Plan for Unit Grade 7 Unit 8: Probability and Sampling

Relevant Unit(s) to review: Grade 6, Unit 8: Data Sets and Distributions

Essential prior concepts to engage with this unit	• interpreting dot plots, histograms, and box plots • describing distributions using measures of center (mean and median) and measures of variability (mean absolute deviation and interquartile range)
Brief narrative of approach	Begin by looking at data representations (6.8.B), then get more formal with their interpretation of typical values and variability (6.8.C, 6.8.D) to describe distributions. This leads to analyzing data when only a sample is accessible (7.8.C).

Lessons to Add	Lessons to Remove or Modify
1. Combine 6.8.2 and 6.8.3: introduce variability and distributions 2. 6.8.5: describe distributions using center and spread 3. Combine 6.8.6 and 6.8.7: focus on interpreting histograms 4. Combine 6.8.9 and 6.8.10: focus on understanding mean 5. Combine 6.8.11 and 6.8.12: focus on understanding MAD 6. Combine 6.8.13 and 6.8.14: focus on understanding median 7. Combine 6.8.15 and 6.8.16: focus on box plots and interquartile range	1. Remove 7.8.1: simulation introduction that has analogs later in the unit 2. Remove 7.8.5: additional practice that could be done outside of class 3. Remove 7.8.11: reminder of mean and median that will be more fresh after doing 6.8 lessons 4. Remove 7.8.16: inferring another measure from the population based on a sample 5. Remove 7.8.17: an optional lesson examining how sample size affects the variability of sample means 6. Remove 7.8.19: a lesson comprised primarily of an info gap that practices comparing populations 7. Move to outside of class 7.8.20: culminating lesson incorporating work from the unit
Lessons added: 7	Lessons removed: 7

Modified Plan for Grade 7 Unit 8

Day	IM lesson	Notes
	assessment	6.8 Check Your Readiness assessment Note that the Check Your Readiness assessment includes item-by-item guidance to inform just-in-time adjustments to instruction within the lessons in 7.8.
1	6.8.2 6.8.3	Use Activity 6.8.2.2 to begin a conversation about variability and Activity 6.8.3.3 to remind students of how dot plots can be used to examine distributions.
2	6.8.5	Describing distributions using center and spread.
3	7.8.2	
4	7.8.3	
5	7.8.4	
6	7.8.6	
7	7.8.7	
8	7.8.8	
9	7.8.9	
10	7.8.10	
11	6.8.6 6.8.7	If the initial assessment shows that students are not familiar with interpreting histograms, consider including activities from these lessons before continuing with grade-level content.
12	6.8.9 6.8.10	If the initial assessment shows that students are not familiar with the mean, consider including activities from these lessons before continuing with grade-level content.

13	6.8.11 6.8.12	If the initial assessment shows that students are not familiar with mean absolute deviation, consider including activities from these lessons before continuing with grade-level content.
14	6.8.13 6.8.14	If the initial assessment shows that students are not familiar with the median, consider including activities from these lessons before continuing with grade-level content.
15	6.8.15 6.8.16	If the initial assessment shows that students are not familiar with box plots and interquartile range, consider including activities from these lessons before continuing with grade-level content.
16	7.8.12	
17	7.8.13	
18	7.8.14	
19	7.8.18	
20	7.8 End Assessment	

Priority and Category List for Lessons

High priority (+), Medium priority (0), Low priority (-)

E: Explore, Play, and Discuss, D: Deep Dive, A: Synthesize and Apply

Lesson	Priority (+, 0, -)	Category (E, D, A)	Notes
7.8.1	0	E	To introduce the unit on probability, students play a game to collect data about what is inside bags and then make a decision based on the information they have collected. The process of using previous results from repeated trials to inform the likelihood of future events is one way to estimate probabilities that will be revisited later.
7.8.2	+	E	In this lesson, students investigate chance **events**. They use language like impossible, unlikely, equally likely as not, likely, or certain to describe a likelihood of a **chance event**.

7.8.3	+	D	In this lesson, students begin to assign probabilities to chance events. They understand that the greater the **probability**, the more likely the event will occur. They define an **outcome** as a possible result for a **chance experiment**. They learn that the **sample space** is the set of all possible outcomes, and understand that a process is called **random** when the outcome of an experiment is based on chance. They reason that if there are n equally likely outcomes for a chance experiment, they construct the argument that the probability of each of these outcomes is $\frac{1}{n}$.
7.8.4	+	D	In this lesson, students roll a number cube many times and calculate the cumulative fraction of the time that an event occurs to see that in the long run this relative frequency approaches the probability of the chance event.
7.8.5	+	A	In this lesson, students compare the results from running actual trials of an experiment to the expected, calculated probabilities. They also use their data to see that additional trials usually produce more accurate results as minor differences even out after many trials.
7.8.6	+	A	This lesson introduces the idea of **simulation**. Different groups of students use different chance experiments that are designed to enable them to approximate the probability of a real-world event. Students follow a process similar to what they used in previous lessons for calculating relative frequencies. The distinction in this lesson is that the outcomes students are tracking are from an experiment designed to represent the outcome of some other experiment that would be harder to study directly.
7.8.7	0	E	In this lesson, students see that compound events can be simulated by using multiple chance experiments. In this case, it is important for students to communicate precisely what represents one outcome of the simulation.
7.8.8	+	D	In this lesson, students practice listing the sample space for a compound event. They make use of the structure of tree diagrams, tables, and organized lists as methods of organizing this information.
7.8.9	0	D	In this lesson, students continue writing out the sample spaces for chance experiments that have multiple steps and also begin using those sample spaces to calculate the probability of certain events. Students may start listing the sample space using one method and then decide to switch to a different method when they get stuck in the middle

			of the problem or they might recognize certain aspects of the situation that would lead them to choose a particular method from the beginning.
7.8.10	+	A	In this lesson, students are shown that the probability of compound events can also be estimated using simulations.
7.8.11	0	E	In this lesson, students review measures of center and variability from grade 6. They also work at deciding whether or not two distributions are very different from each other. This lesson introduces the idea of expressing the difference between the centers of two distributions as a multiple of a measure of variability as a way to help students make this determination.
7.8.12	+	D	This lesson introduces the idea of using data from a sample of a **population** when it is impractical or impossible to gather data from every individual in the populations under study.
7.8.13	+	D	In this lesson, students examine multiple samples of the same population and learn what it means for a sample to be **representative** of the population. Students look at the structure of dot plots, attending to center, shape, and spread, to help them compare the samples and the population.
7.8.14	+	A	In this lesson, students consider different methods of selecting a sample. Students begin by critiquing different sampling methods for their benefits and drawbacks. In particular, students notice that some sampling methods are more biased than others.
7.8.15	+	E	In this lesson students calculate measures of center and variation for samples from different populations and consider the meaning of these quantities in terms of the situation.
7.8.16	0	E	In the previous lesson, students used samples to estimate measures of the center of a population. In this lesson, students estimate **population proportions**.
7.8.17	-	D	This lesson is optional. It goes beyond necessary grade-level standards to examine the accuracy of estimates for population characteristics based on many samples. The lesson builds a solid foundation for future grades to build upon, but may be shortened or skipped

			due to time constraints.
7.8.18	+	A	In previous lessons, students examined the distributions of two entire populations to decide whether or not they were very different. In this lesson, students use samples to make comparative inferences about populations.
7.8.19	+	A	Students continue to practice comparing populations by using samples from each population.
7.8.20	-	A	This lesson is optional. In this lesson, students apply what they have learned about probability, sampling, and comparing populations to analyze two data sets.

Lesson 2: Statistical Questions

2.1: Pencils on A Plot

1. Measure your pencil to the nearest $\frac{1}{4}$-inch. Then, plot your measurement on the class dot plot.

2. What is the difference between the longest and shortest pencil lengths in the class?

3. What is the most common pencil length?

4. Find the difference in lengths between the most common length and the shortest pencil.

2.2: What’s in the Data?

Ten sixth-grade students at a school were each asked five survey questions. Their answers to each question are shown here.

data set A	0	1	1	3	0	0	0	2	1	1
data set B	12	12	12	12	12	12	12	12	12	12
data set C	6	5	7	6	4	5	3	4	6	8
data set D	6	6	6	6	6	6	6	6	6	6
data set E	3	7	9	11	6	4	2	16	6	10

1. Here are the five survey questions. Match each question to a data set that could represent the students’ answers. Explain your reasoning.
    - Question 1: Flip a coin 10 times. How many heads did you get?
    - Question 2: How many books did you read in the last year?
    - Question 3: What grade are you in?
    - Question 4: How many dogs and cats do you have?
    - Question 5: How many inches are in 1 foot?
2. How are survey questions 3 and 5 different from the other questions?

2.3: What Makes a Statistical Question?

These three questions are examples of **statistical questions**:

- What is the most common color of the cars in the school parking lot?
- What percentage of students in the school have a cell phone?
- Which kind of literature—fiction or nonfiction—is more popular among students in the school?

These three questions are not examples of **statistical questions**:

- What color is the principal's car?
- Does Elena have a cell phone?
- What kind of literature—fiction or nonfiction—does Diego prefer?

1. Study the examples and non-examples. Discuss with your partner:
    a. How are the three statistical questions alike? What do they have in common?
    b. How are the three non-statistical questions alike? What do they have in common?
    c. How can you find answers to the statistical questions? How about answers to non-statistical questions?
    d. What makes a question a statistical question?

Pause here for a class discussion.

2. Read each question. Think about the data you might collect to answer it and whether you expect to see **variability** in the data. Complete each blank with “Yes” or “No.”

   a. How many cups of water do my classmates drink each day?

      - Is variability expected in the data? ______
      - Is the question statistical? _____

   b. Where in town does our math teacher live?

      - Is variability expected in the data? ______
      - Is the question statistical? _____

   c. How many minutes does it take students in my class to get ready for school in the morning?

      - Is variability expected in the data? ______
      - Is the question statistical? _____

   d. How many minutes of recess do sixth-grade students have each day?

      - Is variability expected in the data? ______
      - Is the question statistical? _____

   e. Do all students in my class know what month it is?

      - Is variability expected in the data? ______
      - Is the question statistical? _____

2.4: Sifting for Statistical Questions

1. Your teacher will give you and your partner a set of cards with questions. Sort them into three piles: Statistical Questions, Not Statistical Questions, and Unsure.

2. Compare your sorting decisions with another group of students. Start by discussing the two piles that your group sorted into the Statistical Questions and Not Statistical Questions piles. Then, review the cards in the Unsure pile. Discuss the questions until both groups reach an agreement and have no cards left in the Unsure pile. If you get stuck, think about whether the question could be answered by collecting data and if there would be variability in that data.

3. Record the letter names of the questions in each pile.

    - Statistical questions:
    - Non-statistical questions:

Are you ready for more?

Tyler and Han are discussing the question, “Which sixth-grade student lives the farthest from school?”

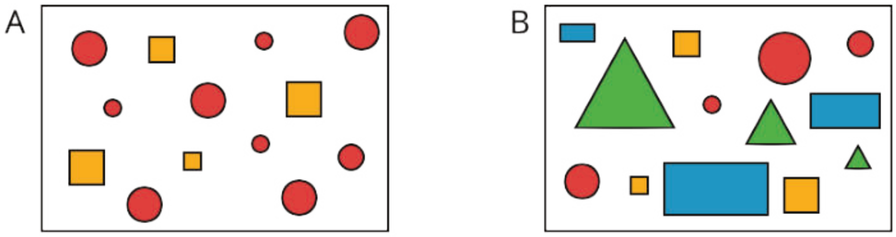
- Tyler says, “I don’t think the question is a statistical question. There is only one person who lives the farthest from school, so there would not be variability in the data we collect.”
- Han says: “I think it is a statistical question. To answer the question, we wouldn’t actually be asking everyone, 'Which student lives the farthest from school?' We would have to ask each student how far away from school they live, and we can expect their responses to have variability.”

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

Lesson 2 Summary

We often collect data to answer questions about something. The data we collect may show **variability**, which means the data values are not all the same.

Some data sets have more variability than others. Here are two sets of figures.



Set A has more figures with the same shape, color, or size. Set B shows more figures with different shapes, colors, or sizes, so set B has greater variability than set A.

Both numerical and categorical data can show variability. Numerical sets can contain different numbers, and categorical sets can contain different categories or types.

When a question can only be answered by using data and we expect that data to have variability, we call it a **statistical question**. Here are some examples.

- Who is the most popular musical artist at your school?
- When do students in your class typically eat dinner?
- Which classroom in your school has the most books?

To answer the question about books, we may need to count all of the books in each classroom of a school. The data we collect would likely show variability because we would expect each classroom to have a different number of books.

In contrast, the question “How many books are in your classroom?” is *not* a statistical question. If we collect data to answer the question (for example, by asking everyone in the class to count books), the data can be expected to show the same value. Likewise, if we ask all of the students at a school where they go to school, that question is not a statistical question because the responses will all be the same.

Sifting for Statistical Questions

A. What fraction of the people in your school wear glasses?

Sifting for Statistical Questions

B. How many centimeters tall is the door of your classroom?

Sifting for Statistical Questions

C. What percentage of students in your class say they like dogs better than cats?

Sifting for Statistical Questions

D. What is the average age of the teachers at your school?

Sifting for Statistical Questions

E. What do students in your class prefer to have on a hot dog: ketchup, mustard, both ketchup and mustard, or neither?

Sifting for Statistical Questions

F. What is a typical number of students per class in your school?

Sifting for Statistical Questions

G. Who is the current Vice President of the United States?

Sifting for Statistical Questions

H. How old is the principal at your school?

Sifting for Statistical Questions

I. What is the most common favorite color for the students in your class?

Sifting for Statistical Questions

J. Do students at your school generally get more sleep on school nights than on weekends and holidays?

Sifting for Statistical Questions

K. Who is is the oldest staff member at the school?

Sifting for Statistical Questions

L. What day of the week is today?

Lesson 2: Statistical Questions

Cool Down: Questions about Temperature

Here are two questions:

Question A: Over the past 10 years, what is the warmest temperature recorded, in degrees Fahrenheit, for the month of December in Miami, Florida?

Question B: At what temperature does water freeze in Miami, Florida?

1. Decide if each question is statistical or non-statistical. Explain your reasoning.

2. If you decide that a question is statistical, describe how you would find the answer. What data would you collect?

Unit 8 Lesson 2 Cumulative Practice Problems

1. Sixth-grade students were asked, "What grade are you in?" Explain why this is *not* a statistical question.

2. Lin and her friends went out for ice cream after school. The following questions came up during their trip. Select **all** the questions that are statistical questions.

   A. How far are we from the ice cream shop?

   B. What is the most popular ice cream flavor this week?

   C. What does a group of 4 people typically spend on ice cream at this shop?

   D. Do kids usually prefer to get a cup or a cone?

   E. How many toppings are there to choose from?

3. Here is a list of questions about the students and teachers at a school. Select **all** the questions that are statistical questions.

   A. What is the most popular lunch choice?

   B. What school do these students attend?

   C. How many math teachers are in the school?

   D. What is a common age for the teachers at the school?

   E. About how many hours of sleep do students generally get on a school night?

   F. How do students usually travel from home to school?

4. Here is a list of statistical questions. What data would you collect and analyze to answer each question? For numerical data, include the unit of measurement that you would use.

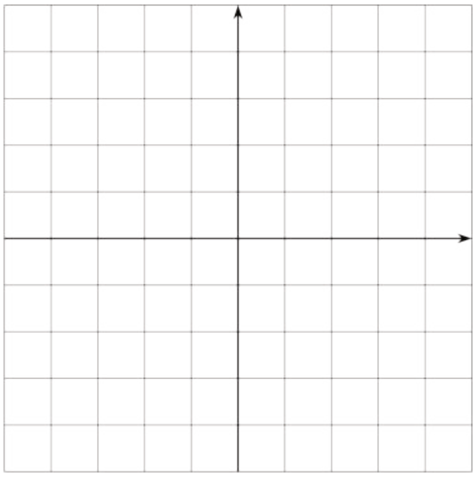
    a. What is a typical height of female athletes on a team in the most recent international sporting event?

    b. Are most adults in the school football fans?

    c. How long do drivers generally need to wait at a red light in Washington, DC?

5. Describe the scale you would use on the coordinate plane to plot each set of points. What value would you assign to each unit of the grid?

    a. (1, -6), (-7, -8), (-3, 7), (0, 9)

    b. (-20, -30), (-40, 10), (20, -10), (5, -20)

    c. $(\frac{-1}{3}, -1), (\frac{2}{3}, -1\frac{1}{3}), (\frac{-4}{3}, \frac{2}{3}), (\frac{1}{6}, 0)$

(From Unit 7, Lesson 13.)

6. Noah's water bottle contains more than 1 quart of water but less than $1\frac{1}{2}$ quarts. Let w be the amount of water in Noah's bottle, in quarts. Select **all** the true statements.

   A. w could be $\frac{3}{4}$.

   B. w could be 1.

   C. $w > 1$

   D. w could be $\frac{4}{3}$.

   E. w could be $\frac{5}{4}$.

   F. w could be $\frac{5}{3}$.

   G. $w > 1.5$

   (From Unit 7, Lesson 9.)

7. Order these numbers from least to greatest:

   $|\text{-}17|$ $|\text{-}18|$ -18 $|19|$ 20

   (From Unit 7, Lesson 7.)

Lesson 3: Representing Data Graphically

3.1: Curious about Caps

Clare collects bottle caps and keeps them in plastic containers.

Write one statistical question that someone could ask Clare about her collection. Be prepared to explain your reasoning.

3.2: Estimating Caps

1. Write down the statistical question your class is trying to answer.

2. Look at the dot plot that shows the data from your class. Write down one thing you notice and one thing you wonder about the dot plot.

3. Use the dot plot to answer the statistical question. Be prepared to explain your reasoning.

3.3: Been There, Done That!

Priya wants to know if basketball players on a men's team and a women's team have had prior experience in international competitions. She gathered data on the number of times the players were on a team before 2016.

men's team

3 0 0 0 0 1 0 0 0 0 0 0

women's team

2 3 3 1 0 2 0 1 1 0 3 1

1. Did Priya collect categorical or numerical data?
2. Organize the information on the two basketball teams into these tables.

Men's Basketball Team Players

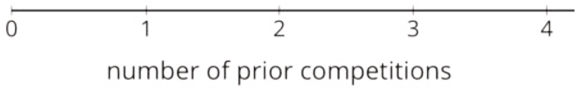
number of prior competitions	frequency (number)
0	
1	
2	
3	
4	

Women's Basketball Team Players

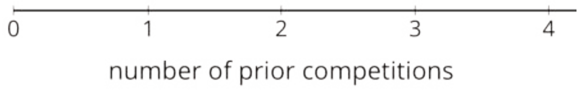
number of prior competitions	frequency (number)
0	
1	
2	
3	
4	

3. Make a dot plot for each table.

Men's Basketball Team Players



Women's Basketball Team Players



4. Study your dot plots. What do they tell you about the competition participation of:

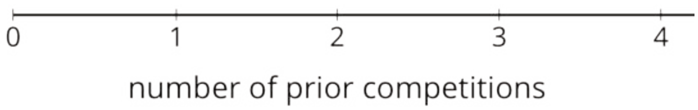
    a. the players on the men's basketball team?

    b. the players on the women's basketball team?

5. Explain why a dot plot is an appropriate representation for Priya's data.

Are you ready for more?

Combine the data for the players on the men's and women's teams and represent it as a single dot plot. What can you say about the repeat participation of the basketball players?



3.4: Favorite Summer Sports

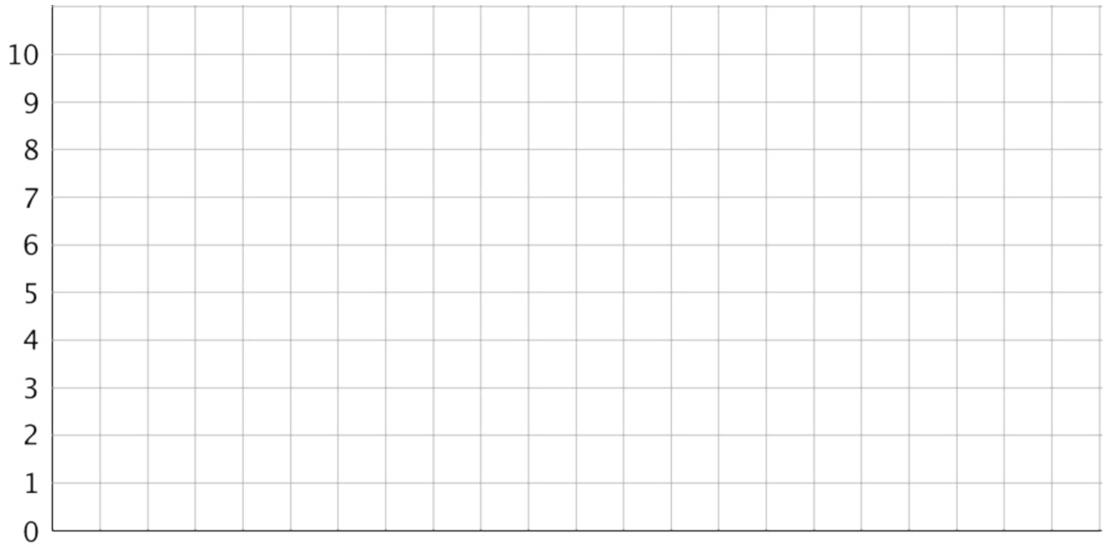
Kiran wants to know which three summer sports are most popular in his class. He surveyed his classmates on their favorite summer sport. Here are their responses.

swimming	gymnastics	track and field	volleyball
swimming	swimming	diving	track and field
gymnastics	basketball	basketball	volleyball
track and field	track and field	volleyball	gymnastics
diving	gymnastics	volleyball	rowing
track and field	track and field	soccer	swimming
gymnastics	track and field	swimming	rowing
diving	soccer		

1. Did Kiran collect categorical or numerical data?
2. Organize the responses in a table to help him find which summer sports are most popular in his class.

sport	frequency

3. Represent the information in the table as a bar graph.



4. a. How can you use the bar graph to find how many classmates Kiran surveyed?

b. Which three summer sports are most popular in Kiran's class?

c. Use your bar graph to describe at least one observation about Kiran's classmates' preferred summer sports.

5. Could a dot plot be used to represent Kiran's data? Explain your reasoning.

Lesson 3 Summary

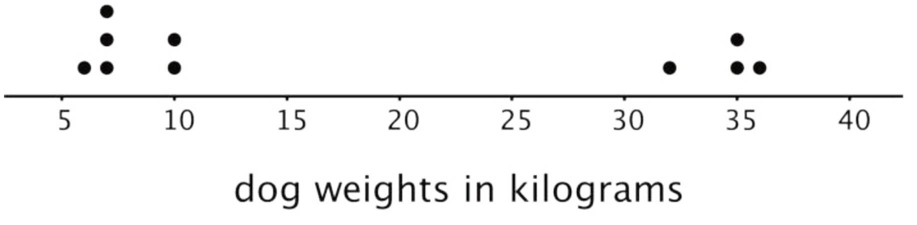
When we analyze data, we are often interested in the **distribution**, which is information that shows all the data values and how often they occur.

In a previous lesson, we saw data about 10 dogs. We can see the distribution of the dog weights in a table such as this one.

weight in kilograms	frequency
6	1
7	3
10	2
32	1
35	2
36	1

The term **frequency** refers to the number of times a data value occurs. In this case, we see that there are three dogs that weigh 7 kilograms, so “3” is the frequency for the value “7 kilograms.”

Recall that dot plots are often used to to represent numerical data. Like a frequency table, a dot plot also shows the distribution of a data set. This dot plot, which you saw in an earlier lesson, shows the distribution of dog weights.

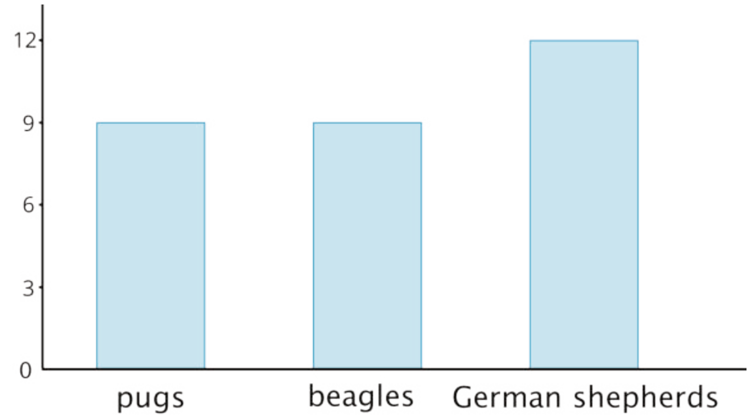


A dot plot uses a horizontal number line. We show the frequency of a value by the number of dots drawn above that value. Here, the two dots above the number 35 tell us that there are two dogs weighing 35 kilograms.

The distribution of categorical data can also be shown in a table. This table shows the distribution of dog breeds.

breed	frequency
pug	9
beagle	9
German shepherd	12

We often represent the distribution of categorical data using a bar graph.



A bar graph also uses a horizontal line. Above it we draw a rectangle (or "bar") to represent each category in the data set. The height of a bar tells us the frequency of the category. There are 12 German shepherds in the data set, so the bar for this category is 12 units tall. Below the line we write the labels for the categories.

In a dot plot, a data value is placed according to its position on the number line. A weight of 10 kilograms must be shown as a dot above 10 on the number line.

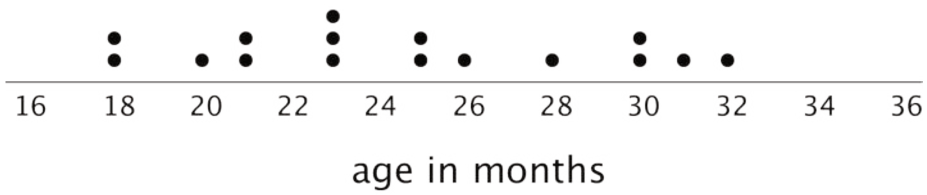
In a bar graph, however, the categories can be listed in any order. The bar that shows the frequency of pugs can be placed anywhere along the horizontal line.

Lesson 3: Representing Data Graphically

Cool Down: Swimmers and Swimming Class

1. Noah gathered information on the home states of the swimmers on Team USA. He organized the data in a table. Would a dot plot be appropriate to display his data? Explain your reasoning.

2. This dot plot shows the ages of students in a swimming class. How many students are in the class?



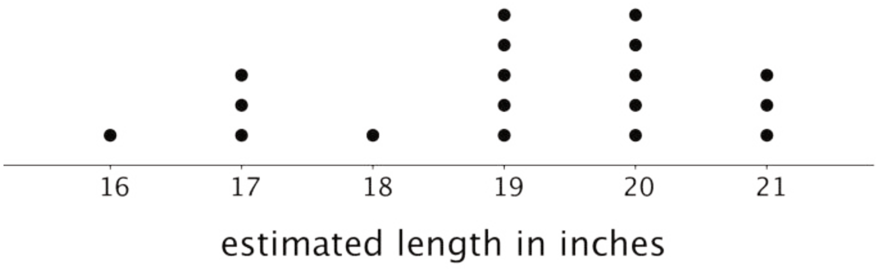
3. Based on the dot plot, do you agree with each of the following statements? Explain your reasoning.

   a. The class is an adult swimming class.

   b. Half of the students are between 2 and 3 years old.

Unit 8 Lesson 3 Cumulative Practice Problems

1. A teacher drew a line segment that was 20 inches long on the blackboard. She asked each of her students to estimate the length of the segment and used their estimates to draw this dot plot.



   a. How many students were in the class?

   b. Were students generally accurate in their estimates of the length of the line? Explain your reasoning.

2. Here are descriptions of data sets. Select **all** descriptions of data sets that could be graphed as dot plots.

   A. Class size for the classes at an elementary school

   B. Colors of cars in a parking lot

   C. Favorite sport of each student in a sixth-grade class

   D. Birth weights for the babies born during October at a hospital

   E. Number of goals scored in each of 20 games played by a school soccer team

3. Priya recorded the number of attempts it took each of 12 of her classmates to successfully throw a ball into a basket. Make a dot plot of Priya's data.

   1 2 1 3 1 4 4 3 1 2 5 2

4. Solve each equation.

   a. $9v = 1$

   b. $1.37w = 0$

   c. $1 = \frac{7}{10}x$

   d. $12.1 = 12.1 + y$

   e. $\frac{3}{5} + z = 1$

(From Unit 6, Lesson 4.)

5. Find the quotients.

   a. $\frac{2}{5} \div 2$

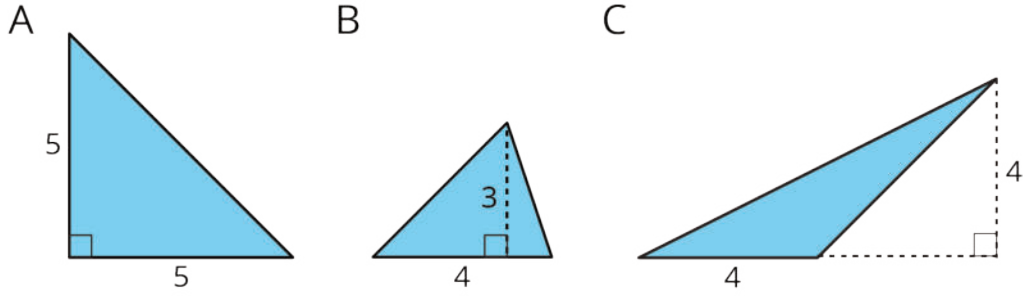
   b. $\frac{2}{5} \div 5$

   c. $2 \div \frac{2}{5}$

   d. $5 \div \frac{2}{5}$

(From Unit 4, Lesson 11.)

6. Find the area of each triangle.

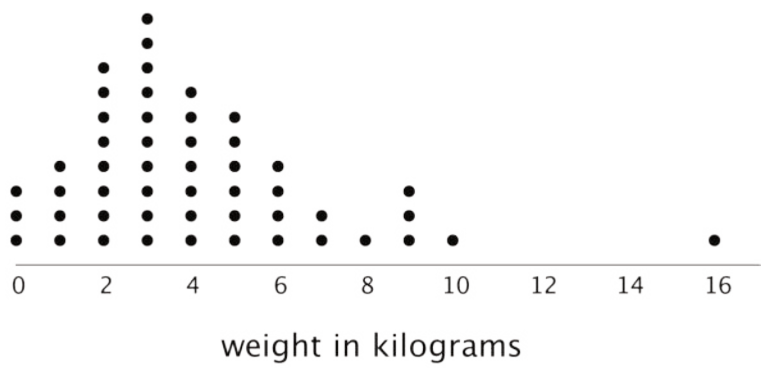


(From Unit 1, Lesson 9.)

Lesson 5: Using Dot Plots to Answer Statistical Questions

5.1: Packs on Backs

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?

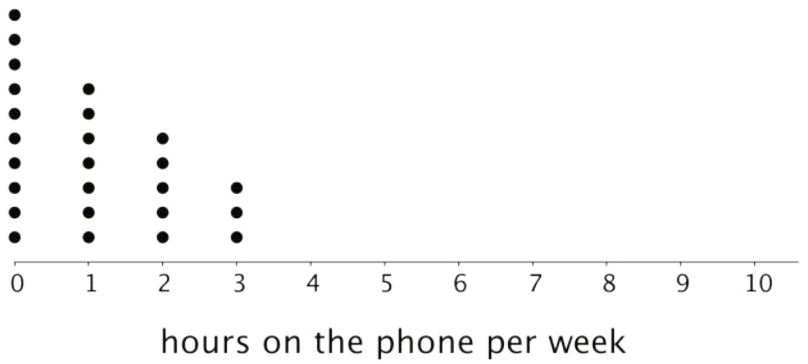
2. Clare and Tyler studied the dot plot.

    - 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.

5.2: On the Phone

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?

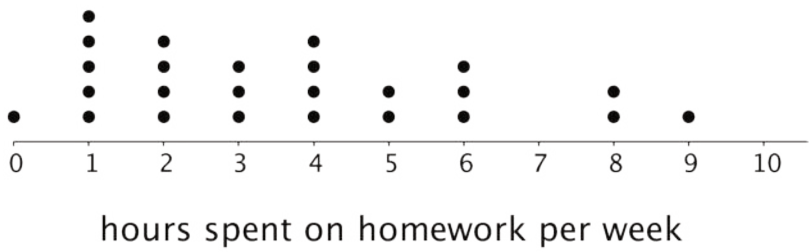
   b. What percentage of the group reported talking on the phone for this amount of time?

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.



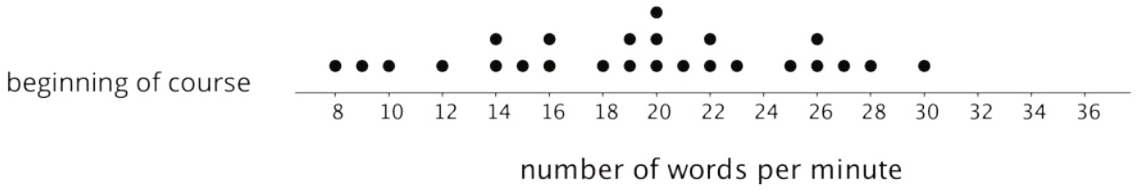
   Overall, are these students more alike in the amount of time they spend talking on the phone or in the amount of time they spend on homework? Explain your reasoning.

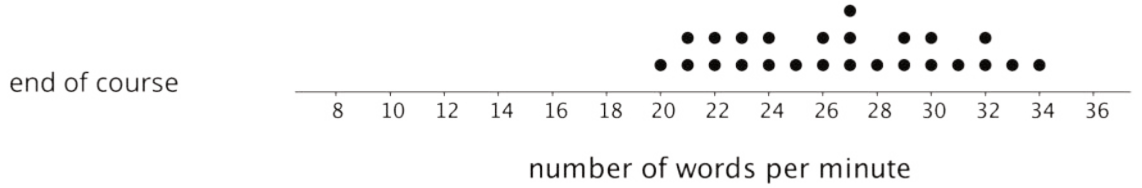
5. Suppose someone claimed that these sixth-grade students spend too much time on the phone. Do you agree? Use your analysis of the dot plot to support your answer.

5.3: Click-Clack

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?

Are you ready for more?

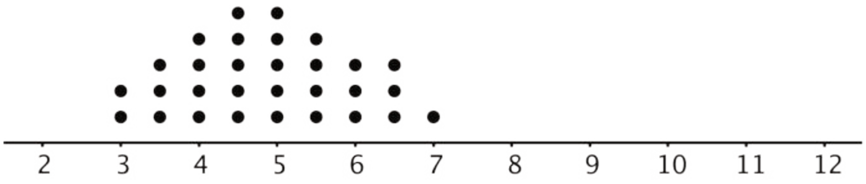
Use one of these suggestions (or make up your own). Research to create a dot plot with at least 10 values. Then, describe the center and spread of the distribution.

- 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

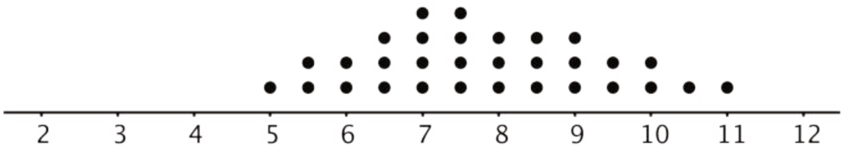
Lesson 5 Summary

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.

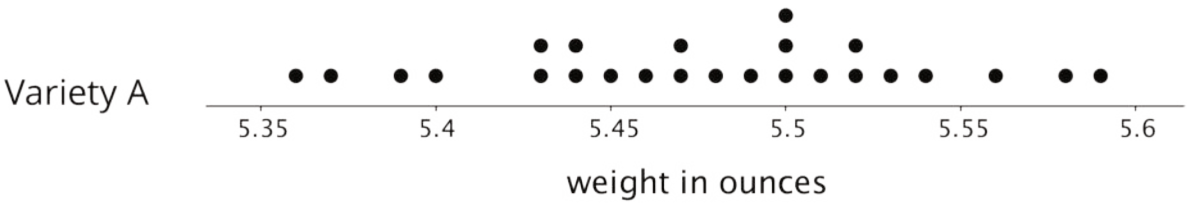
A distribution with greater spread tells us that the data have greater variability. In this case, we could say that the cats are more similar in their weights than the dogs.

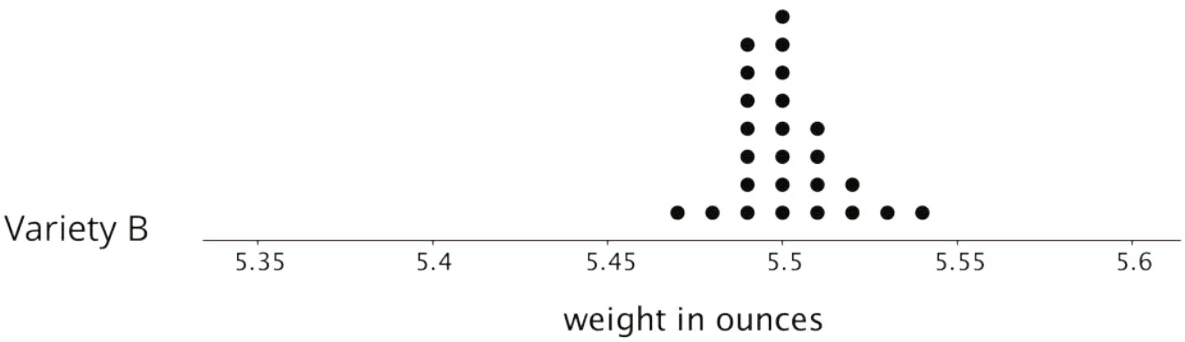
In future lessons, we will discuss how to measure the center and spread of a distribution.

Lesson 5: Using Dot Plots to Answer Statistical Questions

Cool Down: Packing Tomatoes

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.



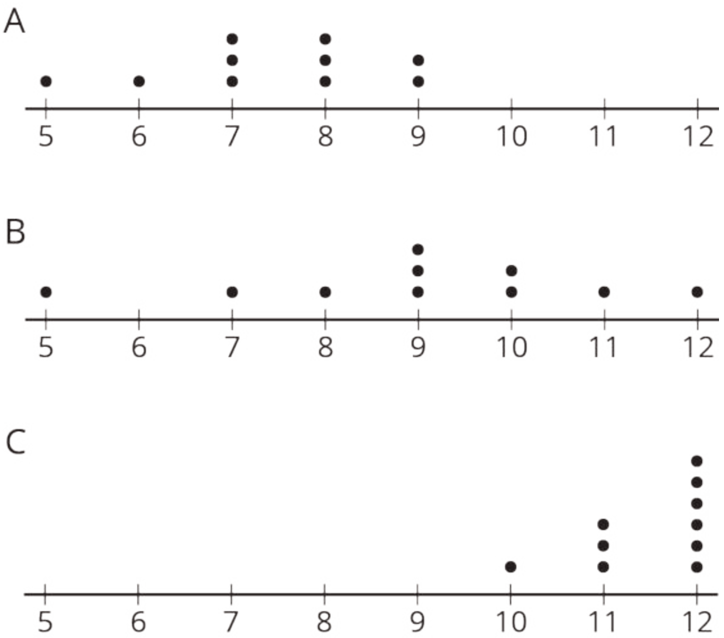


1. What would be a good description for the weight of Variety A tomatoes, in general? What about for the weight of Variety B tomatoes, in general?

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

Unit 8 Lesson 5 Cumulative Practice Problems

1. 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.



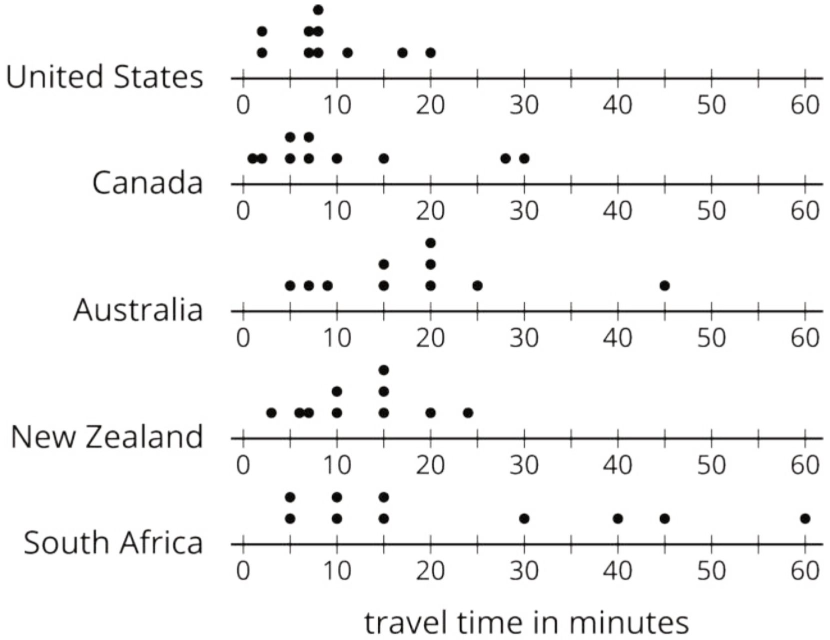
A. Dot Plot A

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

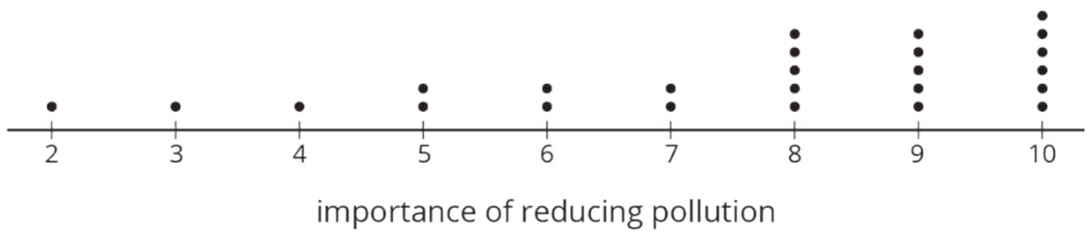
2. The dot plots show the time it takes to get to school for ten sixth-grade students from the United States, Canada, Australia, New Zealand, and South Africa.



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.

3. 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.

4. 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.

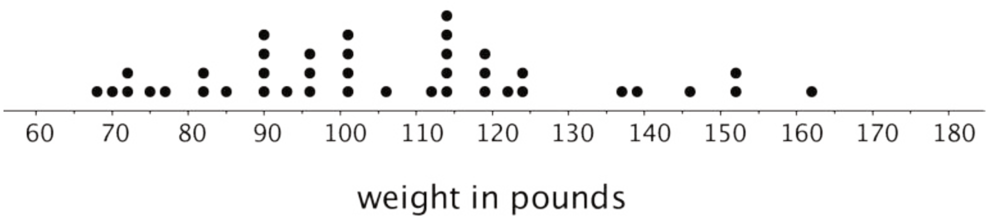
   d. Can 8 be a value of m? Can 2 be a value of m? What about 10.5? Explain your reasoning.

(From Unit 7, Lesson 10.)

Lesson 6: Histograms

6.1: Dog Show (Part 1)

Here is a dot plot showing the weights, in pounds, of 40 dogs at a dog show.

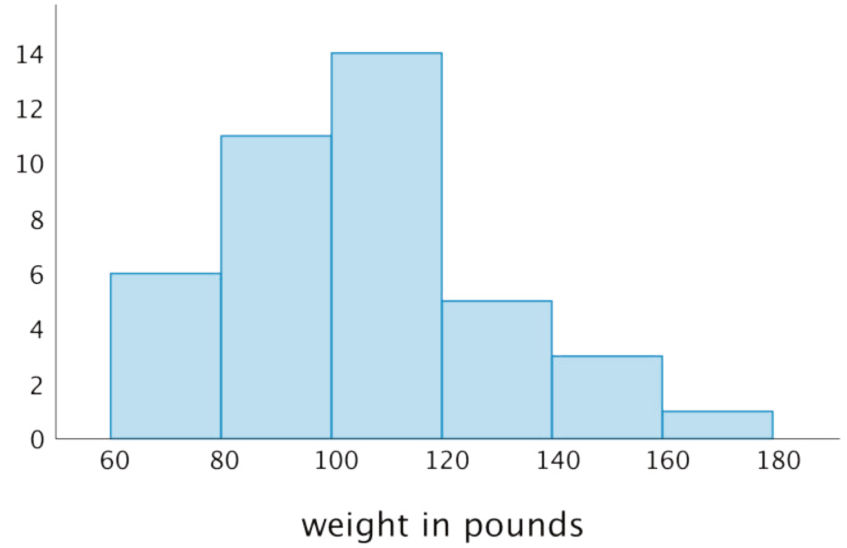


1. Write two statistical questions that can be answered using the dot plot.

2. What would you consider a typical weight for a dog at this dog show? Explain your reasoning.

6.2: Dog Show (Part 2)

Here is a **histogram** that shows some dog weights in pounds.



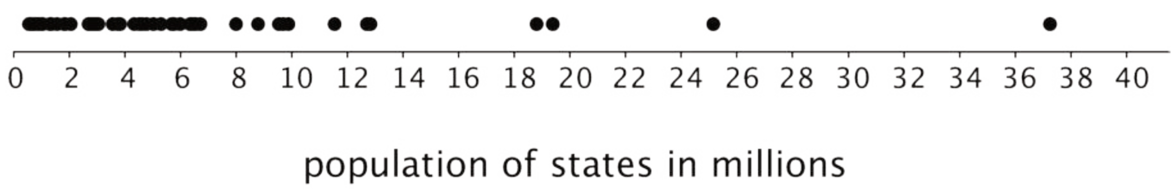
Each bar includes the left-end value but not the right-end value. For example, the first bar includes dogs that weigh 60 pounds and 68 pounds but not 80 pounds.

1. Use the histogram to answer the following questions.
    a. How many dogs weigh at least 100 pounds?
    b. How many dogs weigh exactly 70 pounds?
    c. How many dogs weigh at least 120 and less than 160 pounds?
    d. How much does the heaviest dog at the show weigh?
    e. What would you consider a typical weight for a dog at this dog show? Explain your reasoning.

2. Discuss with a partner:
    - If you used the dot plot to answer the same five questions you just answered, how would your answers be different?
    - How are the histogram and the dot plot alike? How are they different?

6.3: Population of States

Every ten years, the United States conducts a census, which is an effort to count the entire population. The dot plot shows the population data from the 2010 census for each of the fifty states and the District of Columbia (DC).



1. Here are some statistical questions about the population of the fifty states and DC. How difficult would it be to answer the questions using the *dot plot*?

   In the middle column, rate each question with an E (easy to answer), H (hard to answer), or I (impossible to answer). Be prepared to explain your reasoning.

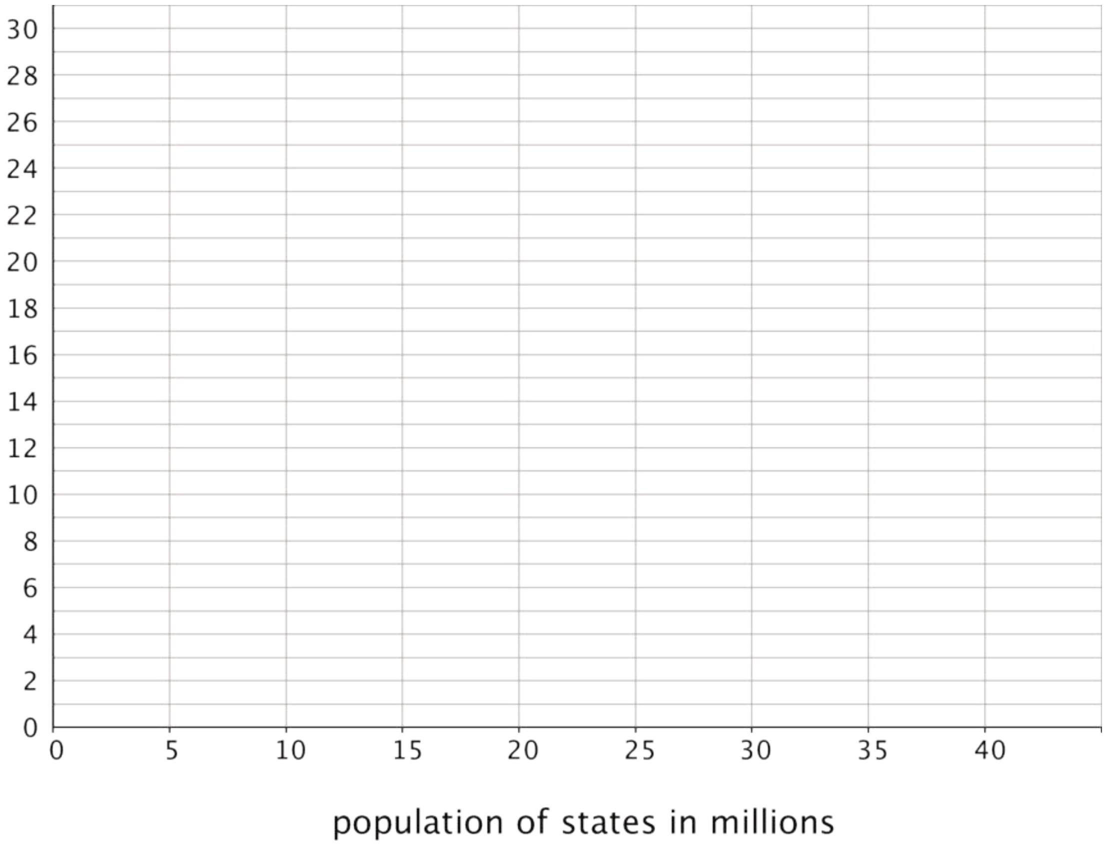
statistical question	using the dot plot	using the histogram
a. How many states have populations greater than 15 million?		
b. Which states have populations greater than 15 million?		
c. How many states have populations less than 5 million?		
d. What is a typical state population?		
e. Are there more states with fewer than 5 million people or more states with between 5 and 10 million people?		
f. How would you describe the distribution of state populations?		

2. Here are the population data for all states and the District of Columbia from the 2010 census. Use the information to complete the table.

Alabama	4.78	Illinois	12.83	Montana	0.99	Rhode Island	1.05
Alaska	0.71	Indiana	6.48	Nebraska	1.83	South Carolina	4.63
Arizona	6.39	Iowa	3.05	Nevada	2.70	South Dakota	0.81
Arkansas	2.92	Kansas	2.85	New Hampshire	1.32	Tennessee	6.35
California	37.25	Kentucky	4.34	New Jersey	8.79	Texas	25.15
Colorado	5.03	Louisiana	4.53	New Mexico	2.06	Utah	2.76
Connecticut	3.57	Maine	1.33	New York	19.38	Vermont	0.63
Delaware	0.90	Maryland	5.77	North Carolina	9.54	Virginia	8.00
District of Columbia	0.60	Massachusetts	6.55	North Dakota	0.67	Washington	6.72
Florida	18.80	Michigan	9.88	Ohio	11.54	West Virginia	1.85
Georgia	9.69	Minnesota	5.30	Oklahoma	3.75	Wisconsin	5.69
Hawaii	1.36	Mississippi	2.97	Oregon	3.83	Wyoming	0.56
Idaho	1.57	Missouri	5.99	Pennsylvania	12.70		

population (millions)	frequency
0–5	
5–10	
10–15	
15–20	
20–25	
25–30	
30–35	
35–40	

3. Use the grid and the information in your table to create a histogram.



4. Return to the statistical questions at the beginning of the activity. Which ones are now easier to answer?

   In the last column of the table, rate each question with an E (easy), H (hard), and I (impossible) based on how difficult it is to answer them. Be prepared to explain your reasoning.

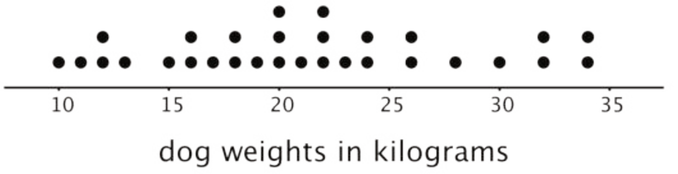
Are you ready for more?

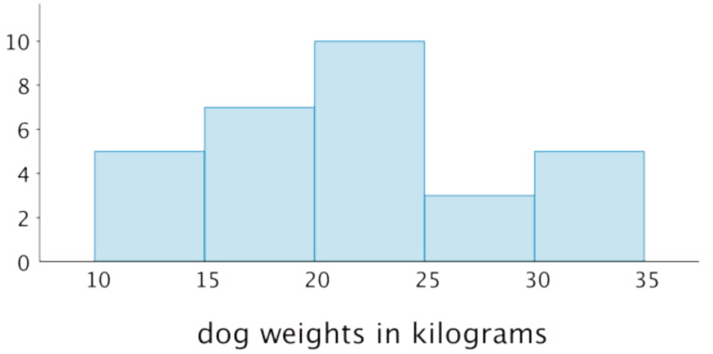
Think of two more statistical questions that can be answered using the data about populations of states. Then, decide whether each question can be answered using the dot plot, the histogram, or both.

Lesson 6 Summary

In addition to using dot plots, we can also represent distributions of numerical data using **histograms**.

Here is a dot plot that shows the weights, in kilograms, of 30 dogs, followed by a histogram that shows the same distribution.



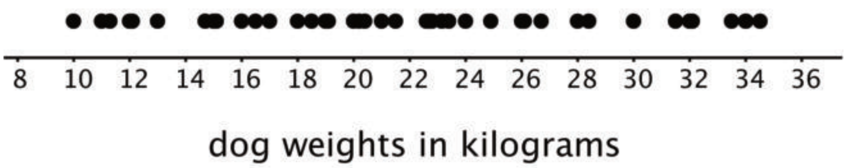


In a histogram, data values are placed in groups or “bins” of a certain size, and each group is represented with a bar. The height of the bar tells us the frequency for that group.

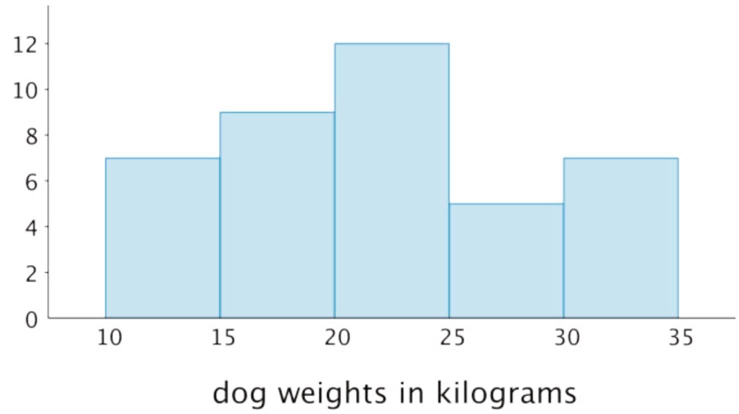
For example, the height of the tallest bar is 10, and the bar represents weights from 20 to less than 25 kilograms, so there are 10 dogs whose weights fall in that group. Similarly, there are 3 dogs that weigh anywhere from 25 to less than 30 kilograms.

Notice that the histogram and the dot plot have a similar shape. The dot plot has the advantage of showing all of the data values, but the histogram is easier to draw and to interpret when there are a lot of values or when the values are all different.

Here is a dot plot showing the weight distribution of 40 dogs. The weights were measured to the nearest 0.1 kilogram instead of the nearest kilogram.



Here is a histogram showing the same distribution.



In this case, it is difficult to make sense of the distribution from the dot plot because the dots are so close together and all in one line. The histogram of the same data set does a much better job showing the distribution of weights, even though we can't see the individual data values.

Lesson 6: Histograms

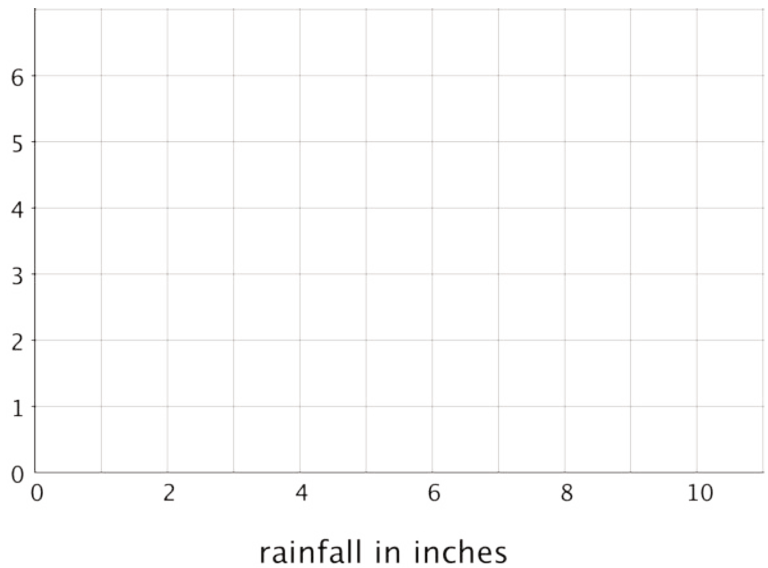
Cool Down: Rain in Miami

Here is the average amount of rainfall, in inches, for each month in Miami, Florida.

month	rainfall (inches)	month	rainfall (inches)
January	1.61	July	6.5
February	2.24	August	8.9
March	2.99	September	9.84
April	3.14	October	6.34
May	5.35	November	3.27
June	9.69	December	2.05

1. Complete the frequency table and use it to make a histogram.

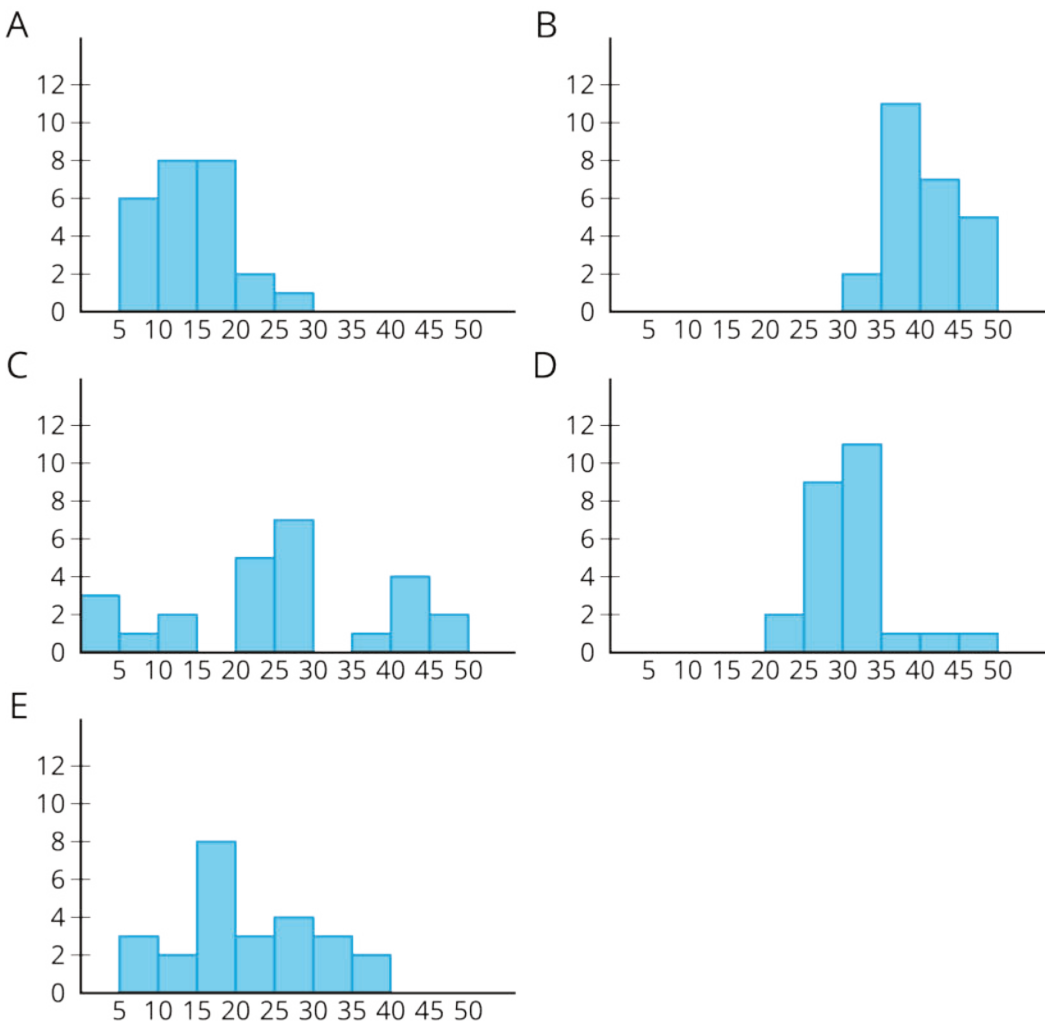
rainfall (inches)	frequency
0–2	
2–4	
4–6	
6–8	
8–10	

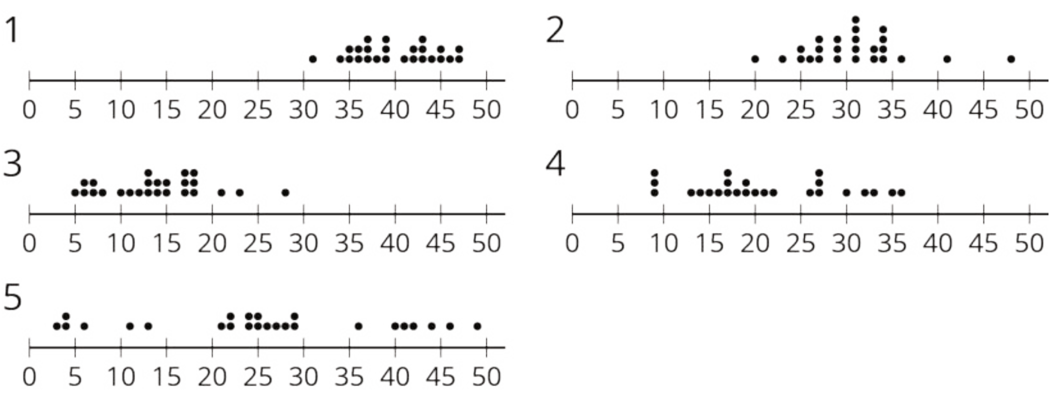


2. What is a typical amount of rainfall in one month in Miami?

Unit 8 Lesson 6 Cumulative Practice Problems

1. Match histograms A through E to dot plots 1 through 5 so that each match represents the same data set.

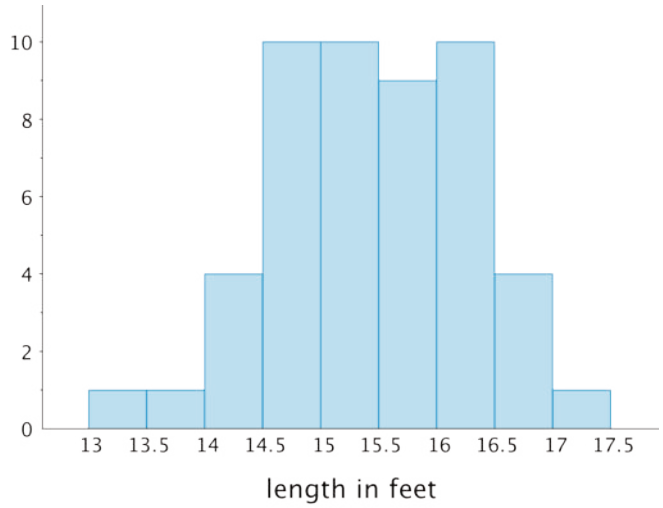




2. $(-2, 3)$ is one vertex of a square on a coordinate plane. Name three points that could be the other vertices.

(From Unit 7, Lesson 12.)

3. Here is a histogram that summarizes the lengths, in feet, of a group of adult female sharks. Select **all** the statements that are true, according to the histogram.



A. A total of 9 sharks were measured.

B. A total of 50 sharks were measured.

C. The longest shark that was measured was 10 feet long.

D. Most of the sharks that were measured were over 16 feet long.

E. Two of the sharks that were measured were less than 14 feet long.

4. This table shows the times, in minutes, it took 40 sixth-grade students to run 1 mile.

time (minutes)	frequency
4 to less than 6	1
6 to less than 8	5
8 to less than 10	13
10 to less than 12	12
12 to less than 14	7
14 to less than 16	2

Draw a histogram for the information in the table.

Lesson 7: Using Histograms to Answer Statistical Questions

7.1: Which One Doesn’t Belong: Questions

Here are four questions about the population of Alaska. Which question does not belong? Be prepared to explain your reasoning.

1. In general, at what age do Alaska residents retire?
2. At what age can Alaskans vote?
3. What is the age difference between the youngest and oldest Alaska residents with a full-time job?
4. Which age group is the largest part of the population: 18 years or younger, 19–25 years, 25–34 years, 35–44 years, 45–54 years, 55–64 years, or 65 years or older?

7.2: Measuring Earthworms

An earthworm farmer set up several containers of a certain species of earthworms so that he could learn about their lengths. The lengths of the earthworms provide information about their ages. The farmer measured the lengths of 25 earthworms in one of the containers. Each length was measured in millimeters.

1. Using a ruler, draw a line segment for each length:
    - 20 millimeters
    - 40 millimeters
    - 60 millimeters
    - 80 millimeters
    - 100 millimeters

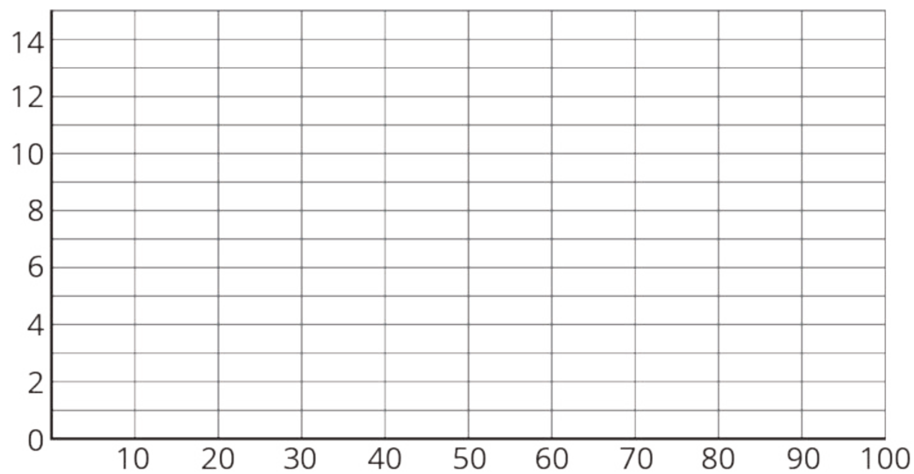
2. Here are the lengths, in millimeters, of the 25 earthworms.

6	11	18	19	20	23	23	25
25	26	27	27	28	29	32	33
41	42	48	52	54	59	60	77
93							

Complete the table for the lengths of the 25 earthworms.

length	frequency
0 millimeters to less than 20 millimeters	
20 millimeters to less than 40 millimeters	
40 millimeters to less than 60 millimeters	
60 millimeters to less than 80 millimeters	
80 millimeters to less than 100 millimeters	

3. Use the grid and the information in the table to draw a histogram for the worm length data. Be sure to label the axes of your histogram.

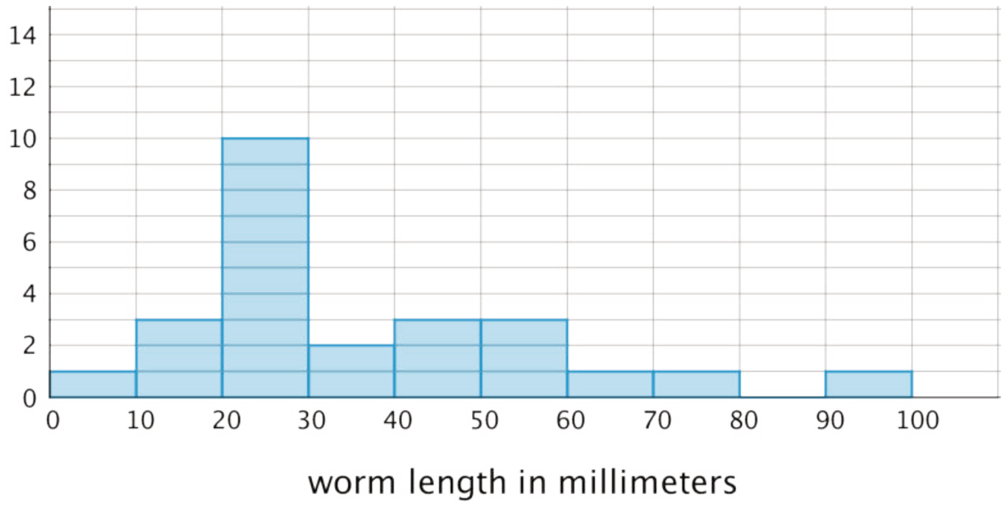


4. Based on the histogram, what is a typical length for these 25 earthworms? Explain how you know.

5. Write 1–2 sentences to describe the spread of the data. Do most of the worms have a length that is close to your estimate of a typical length, or are they very different in length?

Are you ready for more?

Here is another histogram for the earthworm measurement data. In this histogram, the measurements are in different groupings.



1. Based on this histogram, what is your estimate of a typical length for the 25 earthworms?

2. Compare this histogram with the one you drew. How are the distributions of data summarized in the two histograms the same? How are they different?

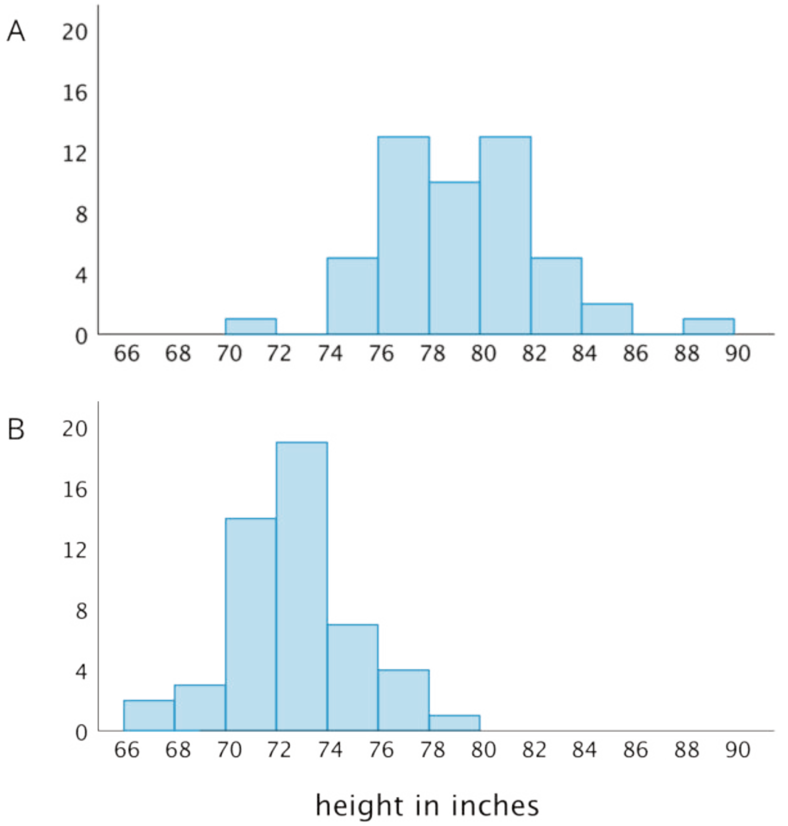
3. Compare your estimates of a typical earthworm length for the two histograms. Did you reach different conclusions about a typical earthworm length from the two histograms?

7.3: Tall and Taller Players

Professional basketball players tend to be taller than professional baseball players.

Here are two histograms that show height distributions of 50 male professional baseball players and 50 male professional basketball players.

1. Decide which histogram shows the heights of baseball players and which shows the heights of basketball players. Be prepared to explain your reasoning.



2. Write 2–3 sentences that describe the distribution of the heights of the basketball players. Comment on the center and spread of the data.

3. Write 2–3 sentences that describe the distribution of the heights of the baseball players. Comment on the center and spread of the data.

Lesson 7 Summary

Here are the weights, in kilograms, of 30 dogs.

10	11	12	12	13	15	16	16
17	18	18	19	20	20	20	21
22	22	22	23	24	24	26	26
28	30	32	32	34	34		

Before we draw a histogram, let’s consider a couple of questions.

- What are the smallest and largest values in our data set? This gives us an idea of the distance on the number line that our histogram will cover. In this case, the minimum is 10 and the maximum is 34, so our number line needs to extend from 10 to 35 at the very least.

  (Remember the convention we use to mark off the number line for a histogram: we include the left boundary of a bar but exclude the right boundary. If 34 is the right boundary of the last bar, it won't be included in that bar, so the number line needs to go a little greater than the maximum value.)

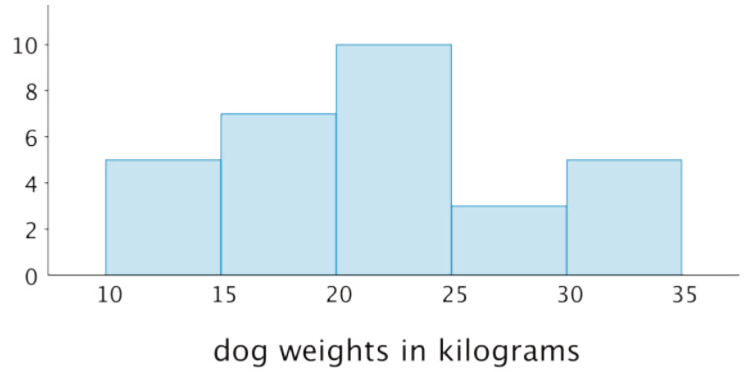
- What group size or bin size seems reasonable here? We could organize the weights into bins of 2 kilograms (10, 12, 14, . . .), 5 kilograms, (10, 15, 20, 25, . . .), 10 kilograms (10, 20, 30, . . .), or any other size. The smaller the bins, the more bars we will have, and vice versa.

Let’s use bins of 5 kilograms for the dog weights. The boundaries of our bins will be: 10, 15, 20, 25, 30, 35. We stop at 35 because it is greater than the maximum.

Next, we find the frequency for the values in each group. It is helpful to organize the values in a table.

weights in kilograms	frequency
10 to less than 15	5
15 to less than 20	7
20 to less than 25	10
25 to less than 30	3
30 to less than 35	5

Now we can draw the histogram.

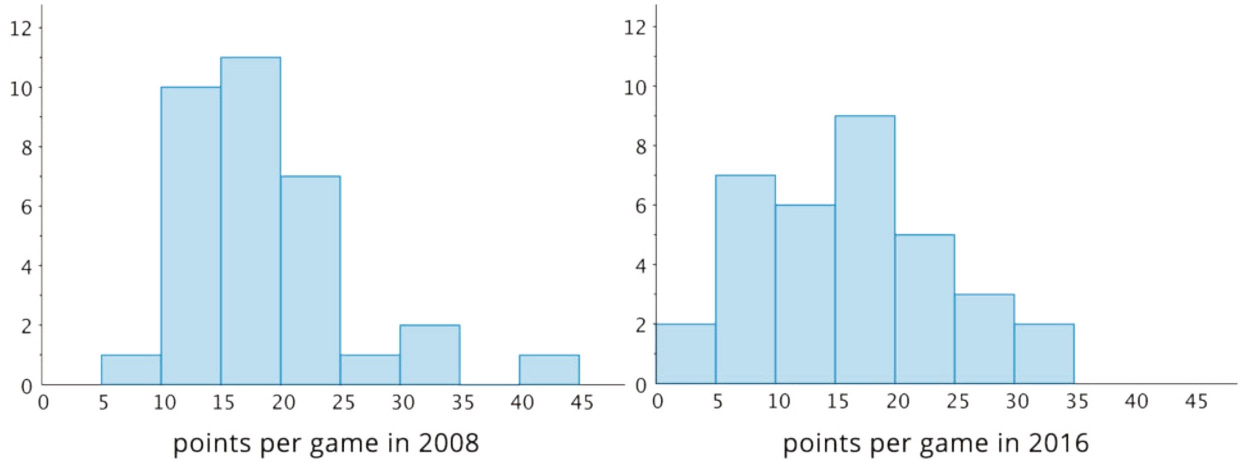


The histogram allows us to learn more about the dog weight distribution and describe its center and spread.

Lesson 7: Using Histograms to Answer Statistical Questions

Cool Down: A Tale of Two Seasons

The two histograms show the points scored per game by a college basketball player in 2008 and 2016.

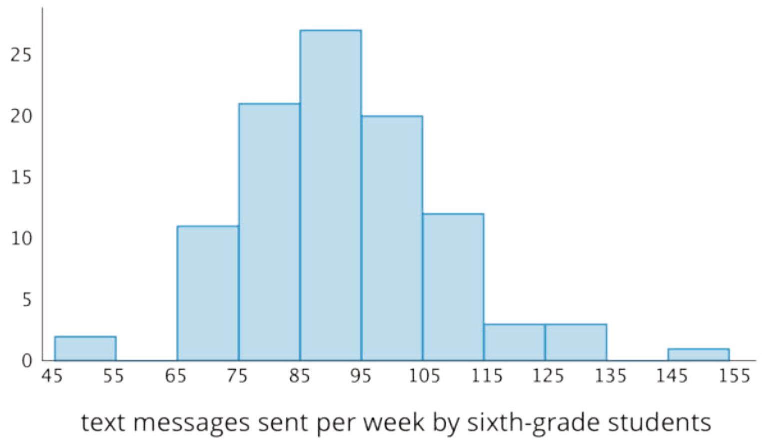


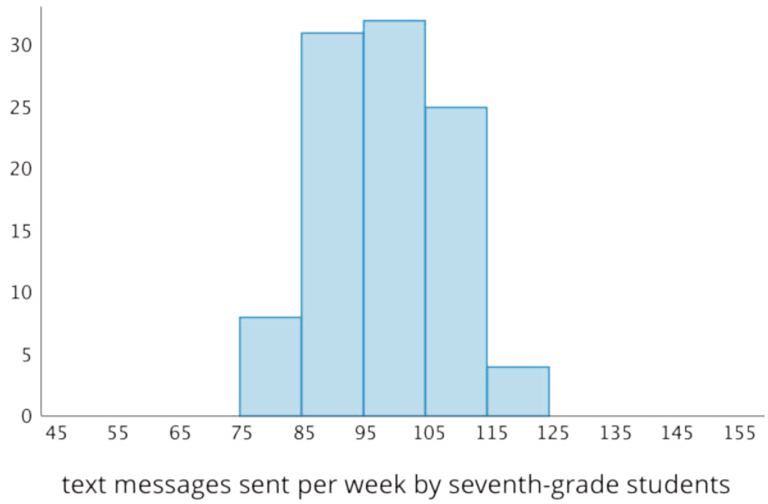
1. What is a typical number of points per game scored by this player in 2008? What about in 2016? Explain your reasoning.

2. Write 2–3 sentences that describe the spreads of the two distributions, including what spreads might tell us in this context.

Unit 8 Lesson 7 Cumulative Practice Problems

1. These two histograms show the number of text messages sent in one week by two groups of 100 students. The first histogram summarizes data from sixth-grade students. The second histogram summarizes data from seventh-grade students.





   a. Do the two data sets have approximately the same center? If so, explain where the center is located. If not, which one has the greater center?

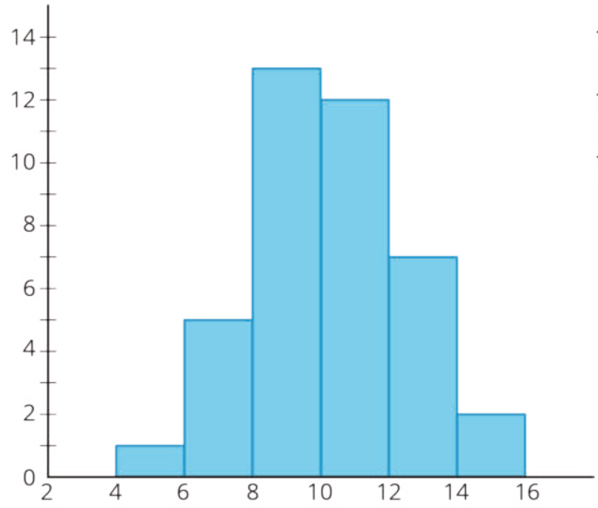
   b. Which data set has greater spread? Explain your reasoning.

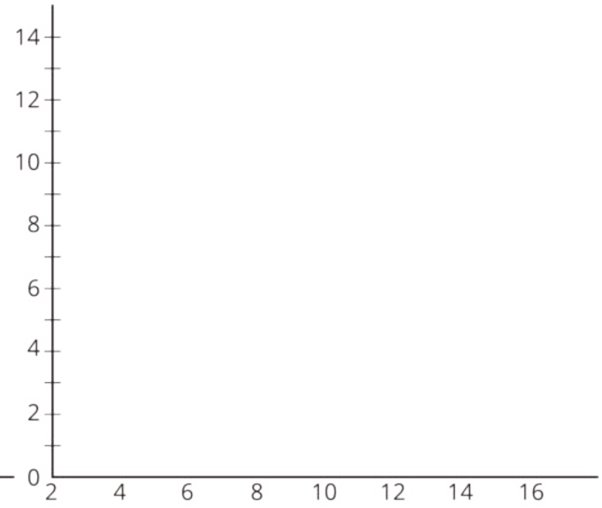
   c. Overall, which group of students—sixth- or seventh-grade—sent more text messages?

2. Forty sixth-grade students ran 1 mile. Here is a histogram that summarizes their times, in minutes. The center of the distribution is approximately 10 minutes.

   On the blank axes, draw a second histogram that has:

   - a distribution of times for a different group of 40 sixth-grade students.
   - a center at 10 minutes.
   - less variability than the distribution shown in the first histogram.





3. Jada has d dimes. She has more than 30 cents but less than a dollar.

   a. Write two inequalities that represent how many dimes Jada has.

   b. Can d be 10?

   c. How many possible solutions make both inequalities true? If possible, describe or list the solutions.

(From Unit 7, Lesson 9.)

4. Order these numbers from greatest to least: -4, $\frac{1}{4}$, 0, 4, $-3\frac{1}{2}$, $\frac{7}{4}$, $-\frac{5}{4}$

(From Unit 7, Lesson 4.)

Lesson 9: Interpreting the Mean as Fair Share

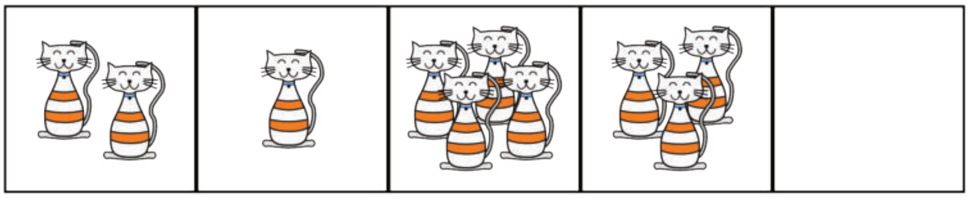
9.1: Close to Four

Use the digits 0–9 to write an expression with a value as close as possible to 4. Each digit can be used only one time in the expression.

$$(\boxed{} + \boxed{} + \boxed{} + \boxed{}) \div 4$$

9.2: Spread Out and Share

1. The kittens in a room at an animal shelter are placed in 5 crates.



   a. The manager of the shelter wants the kittens distributed equally among the crates. How might that be done? How many kittens will end up in each crate?

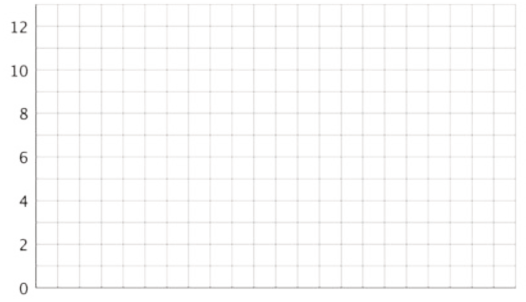
   b. The number of kittens in each crate after they are equally distributed is called the **mean** number of kittens per crate, or the **average** number of kittens per crate. Explain how the expression $10 \div 5$ is related to the average.

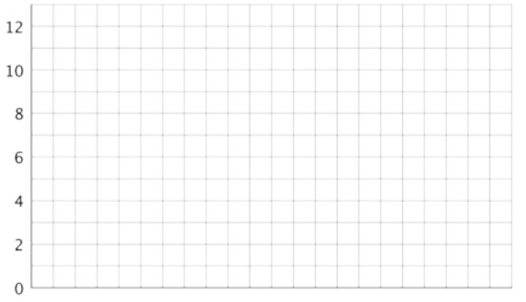
c. Another room in the shelter has 6 crates. No two crates has the same number of kittens, and there is an average of 3 kittens per crate. Draw or describe at least two different arrangements of kittens that match this description.

2. Five servers were scheduled to work the number of hours shown. They decided to share the workload, so each one would work equal hours.

server A: 3 server B: 6 server C: 11 server D: 7 server E: 4

a. On the grid on the left, draw 5 bars whose heights represent the hours worked by servers A, B, C, D, and E.





b. Think about how you would rearrange the hours so that each server gets a fair share. Then, on the grid on the right, draw a new graph to represent the rearranged hours. Be prepared to explain your reasoning.

c. Based on your second drawing, what is the average or mean number of hours that the servers will work?

d. Explain why we can also find the mean by finding the value of the expression $31 \div 5$.

e. Which server will see the biggest change to work hours? Which server will see the least change?

Are you ready for more?

Server F, working 7 hours, offers to join the group of five servers, sharing their workload. If server F joins, will the mean number of hours worked increase or decrease? Explain how you know.

9.3: Getting to School

For the past 12 school days, Mai has recorded how long her bus rides to school take in minutes. The times she recorded are shown in the table.

9	8	6	9	10	7	6	12	9	8	10	8

1. Find the mean for Mai's data. Show your reasoning.

2. In this situation, what does the mean tell us about Mai's trip to school?

3. For 5 days, Tyler has recorded how long his walks to school take in minutes. The mean for his data is 11 minutes. Without calculating, predict if each of the data sets shown could be Tyler's. Explain your reasoning.
    - data set A: 11, 8, 7, 9, 8
    - data set B: 12, 7, 13, 9, 14
    - data set C: 11, 20, 6, 9, 10
    - data set D: 8, 10, 9, 11, 11

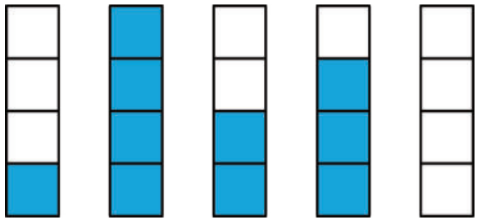
4. Determine which data set is Tyler's. Explain how you know.

Lesson 9 Summary

Sometimes a general description of a distribution does not give enough information, and a more precise way to talk about center or spread would be more useful. The **mean**, or **average**, is a number we can use to summarize a distribution.

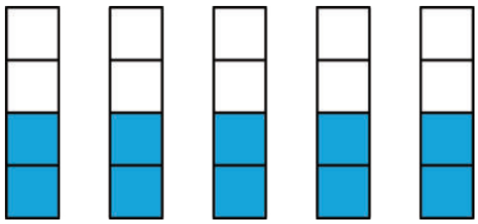
We can think about the mean in terms of “fair share” or “leveling out.” That is, a mean can be thought of as a number that each member of a group would have if all the data values were combined and distributed equally among the members.

For example, suppose there are 5 bottles which have the following amounts of water: 1 liter, 4 liters, 2 liters, 3 liters, and 0 liters.



To find the mean, first we add up all of the values. We can think of as putting all of the water together: $1+4+2+3+0=10$.

To find the “fair share,” we divide the 10 liters equally into the 5 containers: $10 \div 5 = 2$.



Suppose the quiz scores of a student are 70, 90, 86, and 94. We can find the mean (or average) score by finding the sum of the scores $(70+90+86+94=340)$ and dividing the sum by four $(340 \div 4 = 85)$. We can then say that the student scored, on average, 85 points on the quizzes.

In general, to find the mean of a data set with n values, we add all of the values and divide the sum by n.

Lesson 9: Interpreting the Mean as Fair Share

Cool Down: Finding Means

1. Last week, the daily low temperatures for a city, in degrees Celsius, were 5, 8, 6, 5, 10, 7, and 1. What was the average low temperature? Show your reasoning.

2. The mean of four numbers is 7. Three of the numbers are 5, 7, and 7. What is the fourth number? Explain your reasoning.

Unit 8 Lesson 9 Cumulative Practice Problems

1. A preschool teacher is rearranging four boxes of playing blocks so that each box contains an equal number of blocks. Currently Box 1 has 32 blocks, Box 2 has 18, Box 3 has 41, and Box 4 has 9.

   Select **all** the ways he could make each box have the same number of blocks.

   A. Remove all the blocks and make four equal piles of 25, then put each pile in one of the boxes.

   B. Remove 7 blocks from Box 1 and place them in Box 2.

   C. Remove 21 blocks from Box 3 and place them in Box 4.

   D. Remove 7 blocks from Box 1 and place them in Box 2, and remove 21 blocks from Box 3 and place them in Box 4.

   E. Remove 7 blocks from Box 1 and place them in Box 2, and remove 16 blocks from Box 3 and place them in Box 4.

2. In a round of mini-golf, Clare records the number of strokes it takes to hit the ball into the hole of each green.

   2 3 1 4 5 2 3 4 3

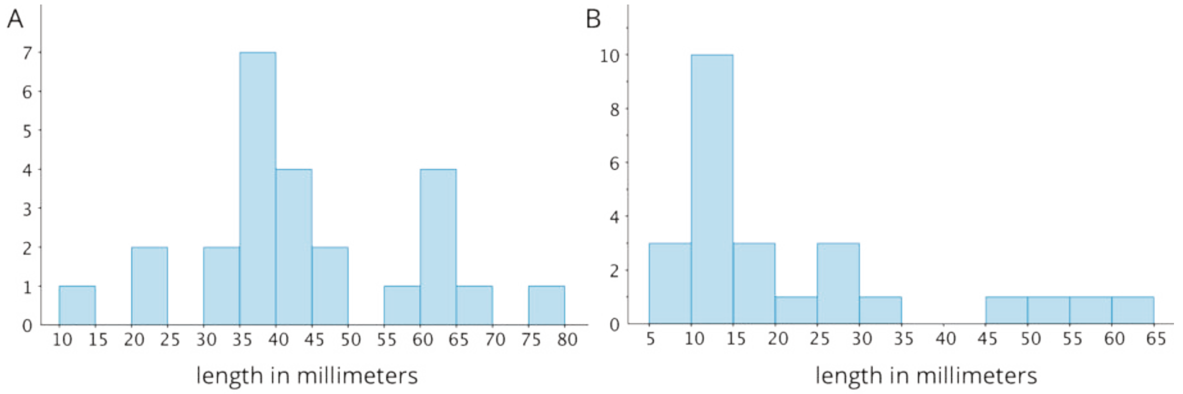
   She said that, if she redistributed the strokes on different greens, she could tell that her average number of strokes per hole is 3. Explain how Clare is correct.

3. Three sixth-grade classes raised \$25.50, \$49.75, and \$37.25 for their classroom libraries. They agreed to share the money raised equally. What is each class's equal share? Explain or show your reasoning.

4. In her English class, Mai's teacher gives 4 quizzes each worth 5 points. After 3 quizzes, she has the scores 4, 3, and 4. What does she need to get on the last quiz to have a mean score of 4? Explain or show your reasoning.

5. An earthworm farmer examined two containers of a certain species of earthworms so that he could learn about their lengths. He measured 25 earthworms in each container and recorded their lengths in millimeters.

   Here are histograms of the lengths for each container.



   a. Which container tends to have longer worms than the other container?

   b. For which container would 15 millimeters be a reasonable description of a typical length of the worms in the container?

   c. If length is related to age, which container had the most young worms?

   (From Unit 8, Lesson 7.)

6. Diego thinks that $x = 3$ is a solution to the equation $x^2 = 16$. Do you agree? Explain or show your reasoning.

   (From Unit 6, Lesson 15.)

Lesson 10: Finding and Interpreting the Mean as the Balance Point

10.1: Which One Doesn’t Belong: Division

Which expression does not belong? Be prepared to explain your reasoning.

$$\frac{8+8+4+4}{4} \qquad \frac{10+10+4}{4} \qquad \frac{9+9+5+5}{4} \qquad \frac{6+6+6+6+6}{5}$$

10.2: Travel Times (Part 1)

Here is the data set from an earlier lesson showing how long it takes for Diego to walk to school, in minutes, over 5 days. The mean number of minutes is 11.

12 7 13 9 14

1. Represent Diego’s data on a dot plot. Mark the location of the mean with a triangle.

2. The mean can also be seen as a **measure of center** that balances the points in a data set. If we find the distance between every point and the mean, add the distances on each side of the mean, and compare the two sums, we can see this balancing.

a. Record the distance between each point and 11 and its location relative to 11.

time in minutes	distance from 11	left of 11 or right of 11?
12		
7		
13		
9		
14		

b. Sum of distances left of 11:__________ Sum of distances right of 11:__________

What do you notice about the two sums?

3. Can another point that is *not* the mean produce similar sums of distances? Let's investigate whether 10 can produce similar sums as those of 11.

a. Complete the table with the distance of each data point from 10.

time in minutes	distance from 10	left of 10 or right of 10?
12		
7		
13		
9		
14		

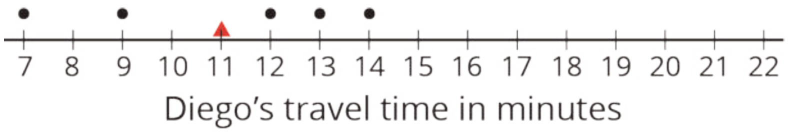
b. Sum of distances left of 10:__________ Sum of distances right of 10:__________

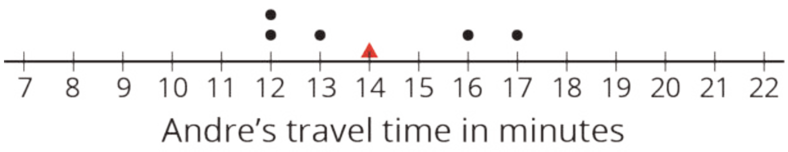
What do you notice about the two sums?

4. Based on your work so far, explain why the mean can be considered a balance point for the data set.

10.3: Travel Times (Part 2)

1. Here are dot plots showing how long Diego's trips to school took in minutes—which you studied earlier—and how long Andre's trips to school took in minutes. The dot plots include the means for each data set, marked by triangles.

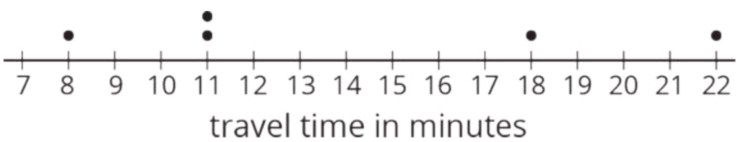




   a. Which of the two data sets has a larger mean? In this context, what does a larger mean tell us?

   b. Which of the two data sets has larger sums of distances to the left and right of the mean? What do these sums tell us about the variation in Diego's and Andre's travel times?

2. Here is a dot plot showing lengths of Lin's trips to school.



   a. Calculate the mean of Lin's travel times.

b. Complete the table with the distance between each point and the mean as well whether the point is to the left or right of the mean.

time in minutes	distance from the mean	left or right of the mean?
22		
18		
11		
8		
11		

c. Find the sum of distances to the left of the mean and the sum of distances to the right of the mean.

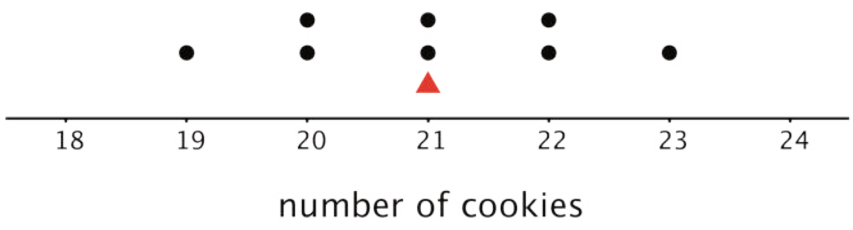
d. Use your work to compare Lin's travel times to Andre's. What can you say about their average travel times? What about the variability in their travel times?

Lesson 10 Summary

The mean is often used as a **measure of center** of a distribution. This is because the mean of a distribution can be seen as the "balance point" for the distribution. Why is this a good way to think about the mean? Let's look at a very simple set of data on the number of cookies that each of eight friends baked:

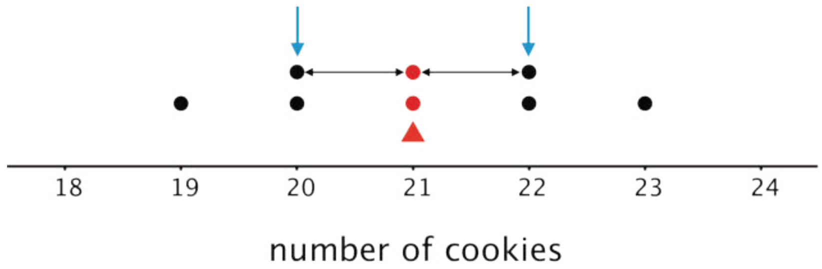
19 20 20 21 21 22 22 23

Here is a dot plot showing the data set.

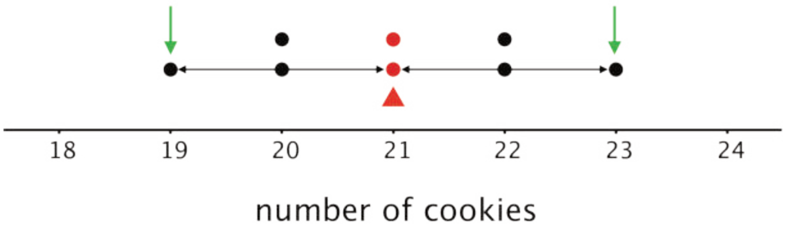


The distribution shown is completely symmetrical. The mean number of cookies is 21, because $(19+20+20+21+21+22+22+23)\div 8=21$. If we mark the location of the mean on the dot plot, we can see that the data points could balance at 21.

In this plot, each point on either side of the mean has a mirror image. For example, the two points at 20 and the two at 22 are the same distance from 21, but each pair is located on either side of 21. We can think of them as balancing each other around 21.



Similarly, the points at 19 and 23 are the same distance from 21 but are on either side of it. They, too, can be seen as balancing each other around 21.



We can say that the distribution of the cookies has a center at 21 because that is its balance point, and that the eight friends, on average, baked 21 cookies.

Even when a distribution is not completely symmetrical, the distances of values below the mean, on the whole, balance the distances of values above the mean.

Lesson 10: Finding and Interpreting the Mean as the Balance Point

Cool Down: Text Messages

The three data sets show the number of text messages sent by Jada, Diego, and Lin over 6 days. One of the data sets has a mean of 4, one has a mean of 5, and one has a mean of 6.

Jada	4	4	4	6	6	6
Diego	4	5	5	6	8	8
Lin	1	1	2	2	9	9

1. Which data set has which mean? What does this tell you about the text messages sent by the three students?

2. Which data set has the greatest variability? Explain your reasoning.

Unit 8 Lesson 10 Cumulative Practice Problems

1. On school days, Kiran walks to school. Here are the lengths of time, in minutes, for Kiran's walks on 5 school days:

   16 11 18 12 13

   a. Create a dot plot for Kiran's data.

   b. Without calculating, decide if 15 minutes would be a good estimate of the mean. If you think it is a good estimate, explain your reasoning. If not, give a better estimate and explain your reasoning.

   c. Calculate the mean for Kiran's data.

   d. In the table, record the distance of each data point from the mean and its location relative to the mean.

time in minutes	distance from the mean	left or right of the mean?
16		
11		
18		
12		
13		

   e. Calculate the sum of all distances to the left of the mean, then calculate the sum of distances to the right of the mean. Explain how these sums show that the mean is a balance point for the values in the data set.

2. Noah scored 20 points in a game. Mai's score was 30 points. The mean score for Noah, Mai, and Clare was 40 points. What was Clare's score? Explain or show your reasoning.

3. Compare the numbers using >, <, or =.

a. -2 _____ 3	a. 3 _____ -4	a. 7 _____ -11
b. $\lvert -12 \rvert$ _____ $\lvert 15 \rvert$	b. $\lvert 15 \rvert$ _____ $\lvert -12 \rvert$	b. -4 _____ $\lvert 5 \rvert$

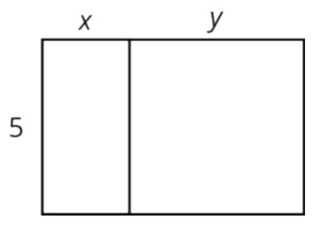
(From Unit 7, Lesson 7.)

4. a. Plot $\frac{2}{3}$ and $\frac{3}{4}$ on a number line.

   b. Is $\frac{2}{3} < \frac{3}{4}$, or is $\frac{3}{4} < \frac{2}{3}$? Explain how you know.

(From Unit 7, Lesson 3.)

5. Select **all** the expressions that represent the total area of the large rectangle.



A. $5(x + y)$

B. $5 + xy$

C. $5x + 5y$

D. $2(5 + x + y)$

E. $5xy$

(From Unit 6, Lesson 10.)

Lesson 11: Deviation from the Mean

11.1: Shooting Hoops (Part 1)

Elena, Jada, and Lin enjoy playing basketball during recess. Lately, they have been practicing free throws. They record the number of baskets they make out of 10 attempts. Here are their data sets for 12 school days.

Elena

4	5	1	6	9	7	2	8	3	3	5	7

Jada

2	4	5	4	6	6	4	7	3	4	8	7

Lin

3	6	6	4	5	5	3	5	4	6	6	7

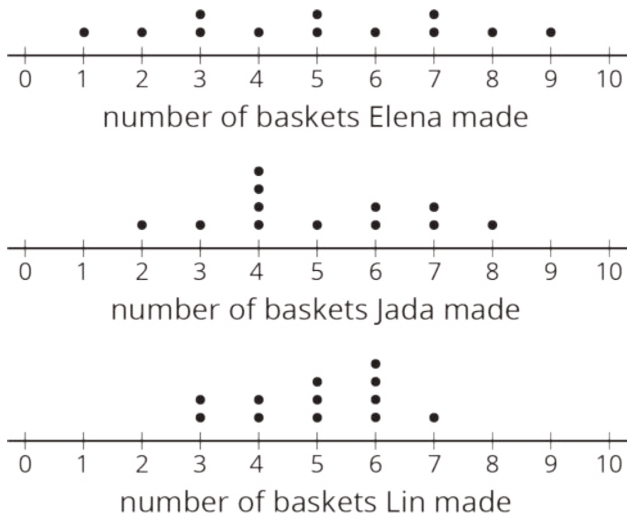
1. Calculate the mean number of baskets each player made, and compare the means. What do you notice?

2. What do the means tell us in this context?

11.2: Shooting Hoops (Part 2)

Here are the dot plots showing the number of baskets Elena, Jada, and Lin each made over 12 school days.

1. On each dot plot, mark the location of the mean with a triangle (Δ). Then, contrast the dot plot distributions. Write 2–3 sentences to describe the shape and spread of each distribution.



2. Discuss the following questions with your group. Explain your reasoning.

   a. Would you say that all three students play equally well?

   b. Would you say that all three students play equally consistently?

   c. If you could choose one player to be on your basketball team based on their records, who would you choose?

11.3: Shooting Hoops (Part 3)

The tables show Elena, Jada, and Lin's basketball data from an earlier activity. Recall that the mean of Elena's data, as well as that of Jada and Lin's data, was 5.

1. Record the distance between each of Elena's scores and the mean.

Elena	4	5	1	6	9	7	2	8	3	3	5	7
distance from 5	1			1								

   Now find *the average of the distances* in the table. Show your reasoning and round your answer to the nearest tenth.

   This value is the **mean absolute deviation (MAD)** of Elena's data.

   Elena's MAD: _________

2. Find the mean absolute deviation of Jada's data. Round it to the nearest tenth.

Jada	2	4	5	4	6	6	4	7	3	4	8	7
distance from 5												

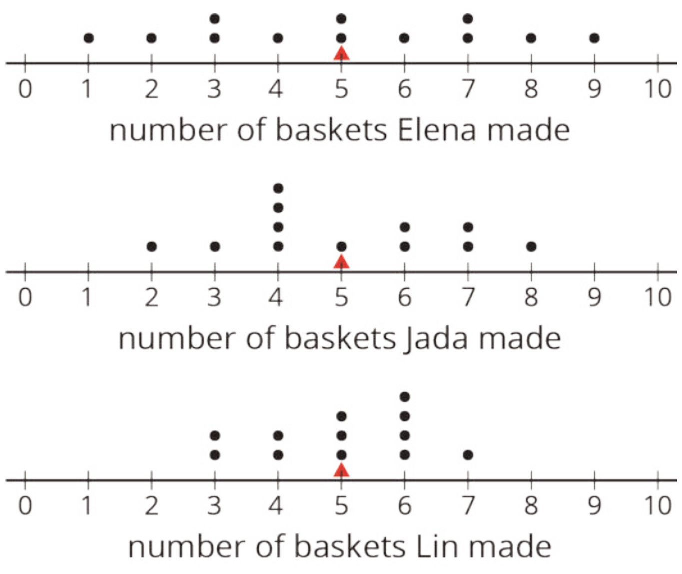
   Jada's MAD: _________

3. Find the mean absolute deviation of Lin's data. Round it to the nearest tenth.

Lin	3	6	6	4	5	5	3	5	4	6	6	7
distance from 5												

Lin's MAD: _________

4. Compare the MADs and dot plots of the three students' data. Do you see a relationship between each student's MAD and the distribution on her dot plot? Explain your reasoning.



Are you ready for more?

Invent another data set that also has a mean of 5 but has a MAD greater than 2. Remember, the values in the data set must be whole numbers from 0 to 10.

11.4: Game of 22

Your teacher will give your group a deck of cards. Shuffle the cards, and put the deck face down on the playing surface.

- To play: Draw 3 cards and add up the values. An ace is a 1. A jack, queen, and king are each worth 10. Cards 2–10 are each worth their face value. If your sum is anything other than 22 (either above or below 22), say: “My sum deviated from 22 by ____ ,” or “My sum was off from 22 by ____ .”
- To keep score: Record each sum and each distance from 22 in the table. After five rounds, calculate the average of the distances. The player with the lowest average distance from 22 wins the game.

player A	round 1	round 2	round 3	round 4	round 5
sum of cards					
distance from 22					

Average distance from 22: ____________

player B	round 1	round 2	round 3	round 4	round 5
sum of cards					
distance from 22					

Average distance from 22: ____________

player C	round 1	round 2	round 3	round 4	round 5
sum of cards					
distance from 22					

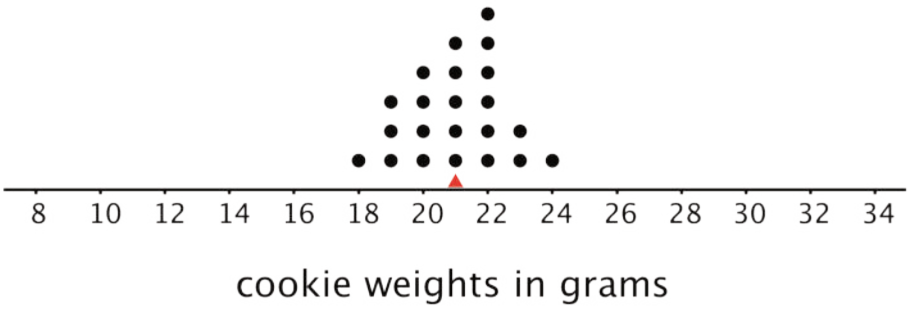
Average distance from 22: ____________

Whose average distance from 22 is the smallest? Who won the game?

Lesson 11 Summary

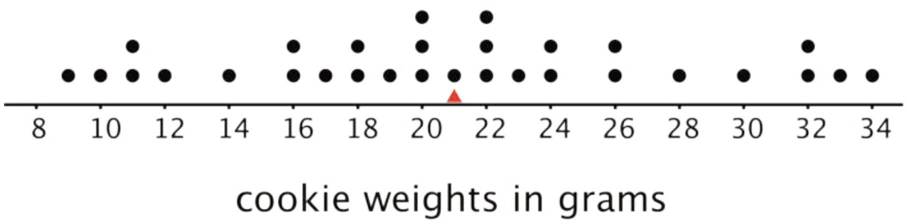
We use the mean of a data set as a measure of center of its distribution, but two data sets with the same mean could have very different distributions.

This dot plot shows the weights, in grams, of 22 cookies.



The mean weight is 21 grams. All the weights are within 3 grams of the mean, and most of them are even closer. These cookies are all fairly close in weight.

This dot plot shows the weights, in grams, of a different set of 30 cookies.

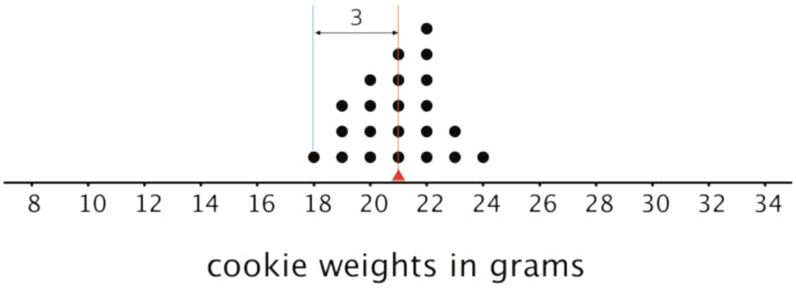


The mean weight for this set of cookies is also 21 grams, but some cookies are half that weight and others are one-and-a-half times that weight. There is a lot more variability in the weight.

There is a number we can use to describe how far away, or how spread out, data points generally are from the mean. This *measure of spread* is called the **mean absolute deviation (MAD)**.

Here the MAD tells us how far cookie weights typically are from 21 grams. To find the MAD, we find the distance between each data value and the mean, and then calculate the mean of those distances.

For instance, the point that represents 18 grams is 3 units away from the mean of 21 grams. We can find the distance between each point and the mean of 21 grams and organize the distances into a table, as shown.



weight in grams	18	19	19	19	20	20	20	20	21	21	21	21	21	22	22	22	22	22	22	23	23	24
distance from mean	3	2	2	2	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	2	2	3

The values in the first row of the table are the cookie weights for the first set of cookies. Their mean, 21 grams, is the *mean of the cookie weights*.

The values in the second row of the table are the *distances* between the values in the first row and 21. The mean of these distances is the *MAD of the cookie weights*.

What can we learn from the averages of these distances once they are calculated?

- In the first set of cookies, the distances are all between 0 and 3. The MAD is 1.2 grams, which tells us that the cookie weights are typically within 1.2 grams of 21 grams. We could say that a typical cookie weighs between 19.8 and 22.2 grams.

- In the second set of cookies, the distances are all between 0 and 13. The MAD is 5.6 grams, which tells us that the cookie weights are typically within 5.6 grams of 21 grams. We could say a typical cookie weighs between 15.4 and 26.6 grams.

The MAD is also called a *measure of the variability* of the distribution. In these examples, it is easy to see that a higher MAD suggests a distribution that is more spread out, showing more variability.

Lesson 11: Deviation from the Mean

Cool Down: Text Messages, Again

These three data sets show the number of text messages sent by Jada, Diego, and Lin over 6 days as well as the mean number of text messages sent by each student per day.

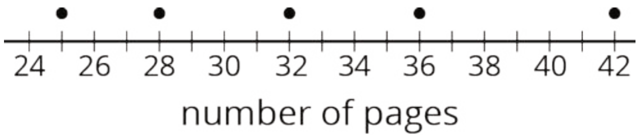
Jada mean: 5	4	4	4	6	6	6
Diego mean: 6	4	5	5	6	8	8
Lin mean: 4	1	1	2	2	9	9

1. Predict which data set has the largest MAD and which has the smallest MAD.

2. Compute the MAD for each data set to check your prediction.

Unit 8 Lesson 11 Cumulative Practice Problems

1. Han recorded the number of pages that he read each day for five days. The dot plot shows his data.



a. Is 30 pages a good estimate of the mean number of pages that Han read each day? Explain your reasoning.

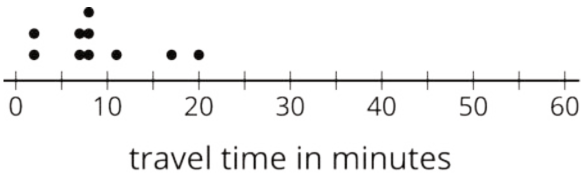
b. Find the mean number of pages that Han read during the five days. Draw a triangle to mark the mean on the dot plot.

c. Use the dot plot and the mean to complete the table.

number of pages	distance from mean	left or right of mean
25		left
28		
32		
36		
42		

d. Calculate the mean absolute deviation (MAD) of the data. Explain or show your reasoning.

2. Ten sixth-grade students recorded the amounts of time each took to travel to school. The dot plot shows their travel times.



The mean travel time for these students is approximately 9 minutes. The MAD is approximately 4.2 minutes.

   a. Which number of minutes—9 or 4.2—is a typical amount of time for the ten sixth-grade students to travel to school? Explain your reasoning.

   b. Based on the mean and MAD, Jada believes that travel times between 5 and 13 minutes are common for this group. Do you agree? Explain your reasoning.

   c. A different group of ten sixth-grade students also recorded their travel times to school. Their mean travel time was also 9 minutes, but the MAD was about 7 minutes. What could the dot plot of this second data set be? Describe or draw how it might look.

3. In an archery competition, scores for each round are calculated by averaging the distance of 3 arrows from the center of the target.

   An archer has a mean distance of 1.6 inches and a MAD distance of 1.3 inches in the first round. In the second round, the archer's arrows are farther from the center but are more consistent. What values for the mean and MAD would fit this description for the second round? Explain your reasoning.

Lesson 12: Using Mean and MAD to Make Comparisons

12.1: Number Talk: Decimal Division

Find the value of each expression mentally.

$42 \div 12$

$2.4 \div 12$

$44.4 \div 12$

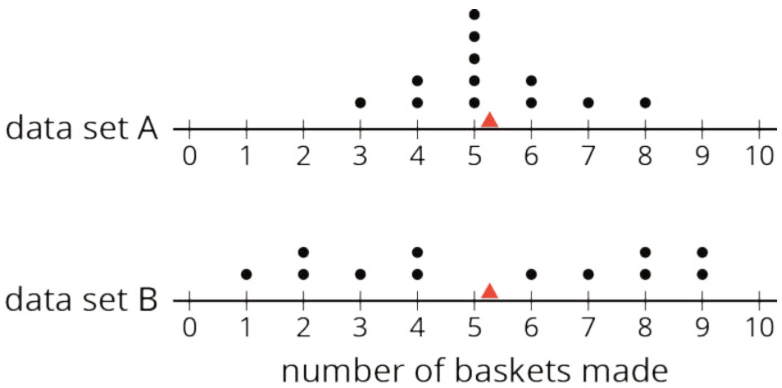
$46.8 \div 12$

12.2: Which Player Would You Choose?

1. Andre and Noah joined Elena, Jada, and Lin in recording their basketball scores. They all recorded their scores in the same way: the number of baskets made out of 10 attempts. Each collected 12 data points.

    - Andre's mean number of baskets was 5.25, and his MAD was 2.6.
    - Noah's mean number of baskets was also 5.25, but his MAD was 1.

    Here are two dot plots that represent the two data sets. The triangle indicates the location of the mean.



    a. Without calculating, decide which dot plot represents Andre's data and which represents Noah's. Explain how you know.

b. If you were the captain of a basketball team and could use one more player on your team, would you choose Andre or Noah? Explain your reasoning.

2. An eighth-grade student decided to join Andre and Noah and kept track of his scores. His data set is shown here. The mean number of baskets he made is 6.

eighth-grade student	6	5	4	7	6	5	7	8	5	6	5	8
distance from 6												

a. Calculate the MAD. Show your reasoning.

b. Draw a dot plot to represent his data and mark the location of the mean with a triangle (Δ).

c. Compare the eighth-grade student’s mean and MAD to Noah’s mean and MAD. What do you notice?

d. Compare their dot plots. What do you notice about the distributions?

e. What can you say about the two players’ shooting accuracy and consistency?

Are you ready for more?

Invent a data set with a mean of 7 and a MAD of 1.

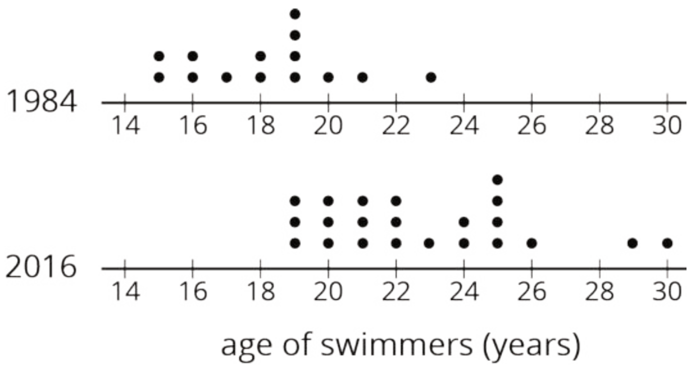
12.3: Swimmers Over the Years

In 1984, the mean age of swimmers on the U.S. women's swimming team was 18.2 years and the MAD was 2.2 years. In 2016, the mean age of the swimmers was 22.8 years, and the MAD was 3 years.

1. How has the average age of the women on the U.S. swimming team changed from 1984 to 2016? Explain your reasoning.

2. Are the swimmers on the 1984 team closer in age to one another than the swimmers on the 2016 team are to one another? Explain your reasoning.

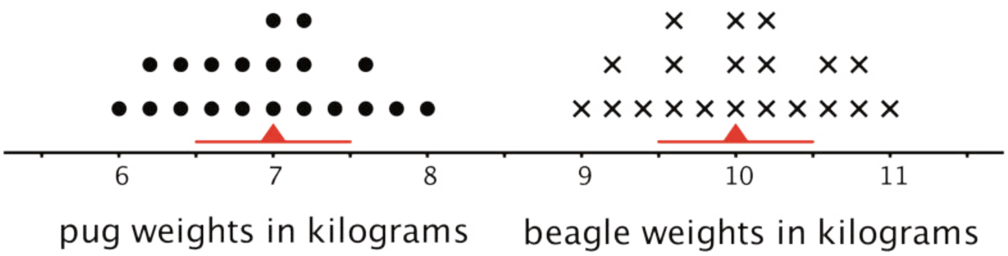
3. Here are dot plots showing the ages of the women on the U.S. swimming team in 1984 and in 2016. Use them to make two other comments about how the women's swimming team has changed over the years.



Lesson 12 Summary

Sometimes two distributions have different means but the same MAD.

Pugs and beagles are two different dog breeds. The dot plot shows two sets of weight data—one for pugs and the other for beagles.



- The mean weight for pugs is 7 kilograms, and the MAD is 0.5 kilogram.
- The mean weight for beagles is 10 kilograms, and the MAD is 0.5 kilogram.

We can say that, in general, the beagles are heavier than the pugs. A typical weight for the beagles in this group is about 3 kilograms heavier than a typical weight for the pugs.

The variability of pug weights, however, is about the same as the variability of beagle weights. In other words, the weights of pugs and the weights of beagles are equally spread out.

Lesson 12: Using Mean and MAD to Make Comparisons

Cool Down: Travel Times Across the World

Ten sixth-grade students in five different countries were asked about their travel times to school. Their responses were organized into five data sets. The mean and MAD of each data set is shown in the table.

	mean (minutes)	MAD (minutes)
United States	9	4.2
Australia	18.1	7.9
South Africa	23.5	16.2
Canada	11	8
New Zealand	12.3	5.5

1. Which group of students has the greatest variability in their travel times? Explain your reasoning.

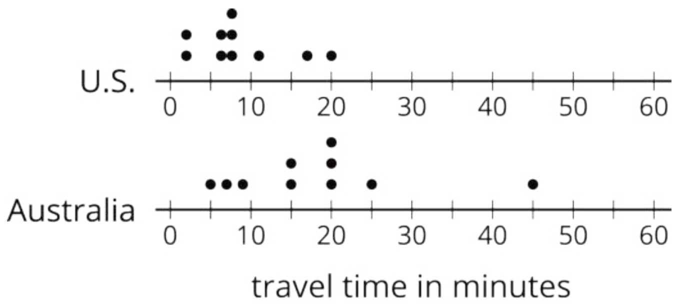
2. a. The mean of the data set for New Zealand is close to that of Canada. What does this tell us about the travel times of students in those two data sets?

   b. The MAD of the data set for New Zealand is quite different than that of Canada. What does this tell us about the travel times of students in those two data sets?

3. The data sets for Australia and Canada have very different means (18.1 and 11 minutes) but very similar MADs. What can you say about the travel times of the students in those two data sets?

Unit 8 Lesson 12 Cumulative Practice Problems

1. The dot plots show the amounts of time that ten U.S. students and ten Australian students took to get to school.



Which statement is true about the MAD of the Australian data set?

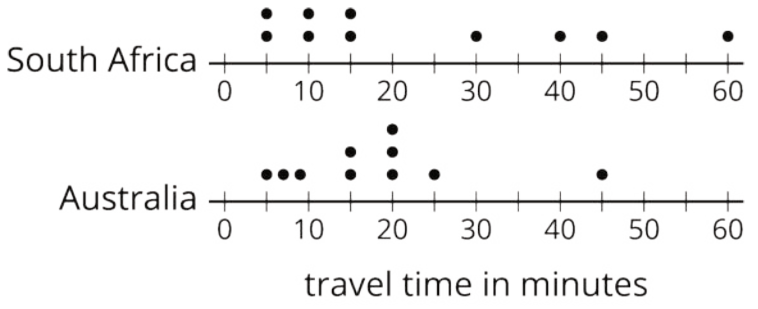
A. It is significantly less than the MAD of the U.S. data set.

B. It is exactly equal to the MAD of the U.S. data set.

C. It is approximately equal to the MAD of the U.S. data set.

D. It is significantly greater than the MAD of the U.S. data set.

2. The dot plots show the amounts of time that ten South African students and ten Australian students took to get to school. Without calculating, answer the questions.



   a. Which data set has the smaller mean? Explain your reasoning.

   b. Which data set has the smaller MAD? Explain your reasoning.

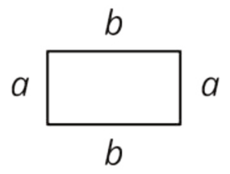
   c. What does a smaller mean tell us in this context?

   d. What does a smaller MAD tell us in this context?

3. Two high school basketball teams have identical records of 15 wins and 2 losses. Sunnyside High School's mean score is 50 points and its MAD is 4 points. Shadyside High School's mean score is 60 points and its MAD is 15 points.

   Lin read the records of each team's score. She likes the team that had nearly the same score for every game it played. Which team do you think Lin likes? Explain your reasoning.

4. Jada thinks the perimeter of this rectangle can be represented with the expression $a + a + b + b$. Andre thinks it can be represented with $2a + 2b$. Do you agree with either of them? Explain your reasoning.



(From Unit 6, Lesson 8.)

5. Draw a number line.

   a. Plot and label three numbers between -2 and -8 (not including -2 and -8).

   b. Use the numbers you plotted and the symbols $<$ and $>$ to write three inequality statements.

(From Unit 7, Lesson 3.)

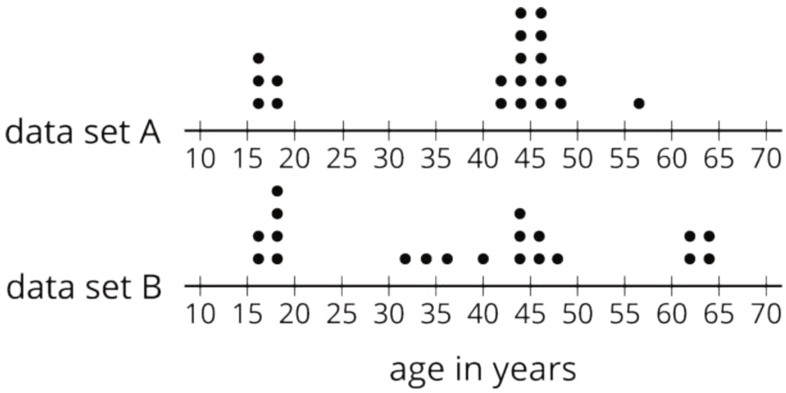
6. Adult elephant seals generally weigh about 5,500 pounds. If you weighed 5 elephant seals, would you expect each seal to weigh exactly 5,500 pounds? Explain your reasoning.

(From Unit 8, Lesson 2.)

Lesson 13: The Median of a Data Set

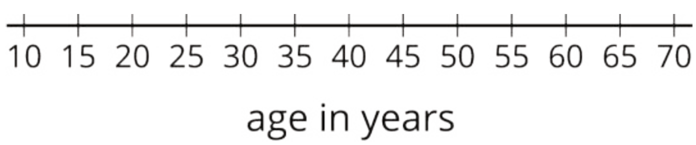
13.1: The Plot of the Story

1. Here are two dot plots and two stories. Match each story with a dot plot that could represent it. Be prepared to explain your reasoning.



   - Twenty people—high school students, teachers, and invited guests—attended a rehearsal for a high school musical. The mean age was 38.5 years and the MAD was 16.5 years.
   - High school soccer team practice is usually watched by supporters of the players. One evening, twenty people watched the team practice. The mean age was 38.5 years and the MAD was 12.7 years.

2. Another evening, twenty people watched the soccer team practice. The mean age was similar to that from the first evening, but the MAD was greater (about 20 years). Make a dot plot that could illustrate the distribution of ages in this story.



13.2: Siblings in the House

Here is data that shows the numbers of siblings of ten students in Tyler's class.

1	0	2	1	7	0	2	0	1	10

1. Represent the data shown with a dot plot.

2. Without making any calculations, estimate the center of the data based on your dot plot. What is a typical number of siblings for these sixth-grade students? Mark the location of that number on your dot plot.

3. Find the mean. Show your reasoning.

4. a. How does the mean compare to the value that you marked on the dot plot as a typical number of siblings? (Is it a little larger, a lot larger, exactly the same, a little smaller, or a lot smaller than your estimate?)

   b. Do you think the mean summarizes the data set well? Explain your reasoning.

Are you ready for more?

Invent a data set with a mean that is significantly lower than what you would consider a typical value for the data set.

13.3: Finding the Middle

1. Your teacher will give you an index card. Write your first and last names on the card. Then record the total number of letters in your name. After that, pause for additional instructions from your teacher.

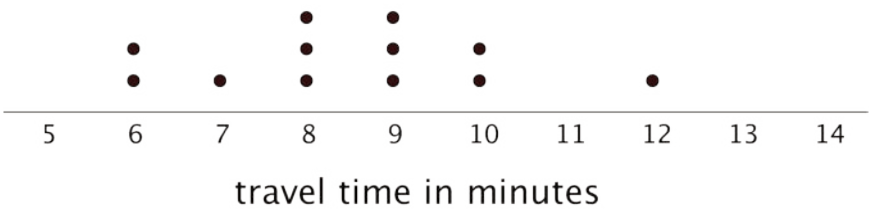
2. Here is the data set on numbers of siblings from an earlier activity.

1	0	2	1	7	0	2	0	1	10

   a. Sort the data from least to greatest, and then find the **median**.

   b. In this situation, do you think the median is a good measure of a typical number of siblings for this group? Explain your reasoning.

3. Here is the dot plot showing the travel time, in minutes, of Elena's bus rides to school.



   a. Find the median travel time. Be prepared to explain your reasoning.

   b. What does the median tell us in this context?

Lesson 13 Summary

The **median** is another measure of center of a distribution. It is the middle value in a data set when values are listed in order. Half of the values in a data set are less than or equal to the median,and half of the values are greater than or equal to the median.

To find the median, we order the data values from least to greatest and find the number in the middle.

Suppose we have 5 dogs whose weights, in pounds, are shown in the table. The median weight for this group of dogs is 32 pounds because three dogs weigh less than or equal to 32 pounds and three dogs weigh greater than or equal to 32 pounds.

20	25	32	40	55

Now suppose we have 6 cats whose weights, in pounds, are as shown in the table. Notice that there are *two* values in the middle: 7 and 8.

4	6	7	8	10	10

The median weight must be between 7 and 8 pounds, because half of the cats weigh less or equal to 7 pounds and half of the cats weigh greater than or equal to 8 pounds.

In general, when we have an even number of values, we take the number exactly in between the two middle values. In this case, the median cat weight is 7.5 pounds because $(7 + 8) \div 2 = 7.5$.

Lesson 13: The Median of a Data Set

Cool Down: Practicing the Piano

Jada and Diego are practicing the piano for an upcoming rehearsal. The tables list the number of minutes each of them practiced in the past few weeks.

Jada's practice times:

10	10	20	15	25	25	8	15
20	20	35	25	40			

Diego's practice times:

25	10	15	30	15	20	20	25
30	45						

1. Find the median of each data set.

2. Explain what the medians tell you about Jada's and Diego's piano practice.

Unit 8 Lesson 13 Cumulative Practice Problems

1. Here is data that shows a student's scores for 10 rounds of a video game.

   130 150 120 170 130 120 160 160 190 140

   What is the median score?

   A. 125

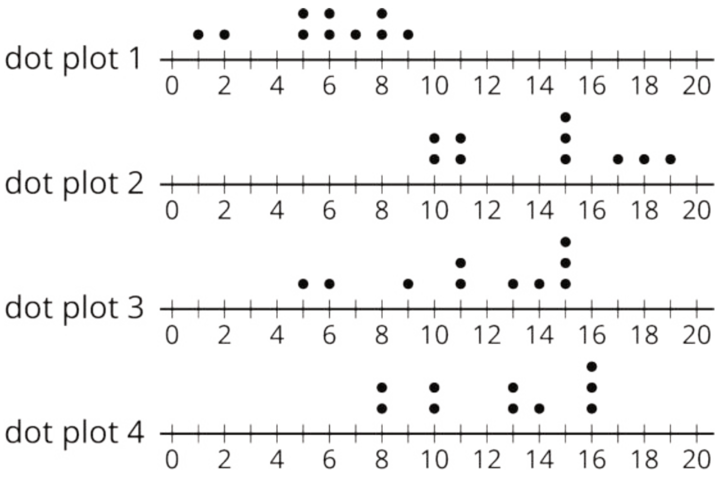
   B. 145

   C. 147

   D. 150

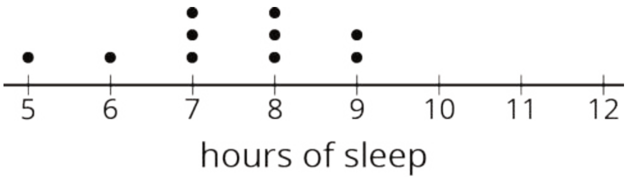
2. When he sorts the class's scores on the last test, the teacher notices that exactly 12 students scored better than Clare and exactly 12 students scored worse than Clare. Does this mean that Clare's score on the test is the median? Explain your reasoning.

3. The medians of the following dot plots are 6, 12, 13, and 15, but not in that order. Match each dot plot with its median.



4. Invent a data set with five numbers that has a mean of 10 and a median of 12.

5. Ten sixth-grade students reported the hours of sleep they get on nights before a school day. Their responses are recorded in the dot plot.

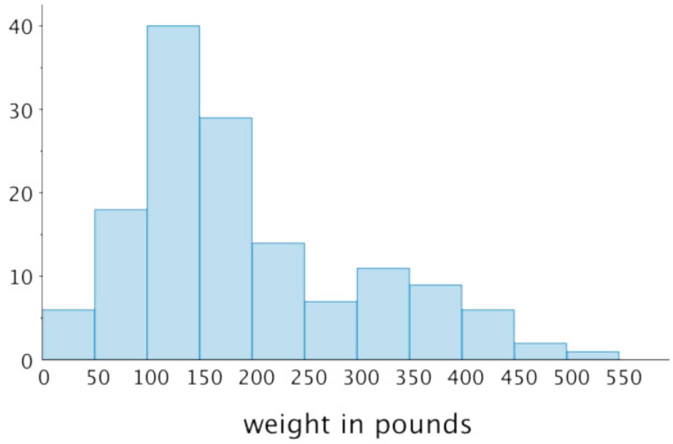


Looking at the dot plot, Lin estimated the mean number of hours of sleep to be 8.5 hours. Noah's estimate was 7.5 hours. Diego's estimate was 6.5 hours.

Which estimate do you think is best? Explain how you know.

(From Unit 8, Lesson 10.)

6. In one study of wild bears, researchers measured the weights, in pounds, of 143 wild bears that ranged in age from newborn to 15 years old. The data were used to make this histogram.



   a. What can you say about the heaviest bear in this group?

   b. What is a typical weight for the bears in this group?

   c. Do more than half of the bears in this group weigh less than 250 pounds?

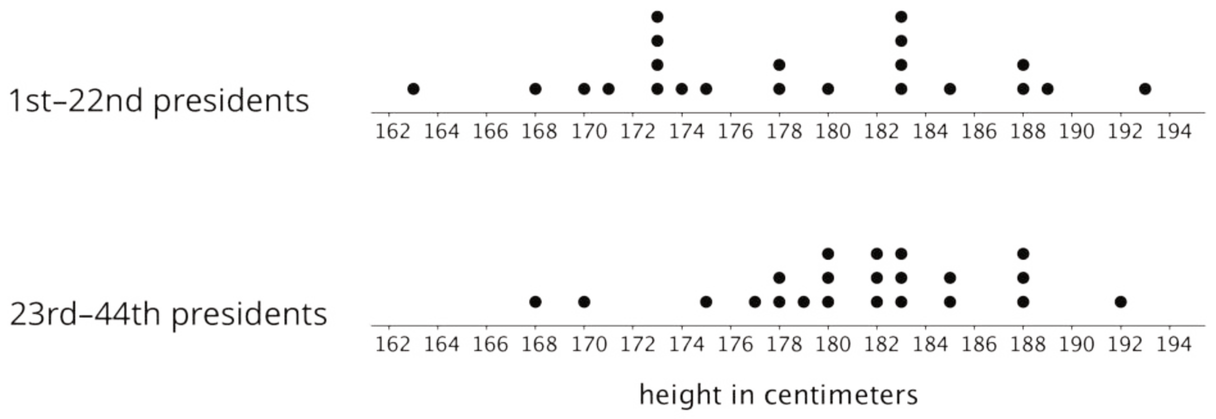
   d. If weight is related to age, with older bears tending to have greater body weights, would you say that there were more old bears or more young bears in the group? Explain your reasoning.

(From Unit 8, Lesson 8.)

Lesson 14: Comparing Mean and Median

14.1: Heights of Presidents

Here are two dot plots. The first dot plot shows the heights of the first 22 U.S. presidents. The second dot plot shows the heights of the next 22 presidents.



Based on the two dot plots, decide if you agree or disagree with each of the following statements. Be prepared to explain your reasoning.

1. The median height of the first 22 presidents is 178 centimeters.
2. The mean height of the first 22 presidents is about 183 centimeters.
3. A typical height for a president in the second group is about 182 centimeters.
4. U.S. presidents have become taller over time.
5. The heights of the first 22 presidents are more alike than the heights of the second 22 presidents.
6. The MAD of the second data set is greater than the MAD of the first set.

14.2: The Tallest and the Smallest in the World

Your teacher will provide the height data for your class. Use the data to complete the following questions.

1. Find the mean height of your class in centimeters.

2. Find the median height in centimeters. Show your reasoning.

3. Suppose that the world’s tallest adult, who is 251 centimeters tall, joined your class.

   a. Discuss the following questions with your group and explain your reasoning.

      - How would the mean height of the class change?

      - How would the median height change?

   b. Find the new mean.

   c. Find the new median.

   d. Which measure of center—the mean or the median—changed more when this new person joined the class? Explain why the value of one measure changed more than the other.

4. The world's smallest adult is 63 centimeters tall. Suppose that the world's tallest and smallest adults both joined your class.

   a. Discuss the following questions with your group and explain your reasoning.

      - How would the mean height of the class change from the original mean?

      - How would the median height change from the original median?

   b. Find the new mean.

   c. Find the new median.

   d. How did the measures of center—the mean and the median—change when these two people joined the class? Explain why the values of the mean and median changed the way they did.

14.3: Mean or Median?

1. Your teacher will give you six cards. Each has either a dot plot or a histogram. Sort the cards into *two* piles based on the distributions shown. Be prepared to explain your reasoning.

2. Discuss your sorting decisions with another group. Did you have the same cards in each pile? If so, did you use the same sorting categories? If not, how are your categories different?

   Pause here for a class discussion.

3. Use the information on the cards to answer the following questions.
    a. Card A: What is a typical age of the dogs being treated at the animal clinic?
    b. Card B: What is a typical number of people in the Irish households?
    c. Card C: What is a typical travel time for the New Zealand students?
    d. Card D: Would 15 years old be a good description of a typical age of the people who attended the birthday party?
    e. Card E: Is 15 minutes or 24 minutes a better description of a typical time it takes the students in South Africa to get to school?
    f. Card F: Would 21.3 years old be a good description of a typical age of the people who went on a field trip to Washington, D.C.?
4. How did you decide which measure of center to use for the dot plots on Cards A–C? What about for those on Cards D–F?

Are you ready for more?

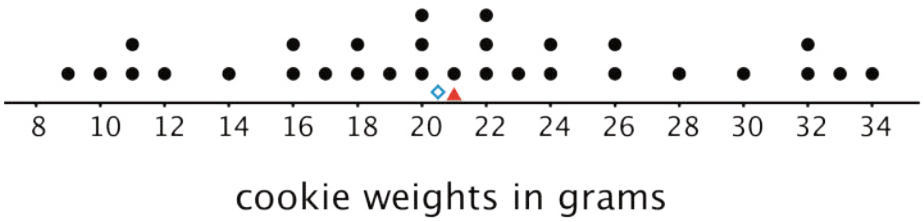
Most teachers use the mean to calculate a student's final grade, based on that student's scores on tests, quizzes, homework, projects, and other graded assignments.

Diego thinks that the median might be a better way to measure how well a student did in a course. Do you agree with Diego? Explain your reasoning.

Lesson 14 Summary

Both the mean and the median are ways of measuring the center of a distribution. They tell us slightly different things, however.

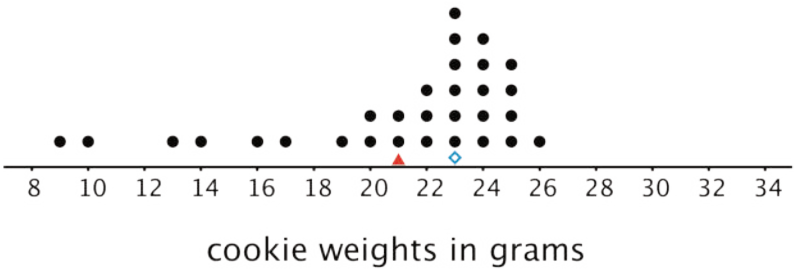
The dot plot shows the weights of 30 cookies. The mean weight is 21 grams (marked with a triangle). The median weight is 20.5 grams (marked with a diamond).



The mean tells us that if the weights of all cookies were distributed so that each one weighed the same, that weight would be 21 grams. We could also think of 21 grams as a balance point for the weights of all of the cookies in the set.

The median tells us that half of the cookies weigh more than 20.5 grams and half weigh less than 20.5 grams. In this case, both the mean and the median could describe a typical cookie weight because they are fairly close to each other and to most of the data points.

Here is a different set of 30 cookies. It has the same mean weight as the first set, but the median weight is 23 grams.



In this case, the median is closer to where most of the data points are clustered and is therefore a better measure of center for this distribution. That is, it is a better description of a typical cookie weight. The mean weight is influenced (in this case, pulled down) by a handful of much smaller cookies, so it is farther away from most data points.

In general, when a distribution is symmetrical or approximately symmetrical, the mean and median values are close. But when a distribution is not roughly symmetrical, the two values tend to be farther apart.

Mean or Median?

Card A

Ages of 20 dogs being treated at an animal hospital

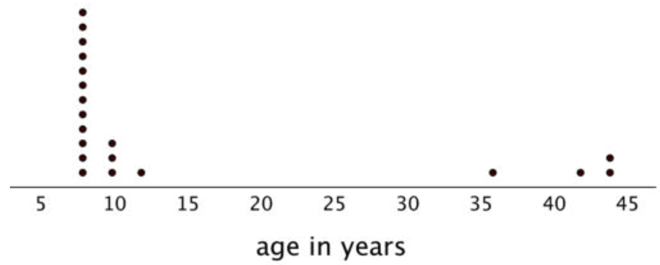
Mean: Approximately 7.8 years old

Median: 8 years old

Mean or Median?

Card D

Ages of 20 people at an 8-year-old's birthday party



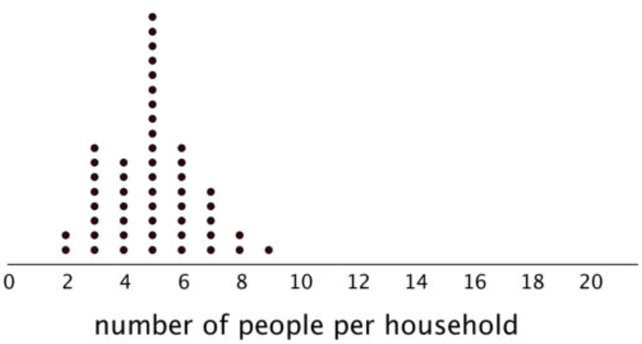
Mean: Approximately 15 years old

Median: 8 years old

Mean or Median?

Card B

Size of household of 50 people in Ireland



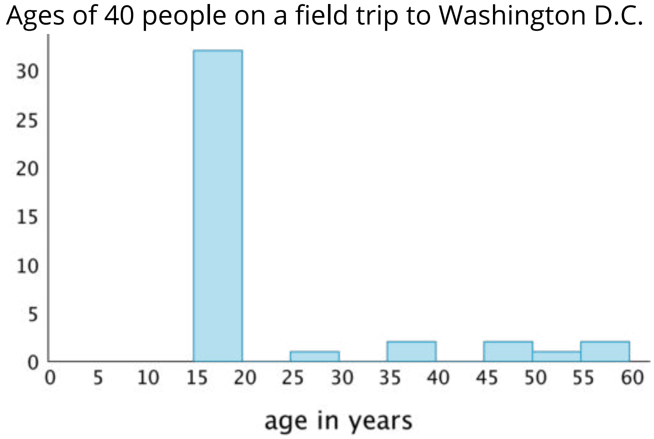
Mean: Approximately 5 people per household

Median: 5 people per household

Mean or Median?

Card F

Ages of 40 people on a field trip to Washington D.C.



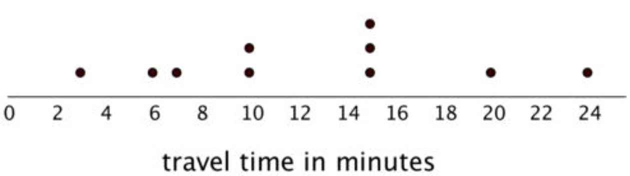
Mean: Approximately 21.3 years old

Median: 15 years old

Mean or Median?

Card C

Time it took ten sixth-grade students in New Zealand to travel to school



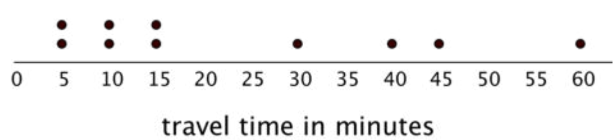
Mean: Approximately 12.5 minutes

Median: 12.5 minutes

Mean or Median?

Card E

Time it took ten sixth-grade students in South Africa to travel to school



Mean: Approximately 24 minutes

Median: 15 minutes

Lesson 14: Comparing Mean and Median

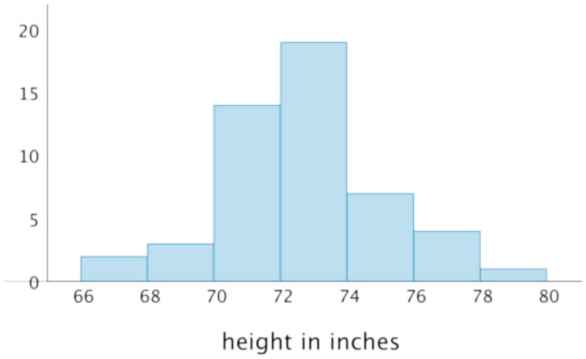
Cool Down: Which Measure of Center to Use?

For each dot plot or histogram:

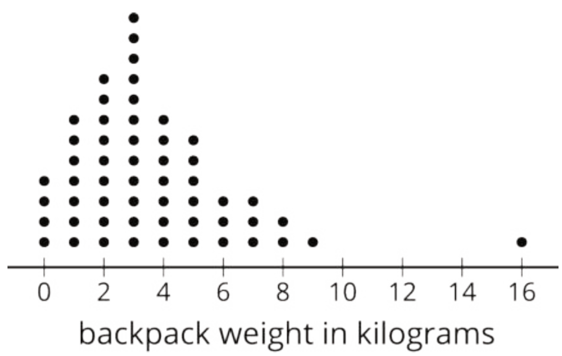
a. Predict if the mean is greater than, less than, or approximately equal to the median. Explain your reasoning.

b. Which measure of center—the mean or the median—better describes a typical value for the following distributions?

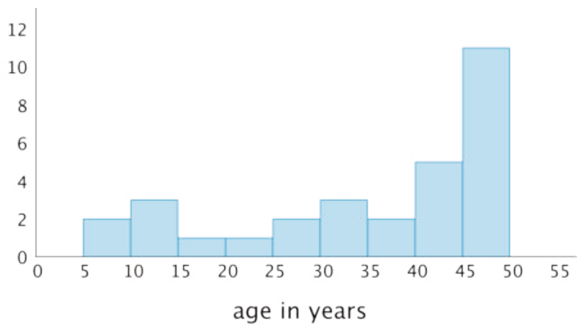
1. Heights of 50 NBA basketball players



2. Backpack weights of 55 sixth-grade students



3. Ages of 30 people at a family dinner party



Unit 8 Lesson 14 Cumulative Practice Problems

1. Here is a dot plot that shows the ages of teachers at a school.

   Which of these statements is true of the data set shown in the dot plot?

   A. The mean is less than the median.

   B. The mean is approximately equal to the median.

   C. The mean is greater than the median.

   D. The mean cannot be determined.

2. Priya asked each of five friends to attempt to throw a ball in a trash can until they succeeded. She recorded the number of unsuccessful attempts made by each friend as: 1, 8, 6, 2, 4. Priya made a mistake: The 8 in the data set should have been 18.

   How would changing the 8 to 18 affect the mean and median of the data set?

   A. The mean would decrease and the median would not change.

   B. The mean would increase and the median would not change.

   C. The mean would decrease and the median would increase.

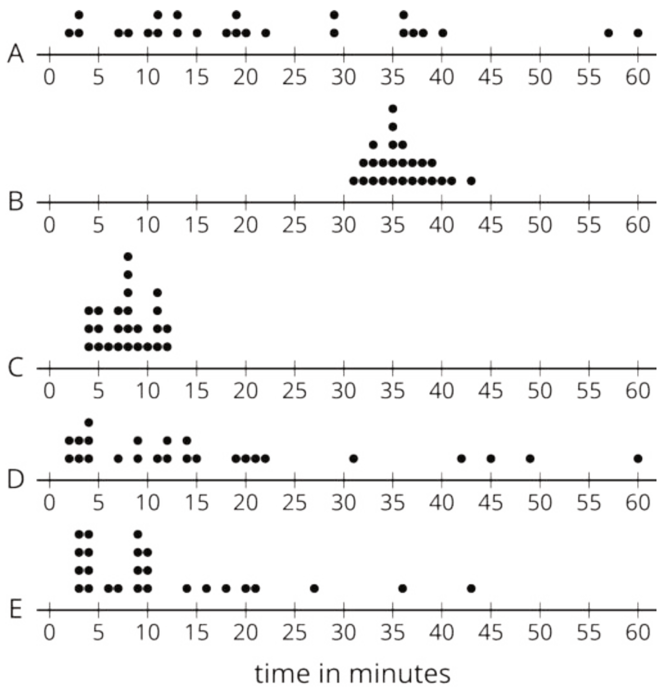
   D. The mean would increase and the median would increase.

3. In his history class, Han's homework scores are:

100	100	100	100	95	100	90	100	0

   The history teacher uses the mean to calculate the grade for homework. Write an argument for Han to explain why median would be a better measure to use for his homework grades.

4. The dot plots show how much time, in minutes, students in a class took to complete each of five different tasks. Select **all** the dot plots of tasks for which the mean time is approximately equal to the median time.



5. Zookeepers recorded the ages, weights, genders, and heights of the 10 pandas at their zoo. Write two statistical questions that could be answered using these data sets.

(From Unit 8, Lesson 2.)

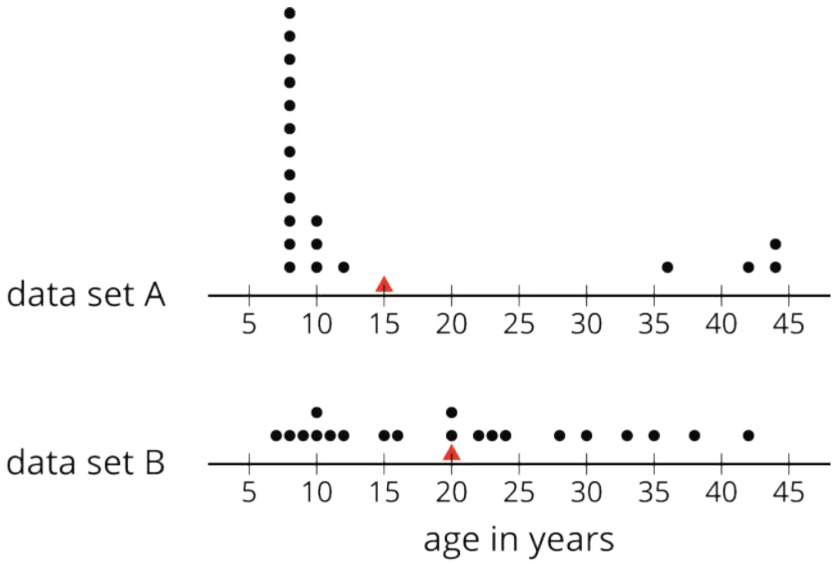
6. Here is a set of coordinates. Draw and label an appropriate pair of axes and plot the points. $A = (1,0)$, $B = (0,0.5)$, $C = (4,3.5)$, $D = (1.5,0.5)$

(From Unit 7, Lesson 12.)

Lesson 15: Quartiles and Interquartile Range

15.1: Notice and Wonder: Two Parties

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?

15.2: The Five-Number Summary

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

Are you ready for more?

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

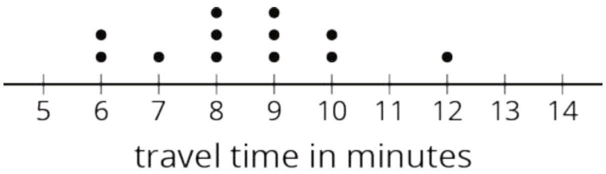
minimum: 5   Q1: 6   Q2: 27   Q3: 32   maximum: 60

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.

15.3: Range and Interquartile Range

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.

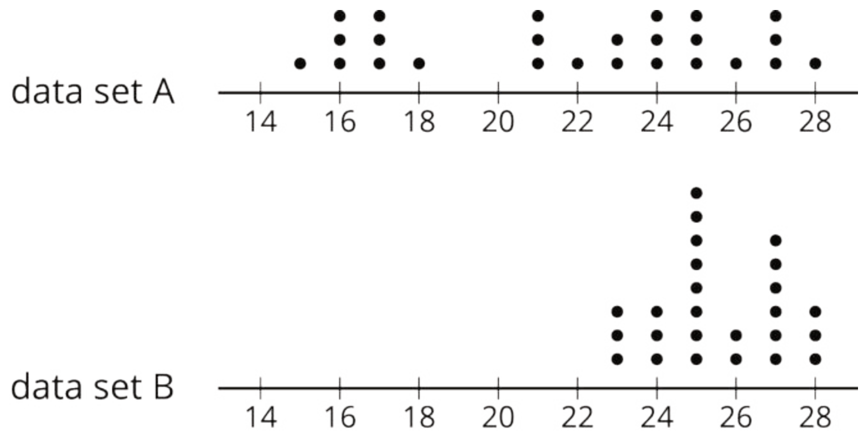
2. The **range** is one way to describe the *spread* of values in a data set. It is the difference between the maximum and minimum. What is the range of Elena’s travel times?

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.

Lesson 15 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.

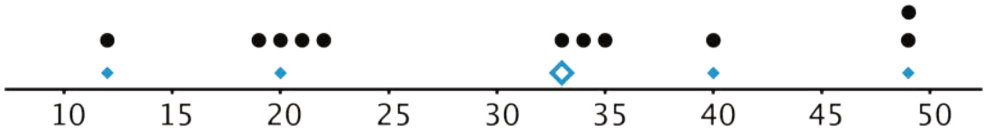
12	19	20	21	22	33	34	35	40	40	49
		Q1			Q2			Q3		

- 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.

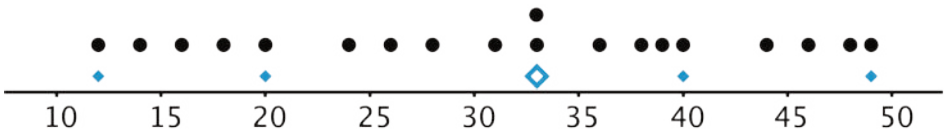
The difference between the maximum and minimum values of a data set is the **range**. The difference between Q3 and Q1 is the **interquartile range (IQR)**. Because the distance between Q1 and Q3 includes the middle two-fourths of the distribution, the values between those two quartiles are sometimes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



Lesson 15: Quartiles and Interquartile Range

Cool Down: How Far Can You Throw?

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.

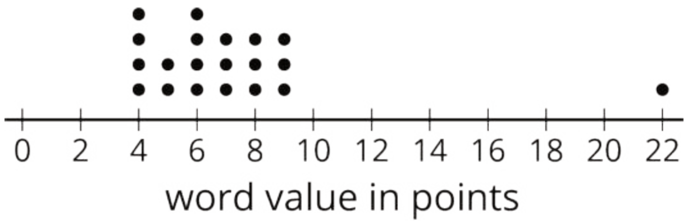
Unit 8 Lesson 15 Cumulative Practice Problems

1. 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?

2. 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?

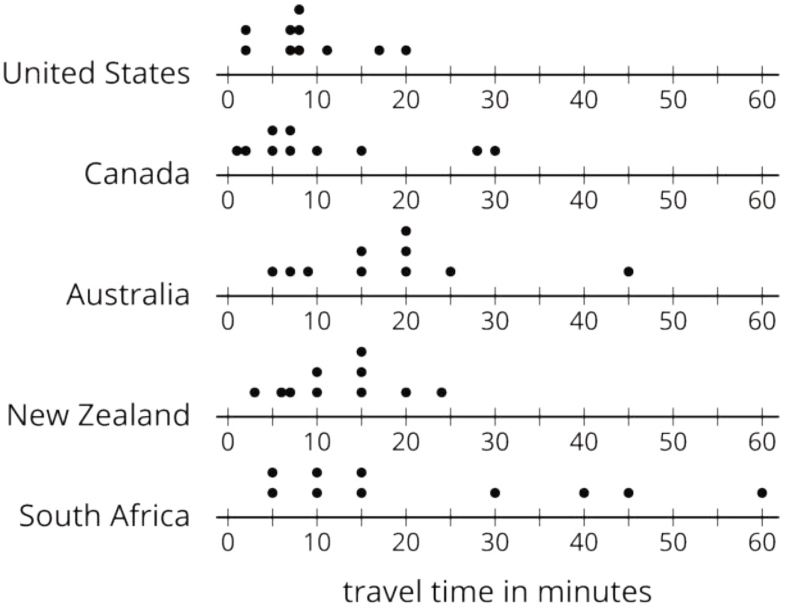
    b. What is the first quartile (Q1)?

    c. What is the third quartile (Q3)?

    d. What is the interquartile range (IQR)?

3. 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.

4. 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

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

(From Unit 7, Lesson 12.)

6. 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?

(From Unit 6, Lesson 7.)

Lesson 16: Box Plots

16.1: Notice and Wonder: Puppy Weights

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?

16.2: Human Box Plot

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.

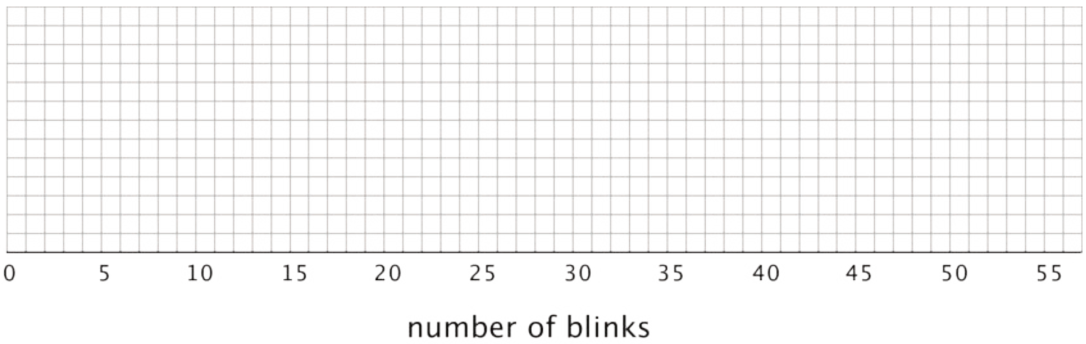
Pause for additional instructions from your teacher.

16.3: Studying Blinks

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				

1. 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?

2. 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.

3. 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

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.

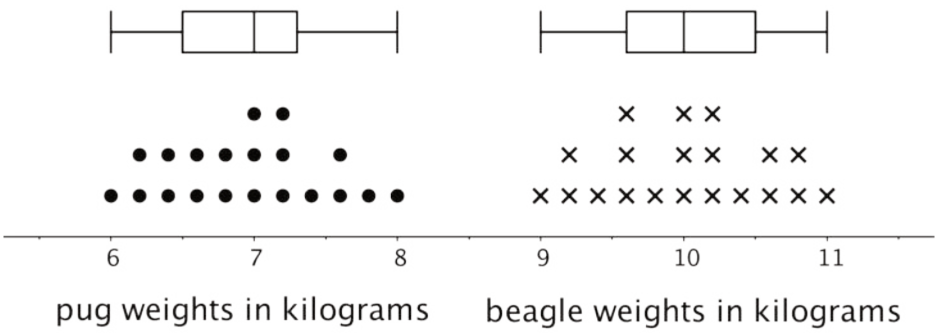
Lesson 16 Summary

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.

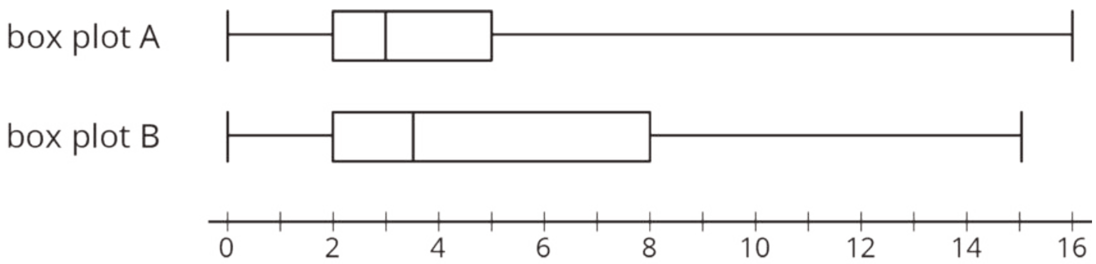


We can tell from the box plots that, in general, the pugs in the group are lighter than the beagles: the median weight of pugs is 7 kilograms and the median weight of beagles is 10 kilograms. Because the two box plots are on the same scale and the rectangles have similar widths, we can also tell that 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.

Lesson 16: Box Plots

Cool Down: Boxes and Dots

1. Here are two box plots that summarize two data sets. Do you agree with each of the following statements?



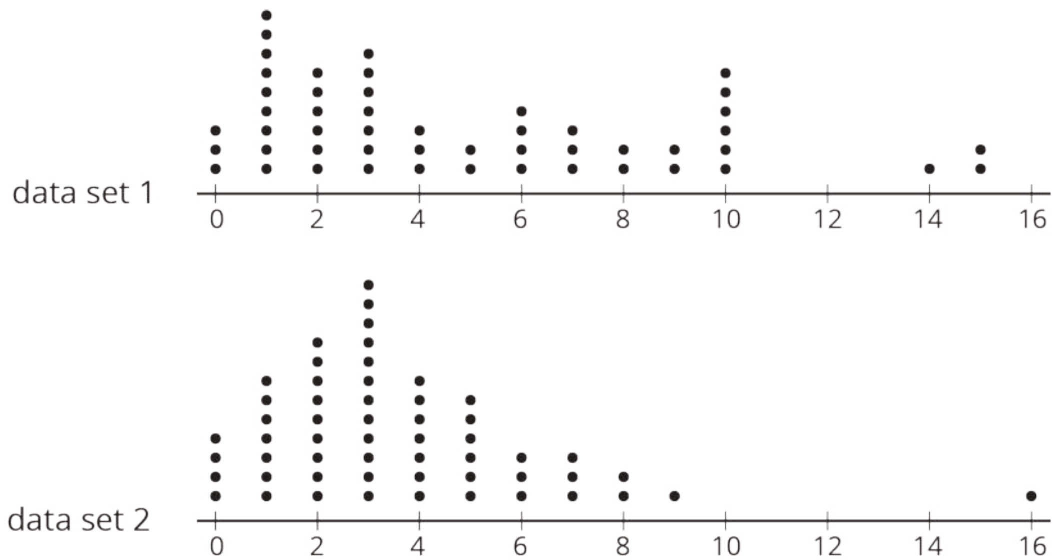
   a. Both data sets have the same range.

   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.

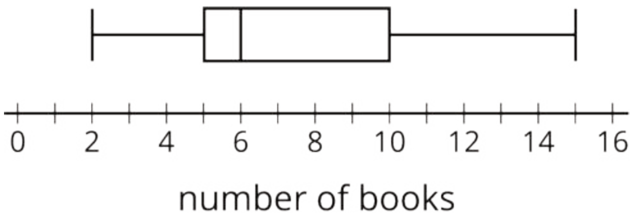
   d. Box plot A shows a data set that has a quarter of its values between 2 and 5.

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.



Unit 8 Lesson 16 Cumulative Practice Problems

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



   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)?

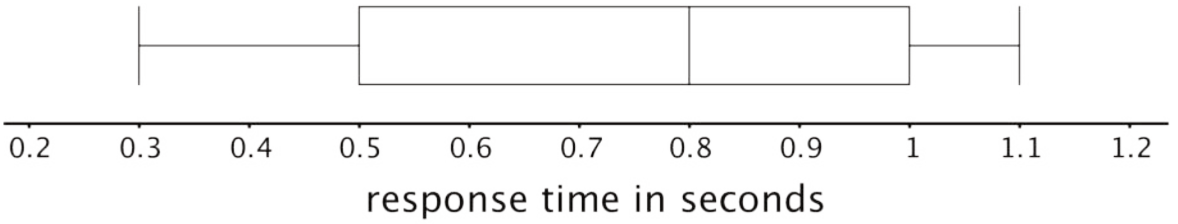
2. Use this five-number summary to draw a box plot. All values are in seconds.

   - Minimum: 40
   - First quartile (Q1): 45
   - Median: 48
   - Third quartile (Q3): 50
   - Maximum: 60

3. The data shows the number of hours per week that each of 13 seventh-grade students spent doing homework. Create a box plot to summarize the data.

3	10	12	4	7	9	5	5
11	11	5	12	11			

4. 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?



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

(From Unit 8, Lesson 14.)

6. Jada earns money from babysitting, walking her neighbor'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.

   (From Unit 8, Lesson 9.)