggestions (or make up your own). Research to create a dot plot with at least 10 values. Then, describe the center and spread of the distribution.

- Points scored by your favorite sports team in its last 10 games
- Length of your 10 favorite movies (in minutes)
- Ages of your favorite 10 celebrities

Student Response

Answers vary.

Activity Synthesis

The purpose of the discussion is for students to deepen their understanding of distributions and using the descriptions to compare two groups.

Focus the whole-class discussion on two ideas:

- The distinctions between the two distributions: Students should see that, overall, the cluster of data points have both shifted up toward a greater number of words per minute (moving its center up) and become more compressed in its spread by the end of the course. Because the center moved up in location, the value we use to describe that center would also increase.
- What the changes in the center and spread tell us in this situation: Students should recognize that a higher center means that, overall, the group has improved in their typing speed. They should see that a narrower spread at the end of the course suggests that there's now less variability in the typing speeds of different students (compared to a much larger variability initially).

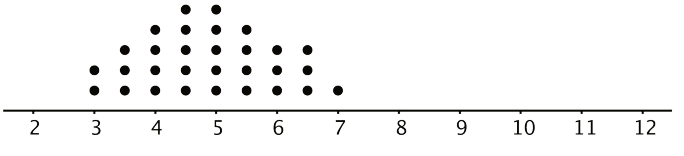
Access for Students with Disabilities

Representation: Internalize Comprehension. Use color coding and annotations to highlight the distinctions between the two distributions. For example, make visible the changes in the center and spread, and note connections to the situation.
Supports accessibility for: Visual-spatial processing

Lesson Synthesis

In this lesson, we talk about using the center and the spread of a distribution to describe a data set.

- "What do we mean by the 'center' of a distribution?"
- "What do we mean by the 'spread' of a distribution?"
- "How does the center and spread of a distribution relate to a typical value that could represent the group?"
- "How can we see center and spread in a dot plot like this?"



cat weights in kilograms

5.4 Packing Tomatoes

Cool Down: 5 minutes

Building On

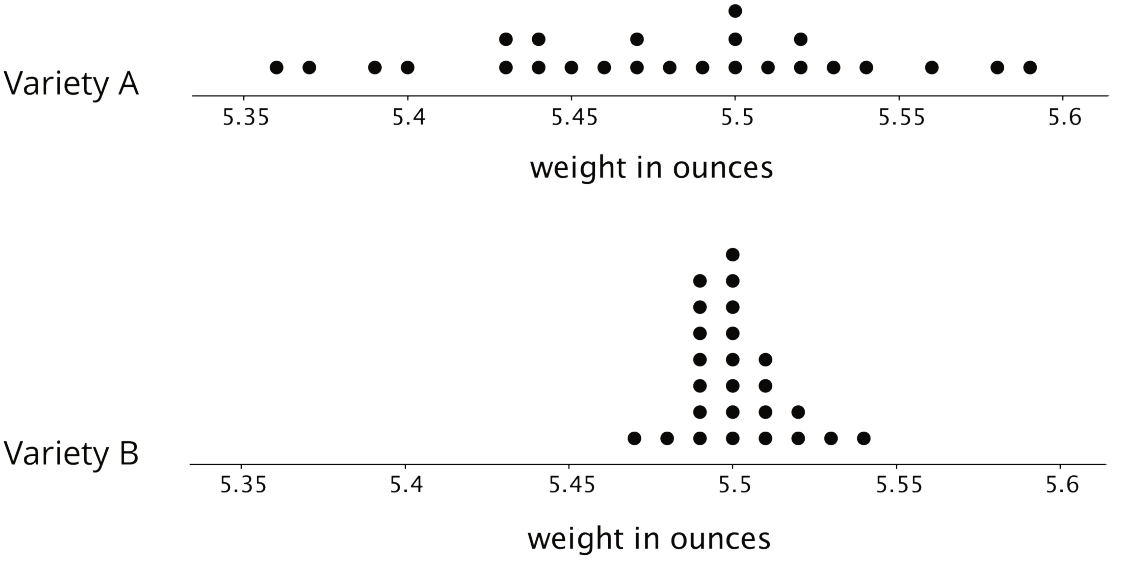
- 6.SP.B.4

Addressing

- 6.SP.A.2
- 6.SP.B.5.b

Student Task Statement

A farmer sells tomatoes in packages of ten. She would like the tomatoes in each package to all be about the same size and close to 5.5 ounces in weight. The farmer is considering two different tomato varieties: Variety A and Variety B. She weighs 25 tomatoes of each variety. These dot plots show her data.



1. What would be a good description for the weight of Variety A tomatoes, in general? What about for the weight of Variety B tomatoes, in general?

2. Which tomato variety should the farmer choose? Explain your reasoning.

Student Response

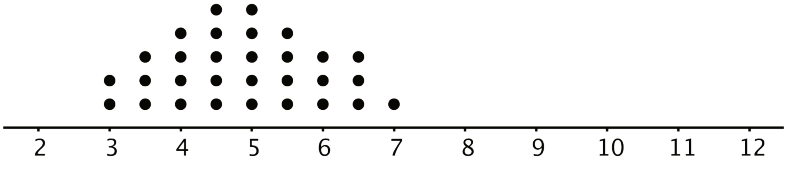
1. In general, Variety A tomatoes are about 5.49 ounces and Variety B tomatoes are about 5.5 ounces.

2. She should choose Variety B. Sample reasoning: The two varieties of tomatoes have about the same typical weight, in general, but there was much less variation in Variety B tomato. The weights were much more consistent than the weights for Variety A, so the tomatoes are more likely to be the same size and closer to 5.5 ounces in weight.

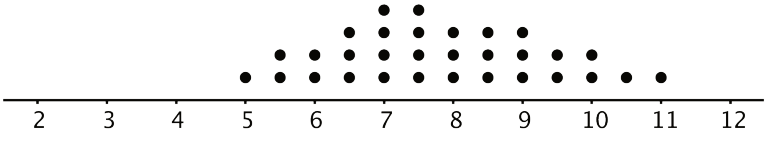
Student Lesson Summary

One way to describe what is typical or characteristic for a data set is by looking at the **center** and **spread** of its distribution.

Let's compare the distribution of cat weights and dog weights shown on these dot plots.



cat weights in kilograms



dog weights in kilograms

The collection of points for the cat data is further to the left on the number line than the dog data. Based on the dot plots, we may describe the center of the distribution for cat weights to be between 4 and 5 kilograms and the center for dog weights to be between 7 and 8 kilograms.

We often say that values at or near the center of a distribution are typical for that group. This means that a weight of 4–5 kilograms is typical for a cat in the data set, and weight of 7–8 kilograms is typical for a dog.

We also see that the dog weights are more spread out than the cat weights. The difference between the heaviest and lightest cats is only 4 kilograms, but the difference between the heaviest and lightest dogs is 6 kilograms.

A distribution with greater spread tells us that the data have greater variability. In this case, we could say that the cats are more similar in their weights than the dogs.

In future lessons, we will discuss how to measure the center and spread of a distribution.

Glossary

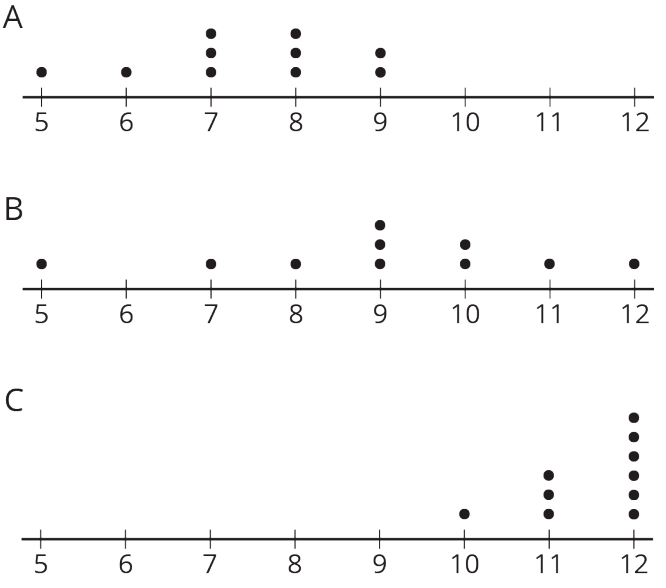
- center
- spread

Lesson 5 Practice Problems

Problem 1

Statement

Three sets of data about ten sixth-grade students were used to make three dot plots. The person who made these dot plots forgot to label them. Match each dot plot with the appropriate label.



A. Dot Plot A

B. Dot Plot B

C. Dot Plot C

1. Ages in years
2. Numbers of hours of sleep on nights before school days
3. Numbers of hours of sleep on nights before non-school days

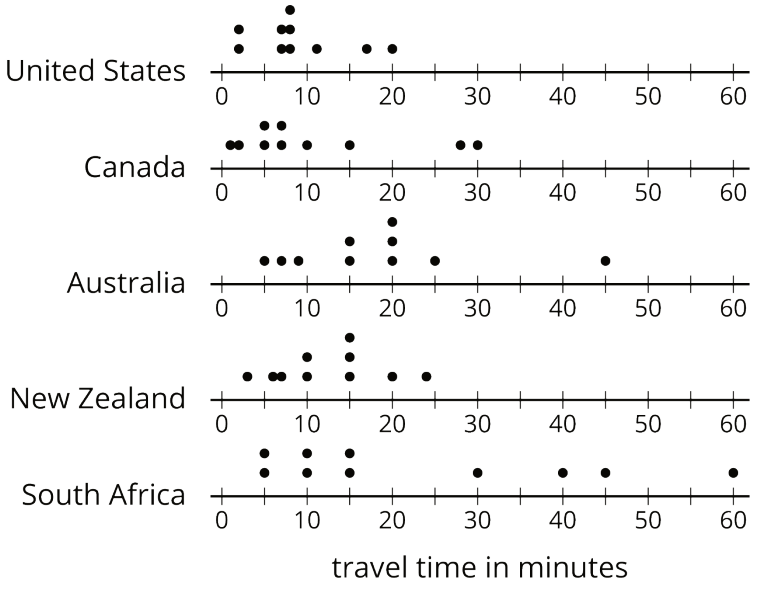
Solution

- A: 2
- B: 3
- C: 1

Problem 2

Statement

The dot plots show the time it takes to get to school for ten sixth-grade students from the United States, Canada, Australia, New Zealand, and South Africa.



a. List the countries in order of *typical travel times*, from shortest to longest.

b. List the countries in order of *variability in travel times*, from the least variability to the greatest.

Solution

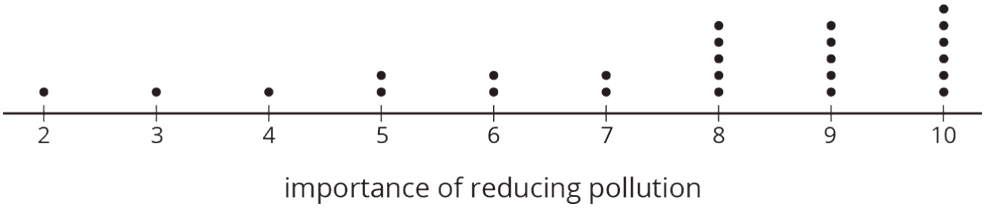
a. U.S., Canada, New Zealand, Australia, South Africa. (The centers for Canada and New Zealand are close, so students may have trouble ordering them. That is acceptable and will be used later to motivate the need for a numerical measure of center, e.g., mean or median.)

b. U.S., New Zealand, Australia, Canada, South Africa. (The spread for the U.S. and New Zealand and the spreads for Canada and Australia are close, so students may have trouble ordering them. That is acceptable and related to the need for numerical measures of variability, e.g., mean absolute deviation or interquartile range, which is the topic of later lessons.)

Problem 3

Statement

Twenty-five students were asked to rate—on a scale of 0 to 10—how important it is to reduce pollution. A rating of 0 means "not at all important" and a rating of 10 means "very important." Here is a dot plot of their responses.



Explain why a rating of 6 is *not* a good description of the center of this data set.

Solution

Responses vary. Sample response: Although 6 is halfway between the largest and the smallest numbers in the data set, most of the values in the data set are larger than 6. Only 7 of the 25 values are less than or equal to 6, but 20 of the data values are greater than or equal to 6.

Problem 4

Statement

Tyler wants to buy some cherries at the farmer's market. He has $10 and cherries cost $4 per pound.

a. If c is the number of pounds of cherries that Tyler can buy, write one or more inequalities or equations describing c.

b. Can 2 be a value of c? Can 3 be a value of c? What about -1? Explain your reasoning.

c. If m is the amount of money, in dollars, Tyler can spend, write one or more inequalities or equations describing m.

d. Can 8 be a value of m? Can 2 be a value of m? What about 10.5? Explain your reasoning.

Solution

a. Answer varies. Sample response: The inequality $c < 2.5$, or $c = 2.5$, says that Tyler cannot spend more than $10 for the cherries. The inequality $c > 0$ means that Tyler actually buys some cherries.

b. Yes, 2 can be a solution because 2 pounds of cherries cost $8. No, 3 could not be a solution because Tyler could not buy 3 pounds of cherries as they would cost $12 and he only has $10. No, -1 could not be a solution because it does not make sense for Tyler to buy -1 pounds of cherries.

c. Answers vary. Sample response: $m < 10$ or $m = 10$ (Tyler can spend $10 or less), and $m > 0$ (Noah spends some money).

d. Yes, 8 and 2 can both be values of m. The value 8 means Tyler buys 2 pounds of cherries ($2 \cdot 4 = 8$), and 2 means he buys $\frac{1}{2}$ pound of cherries ($\frac{1}{2} \cdot 4 = 2$). No, 10.5 cannot be a value of m in this case. Even though 10.5 is greater than 0, m must also be 10 or less.

(From Unit 7, Lesson 10.)