Homework Helpers

Grade 6
Module 6

G6-M6-Lesson 1: Posing Statistical Questions

1. For each of the following, determine whether the question is a statistical question. Give a reason for your answer.

   a. How many bricks are in this wall?

   This is not a statistical question because this question is not answered by collecting data that vary.

   > A *statistical question* is one that can be answered by collecting data, and it is anticipated that the data (information) collected to answer the question will vary.

   > To answer this question, I can just count the bricks. I don't have to collect data since there is just one answer.

   b. On average, how old are the dogs that live on this street?

   This is a statistical question because it would be answered by collecting data on the ages of all the dogs, and there is variability in the ages of the dogs.

   > I anticipate variability in the data because the dogs on the street are likely a variety of ages (e.g., some dogs are young; some dogs are old).

   c. How many days are there until summer break?

   This is not a statistical question because this question is not answered by collecting data that vary.

   d. How many minutes do sixth graders typically spend outside every week?

   This is a statistical question because it would be answered by collecting data on the number of minutes sixth graders spend outside every week, and we expect variability in how many minutes are recorded for each student. They will not all be the same.

2. Identify each of the following data sets as categorical (C) or numerical (N). Explain your answer.
   a. Height of sixth graders

   N; the height can be measured as number of inches, for example, so the data set is numerical.

   In a numerical data set, each value is a number.

   b. The hair color of 20 adults

   C; hair color is categorical because hair colors are categories.

   In a categorical data set, the values are categories, not numbers.

3. Rewrite the following question as a statistical question:
   How many minutes do you spend on homework each week?

   Answers may vary. A sample response is provided below.

   How many minutes do sixth grade students typically spend on homework each week?

   In a numerical data set, the values are numbers, not categories.

4. Write a statistical question that would be answered by collecting data from the animals that reside at the Bambelela Wildlife Sanctuary?

   Answers may vary. A sample response is provided below.

   How long do the cheetahs spend at the watering hole in July?

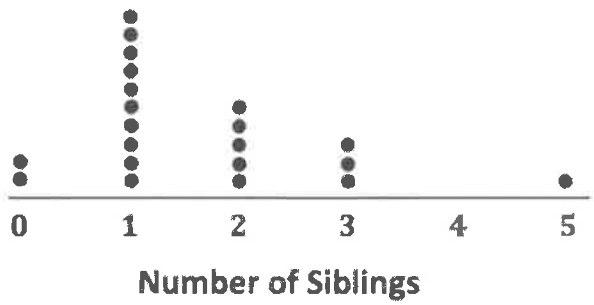
   Because different cheetahs will spend different amounts of time at the watering hole in July, this is a statistical question because the data that will be collected will vary.

5. Are the data you would collect to answer the question you wrote in Problem 4 categorical or numerical? Explain your answer.

   Numerical. The time at the watering hole can be measured in minutes, for example, so each value in the data set is a number, and the data set is numerical.

G6-M6-Lesson 2: Displaying a Data Distribution

1. The dot plot below shows the number of siblings of the sixth grade students in Ms. Baker's class.



a. What statistical question do you think could be answered using these data?

How many siblings does the typical sixth grader have?

> To answer this question, I have to collect data, and I anticipate there is variability in the data set since every sixth grader will not have the same number of siblings.

b. What was the most number of siblings recorded in the class?

5 siblings

c. What was the least number of siblings recorded in the class?

0 siblings

d. What is the most common number of siblings (the number of siblings that occurred most often)?

1 sibling

> I can see that the most common number of siblings students recorded is 1 because the number of dots for that number is more than any other number.

e. How many students recorded the most common number of siblings?

10 students

f. How many students had more than 2 siblings?

4 students

> I can count the number of dots for 3, 4, and 5 siblings.

g. If a new student joins the class and has 1 sibling, how does this student compare with the other students?

The new student would have the most common number of siblings.

2. Read each of the following statistical questions. Write a description of what the dot plot of the data collected to answer the question might look like. Your description should include a description of the spread of the data and the center of the data.

   a. How many minutes do sixth graders spend in the cafeteria eating lunch during a typical school day?

   Most sixth graders are in the cafeteria for the same number of minutes, so the spread would be small. Differences may exist for those students who have to leave early from school and do not eat lunch in the cafeteria. Student responses vary based on their estimates of the number of minutes sixth graders spend in the cafeteria eating lunch during the school day.

   b. What is the number of books owned by the sixth graders in our class?

   These data would have a very big spread. Some students might have very few books, while others could have many books. A typical value of 10 (or something similar) would identify a center. In this case, the center is based on the number most commonly reported by students.

   For the center, students may describe a number that occurs most often, the number in the middle, or the average. It is important to gauge students' thinking about what center means.

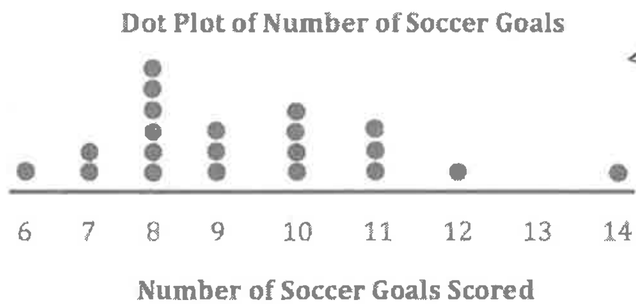
EUREKA MATH

G6-M6-Lesson 3: Creating a Dot Plot

1. The data below are the number of goals scored by a professional indoor soccer team over its last 21 games.

   6 7 7 8 8 8 8 8 8 9 9 9 10 10 10 10 11 11 11 12 14

   a. Make a dot plot of the number of goals scored.

   The data values are already in order from least to greatest. Since the smallest number is 6 and the largest number is 14, I can list the numbers on the scale sequentially, starting at 6, ending at 14, and counting by 1. I can use the given data to plot the points.

   b. What number of goals describes the center of the data?

   The center of the data is around 9. (Answers may vary, but student responses should be around the center of the data distribution.)

   Since there are 21 values in the data set, the center of the data will be the 11^{th} value, which is 9.

   c. What is the least and most number of goals scored by the team?

   The least number of goals scored is 6, and 14 is the most.

   d. Over the 21 games played, the team lost 9 games. Circle the dots on the plot that you think represent the games that the team lost. Explain your answer.

   Students will most likely circle the lowest 9 scores, but answers may vary. Students need to supply an explanation in order to defend their answers.

2. A sixth grader collected data on the number of hours 15 students read independently each week. The following are the number of hours for the 15 students:

3 2 4 4 5 7 6 3 6 3 7 6 6 7 1

a. Complete the frequency table.

Number of Hours	Tally	Frequency
1	|	1
2	|	1
3	| | |	3
4	| |	2
5	|	1
6	| | | |	4
7	| | |	3

A frequency table lists possible data values and how often each value occurs.

b. What number of hours describes the center of the data?

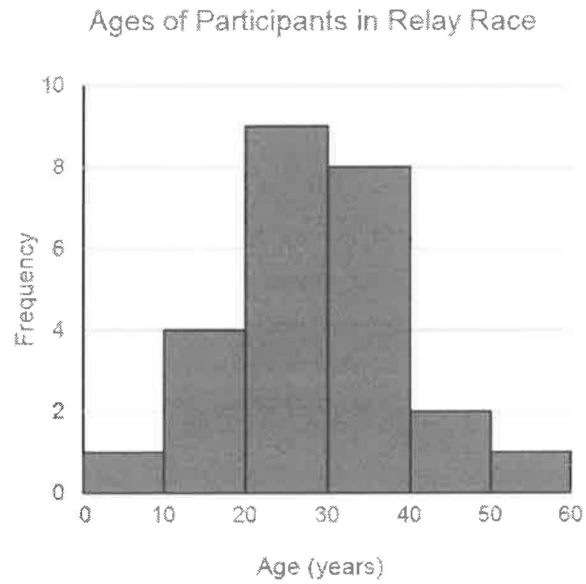
5

c. What number of hours occurs most often for these 15 students?

6

G6-M6-Lesson 4: Creating a Histogram

1. The following histogram summarizes the ages of the participants in a community relay race.



I can look for the interval that has the greatest frequency.

a. Which age interval contains the most participants? How many participants are represented in that interval?

The interval 20 to 30 contains the most participants. There are 9 participants whose ages fall into that category.

b. Describe the shape of the histogram.

The shape of the histogram is approximately symmetrical.

c. What does the histogram tell you about the ages of the participants in the relay race?

Most of the ages are between 20 and 40, with a few people with ages much smaller or larger than the rest.

d. Which interval describes the center of the ages of the participants?

The interval of 20 to 40 describes the center of the ages. Since this data distribution is approximately symmetrical, the center is probably around 30. (Answers may vary, but student responses should be around the center of the data distribution.)

e. An age of 19 would be included in which interval?

An age of 19 is in the interval from 10 to 20.

2. Listed are the prices for various items sold at a garage sale.

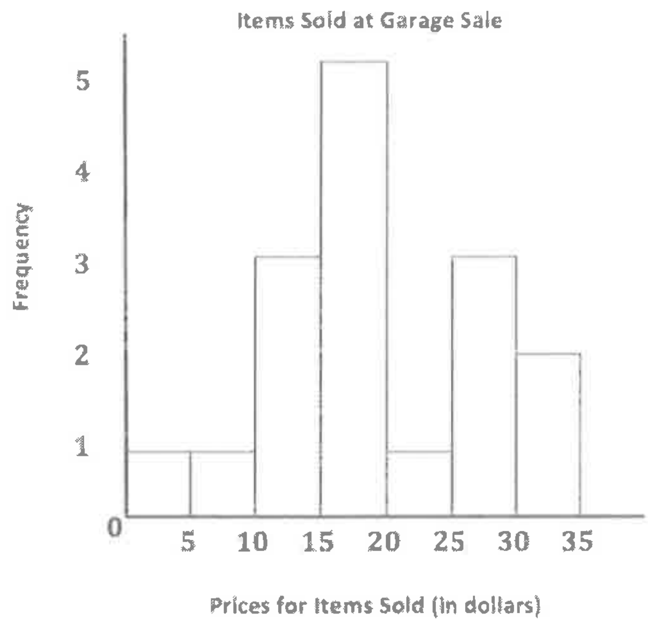
$\$4$	$\$8$	$\$16$	$\$17$	$\$16$	$\$18$	$\$11$	$\$26$
$\$34$	$\$28$	$\$23$	$\$15$	$\$10$	$\$30$	$\$29$	$\$13$

a. Complete the frequency table using the given intervals of width 5.

Prices for Items Sold	Tally	Frequency
$\$0-<\$5$	|	1
$\$5-<\$10$	|	1
$\$10-<\$15$	|||	3
$\$15-<\$20$	~~||||~~	5
$\$20-<\$25$	|	1
$\$25-<\$30$	|||	3
$\$30-<\$35$	||	2

For the interval $\$0-<\$5$, I can look for the prices, $\$0, \$1, \$2, \$3$, and $\$4$ in the data set. There is only one price, $\$4$, that fits in this category.

b. Draw a histogram of the garage sale data.



c. Describe the center and shape of the histogram.

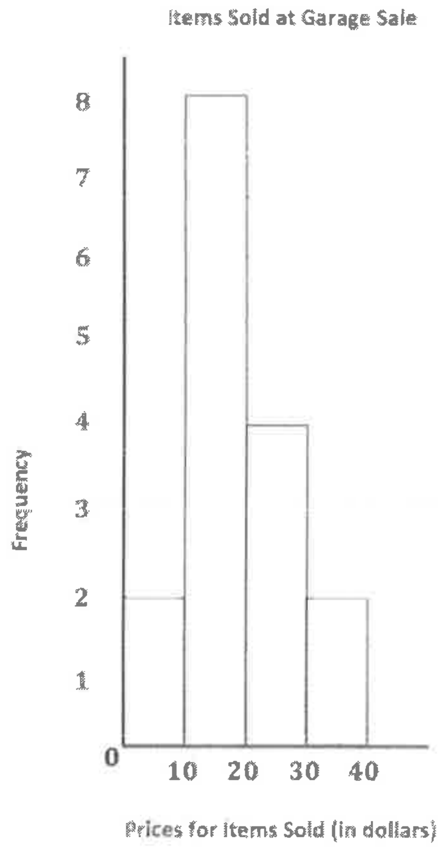
The center is around 18; the histogram is mound shaped and skewed slightly to the left. (Answers may vary, but student responses for describing the center should be around the center of the data distribution.)

d. In the frequency table below, the intervals are changed. Using the garage sale data above, complete the frequency table with intervals of width 10.

Prices for Items Sold	Tally	Frequency
$0–< $10	||	2
$10–< $20	~~||||~~ |||	8
$20–< $30	||||	4
$30–< $40	||	2

Even though the data is the same, I can see how the frequency table looks different because the width of the interval has changed.

e. Draw a histogram.



3. Use the histograms that you constructed in Problem 2 parts (b) and (e) to answer the following questions.
   a. Why are there fewer bars in the histogram in part (e) than the histogram in part (b)?

      There are fewer bars in part (e) because the width of the interval changed from $\$5$ to $\$10$.

   b. Did the shape of the histogram in part (e) change from the shape of the histogram in part (b)?

      Generally, both are mound shaped, but the histogram in part (e) is skewed less to the left.

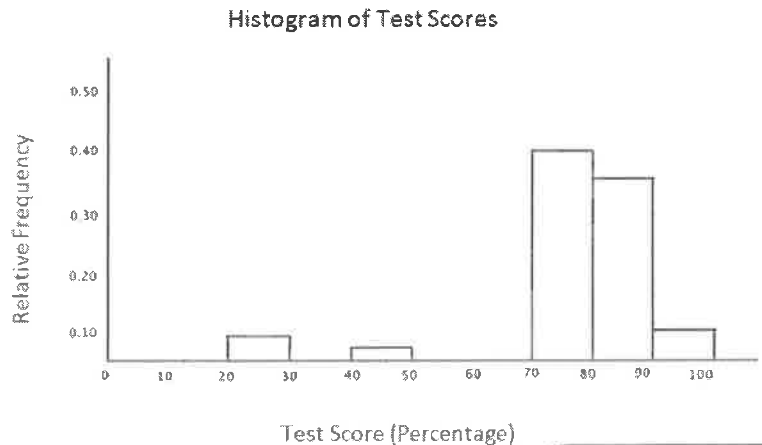
   c. Did your estimate of the center change from the histogram in part (b) to the histogram in part (e)?

      No; the centers of the two histograms are about the same.

EUREKA MATH™

G6-M6-Lesson 5: Describing a Distribution Displayed in a Histogram

1. Below is a relative frequency histogram of the test scores (in percentage) of a selected group of sixth graders.



   a. Describe the shape of the relative frequency histogram.

   The shape is skewed to the left.

   > This graph is skewed left because it has a tail that is longer on the left side. The graph is skewed toward the smaller values.

   b. What does the shape tell you about the maximum test score (in percentage) of the selected group of sixth graders?

   The shape tells us most of the sixth graders have a test score that is between 70% and 90% but that some students have a test score that is quite a bit lower than the others.

   c. Clara said that more than half of the data values are in the interval from 80% to 100%. Do you agree with Clara? Why or why not?

   I do not agree because that interval contains 45% of the data.

   > In the interval 80–90, the relative frequency is 0.35 (or 35%). In the interval 90–100, the relative frequency is 0.10 (or 10%). The cumulative relative frequency is $0.35 + 0.10 = 0.45$, which is 45% and less than half.

2. The frequency table below shows the length of selected professional football games over the past 6 months.

Length of Game (minutes)	Tally	Frequency	Relative Frequency
160–< 170	||	2	$\frac{2}{25} = 0.08$
170–< 180	|||	3	$\frac{3}{25} = 0.12$
180–< 190	~~||||~~ |	6	$\frac{6}{25} = 0.24$
190–< 200	||||	4	$\frac{4}{25} = 0.16$
200–< 210	~~||||~~ |	6	$\frac{6}{25} = 0.24$
210–< 220	|||	3	$\frac{3}{25} = 0.12$
220–< 230	|	1	$\frac{1}{25} = 0.04$

a. Complete the relative frequency column. Round the relative frequencies to the nearest hundredth.

See the table above.

Relative frequency is the frequency for an interval divided by the total number of data values.

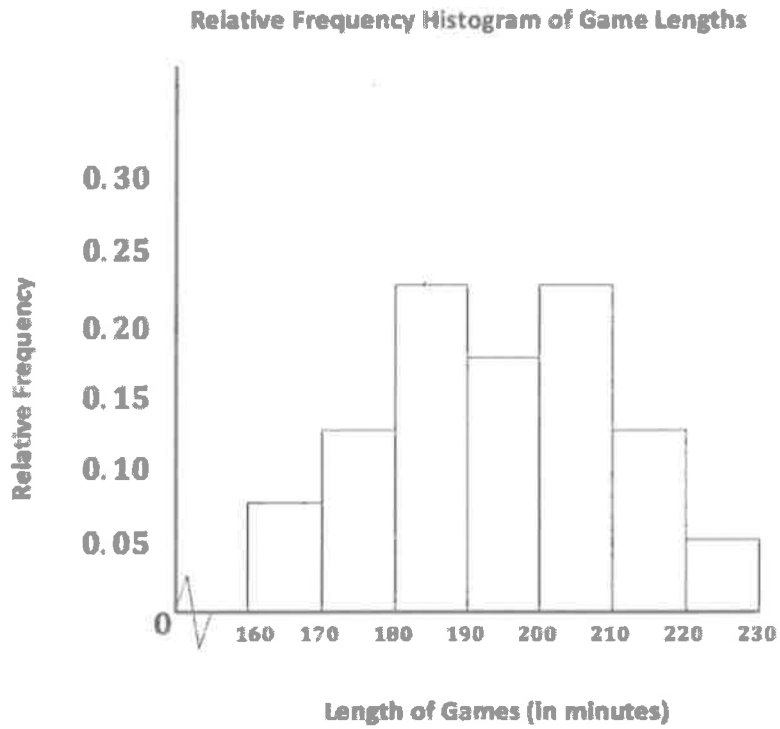
b. What percentage of the football games are greater than or equal to 210 minutes?

$0.12 + 0.04 = 0.16$

16% *of the game lengths are greater than or equal to* 210 *minutes.*

I can add the relative frequencies for the game lengths greater than or equal to 210 minutes (there are two intervals) and then determine the percentage of games in that category.

c. Draw a relative frequency histogram. (Hint: Label the relative frequency scale starting at 0 and going up to 0.30, marking off intervals of 0.05).



d. Describe the shape of the relative frequency histogram.

The histogram is mound shaped and approximately symmetric.

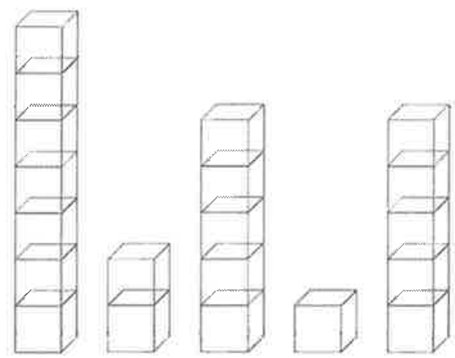
e. What does the shape tell you about the length of football games?

The shape tells us the length of most football games is between 180 and 210 minutes.

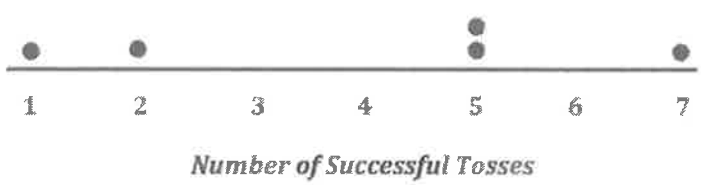
G6-M6-Lesson 6: Describing the Center of a Distribution Using the Mean

1. A game is played where ten tennis balls are tossed into a bucket from a specific distance. The numbers of successful tosses for five students are 7, 2, 5, 1, 5.

   a. Draw a representation of the data using cubes where one cube represents one successful toss of a tennis ball into the bucket.



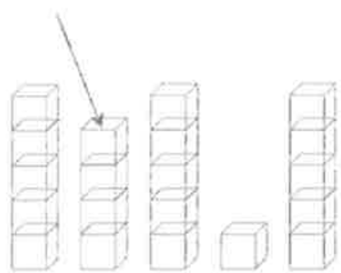
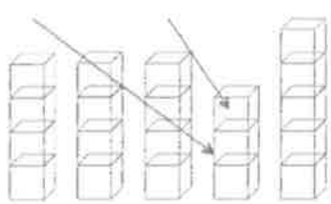
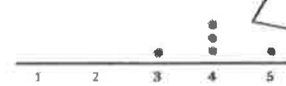
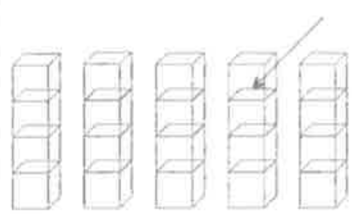
   b. Represent the original data set using a dot plot.



In the data set, there is one student each who successfully tossed the balls 1, 2, and 7 times, so I can place one dot above each of those numbers in the dot plot to represent each student. There are two students who successfully tossed the ball 5 times each, so I can place two dots above the 5 to represent these two students.

2. Find the mean number of successful tosses for this data set using the fair share method. For each step, show the cubes representation and the corresponding dot plot. Explain each step in words in the context of the problem. You may move more than one successful toss in a step, but be sure your explanation is clear. You must show two or more steps.

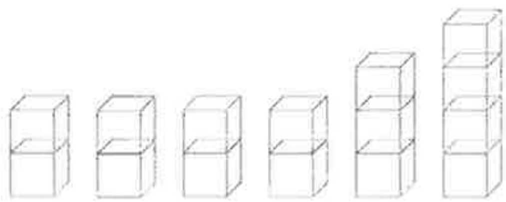
There are several ways of getting to the final fair share cubes representation where each of the five stacks contains four cubes. Ideally, students move one cube at a time because the leveling is seen more easily that way for many students. If a student shortcuts the process by moving several cubes at once, that is okay, as long as the graphic representations are correct and the explanation is clear. The table below provides one possible representation.

Step Described in Words	Fair Share Cubes Representation	Dot Plot
Share two of the cubes in the 7-cube stack with the 2-cube stack. The result would be 5, 4, 5, 1, 5. The 7-cube stack went from 7 successful tosses to 5 successful tosses, and the 2-cube stack went from 2 successful tosses to 4 successful tosses.		
Then, two of the students who have 5 successful tosses share 1 toss with the student who had 1 successful toss. Those two students with 5 successful tosses went down 1 toss each to 4 successful tosses, and the student with 1 successful toss went up 2 tosses to 3 successful tosses. The result would be 4, 4, 4, 3, 5.		
Finally, the last student with 5 successful tosses shares one of them with the student who has 3 successful tosses. The final step of the fair share method shows an even number of tosses for each of the five students. So, the mean number of successful tosses for these five students is 4 tosses.		

I can refer back to the original data set to help me understand the moves in the first step.

I can act this out with actual cubes, coins, paper clips, or anything I find around the house that can help me visualize these steps.

3. The numbers of granola bars six students brought to school today are $1, 2, 3, 1, 4, 4$. Casey produces the following cubes representation as she does the fair share process. Help her decide how to finish the process now that she has stacks of $2, 2, 2, 2, 3, 4$.



To get to this step, I can share one cube from the 3-cube stack with a 1-cube stack. I can also share one cube from one of the 4-cube stacks with the other 1-cube stack.

There are three extra cubes within the stacks of three and four cubes. Since there are six stacks, each extra cube will need to be split in half so that there are six halves. Each of the six stacks will then have a total of two and one half cubes. In the context of the problem, each student will have a fair share mean of two and one half granola bars.

4. Suppose that the mean number of blueberries in 15 blueberry pancakes is 8 blueberries.

   a. Interpret the mean number of blueberries in terms of fair share.

   Answers will vary. If each of the 15 blueberry pancakes were to have the same number of blueberries, each would have 8 blueberries.

   b. Describe the dot plot representation of the fair share mean of 8 blueberries in 15 pancakes.

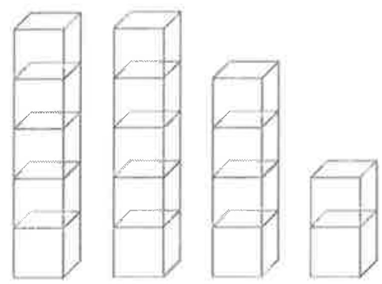
   Answers will vary. There should be 15 dots on the dot plot, all of them stacked up at 8.

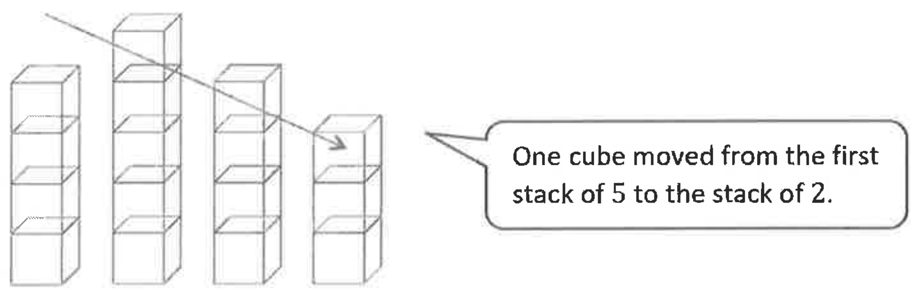
EUREKA MATH™

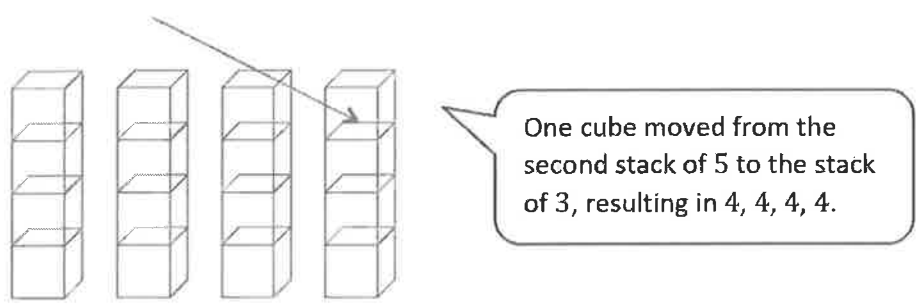
G6-M6-Lesson 7: The Mean as a Balance Point

1. The number of pencils brought to school today by four students is 5, 5, 4, and 2.

   a. Perform the fair share process to find the mean number of pencils for these four students. Sketch the cubes representations for each step of the process.







Each of the students with 5 pencils gives one pencil to the student who started with 2 pencils, yielding four students with four pencils each. Moving the cubes should result in 4 cubes in each of the four stacks. The mean is 4 pencils.

b. Find the total of the distances on each side of the mean to show the mean found in part (a) is correct.

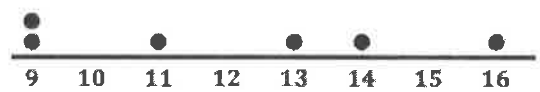
The mean is correct because the total of the distances to the left of 4 is 2, and the total of the distances to the right of 4 is 2 because $1 + 1 = 2$.

The mean represents the balance point of the data set. It is the point that balances the total of the distances to the left of the mean with the total of the distances to the right of the mean.

2 is 2 units away from 4 (the mean). 5 is 1 unit away from 4. Since there are two 5's in the data set, $1 + 1 = 2$, and the total of the distances to the right and left of 4 are both 2.

2. The times (rounded to the nearest minute) it took each of six classmates to run one mile are 9, 9, 11, 13, 14, and 16 minutes.

   a. Draw a dot plot representation for the mile times.



1-Mile Run Times (minutes)

   b. Suppose that Henry thinks the mean is 13 minutes. Is he correct? Explain your answer.

   ***Henry is incorrect.** The total of the distances to the right of 13 is 4 because $1 + 3 = 4$, and the total of the distances to the left of 13 is 10 because $2 + 4 + 4 = 10$. The totals of the distances are not equal; therefore, the mean cannot be 13 minutes.*

11 is 2 units to the left of 13, and 9 is 4 units to the left of 13. There are two 9's in the data set, so $2 + 4 + 4 = 10$.

14 is 1 unit to the right of 13, and 16 is 3 units to the right of 13, so $1 + 3 = 4$.

   c. What is the mean?

   For the total of the distances to be equal on either side of the mean, the mean must be 12 because the total of the distances to the left of 12 is 7 because $1 + 3 + 3 = 7$, and the total of the distances to the right of 12 is 7 because $1 + 2 + 4 = 7$.

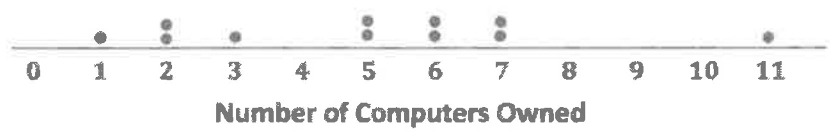
3. The number of computers (laptop and desktop) owned by the members of each of eleven families is $1, 2, 2, 3, 5, 5, 6, 6, 7, 7, 11$.

   a. Use the mathematical formula for the mean (determine the sum of the data points, and divide by the number of data points) to find the mean number of computers owned for these eleven families.

   $$\frac{55}{11} = 5$$

   The mean is* 5 *computers.

   b. Draw a dot plot of the data, and verify your answer in part (a) by using the balancing process.

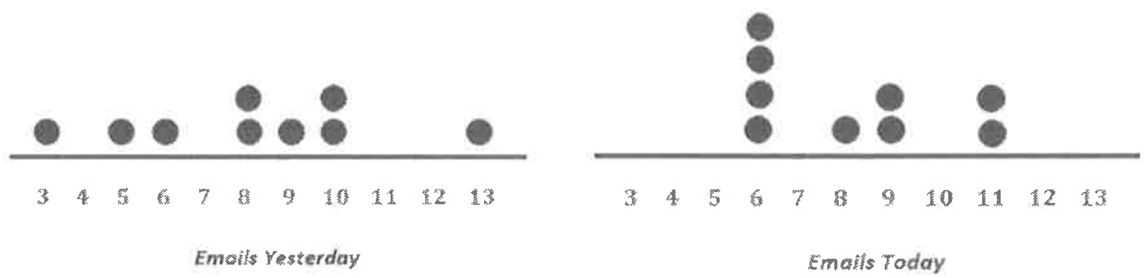


   The total of the distances to the left of 5 *is* 12 *because* $2 + 3 + 3 + 4 = 12$. *The total of the distances to the right of* 5 *is* 12 *because* $1 + 1 + 2 + 2 + 6 = 12$. *Since both totals are equal,* 5 *is the correct mean.*

G6-M6-Lesson 8: Variability in the Data Distribution

1. The number of emails nine employees received in one hour yesterday was 5, 3, 10, 13, 8, 8, 6, 10, and 9. The number of emails received by the same nine employees during the same time period today was 6, 8, 6, 9, 6, 6, 11, 11, and 9.

   a. Draw dot plots for the distributions of the number of emails received yesterday and of the number of emails received today. Be sure to use the same scale on both dot plots.



> I need to have the same scale in both dot plots. So I count from 3 to 13 in both dot plots so that all the points can be plotted.

   b. Do the distributions have the same mean? What is the mean of each dot plot?

   Yes, both distributions have a mean of 8 emails.

> I can use the fair share method that I learned in Lesson 6 to determine the means. Or, I can use the method from Lesson 7, using distances from the center.

EUREKA MATH

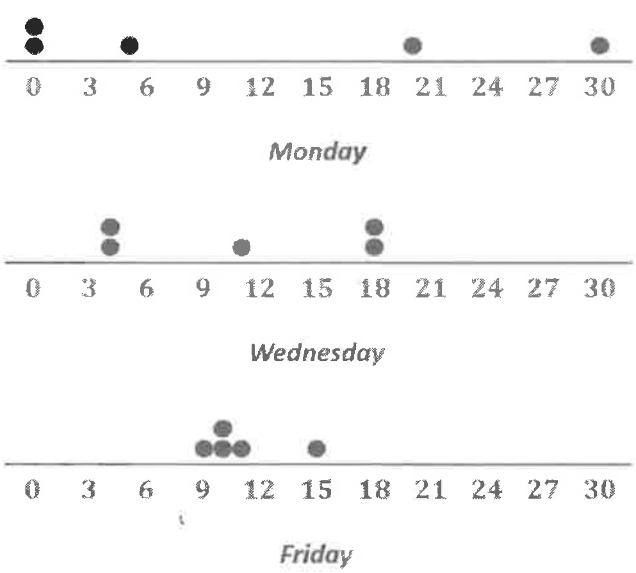
2. The following table shows the wait times, in minutes, at five restaurants at 5:00 p.m. across town as recorded on Monday, Wednesday, and Friday of a certain week.

Day	Marla's Diner	Taco, Taco	Tony's Italian Eatery	The Steak House	China Buffet
Monday	5	0	20	30	0
Wednesday	4	4	11	18	18
Friday	11	10	10	15	9

a. The mean wait time per day for the five restaurants is the same for each of the three days. Without doing any calculations and simply looking at Wednesday's wait times, what must the mean wait time be?

Wednesday's times are centered at 11 minutes. The sum of the distances from 11 for values above 11 is equal to the sum of the distances from 11 for values below 11, so the mean is 11 minutes.

b. For which daily distribution is the mean a better indicator of the typical wait time for the five restaurants? Explain.

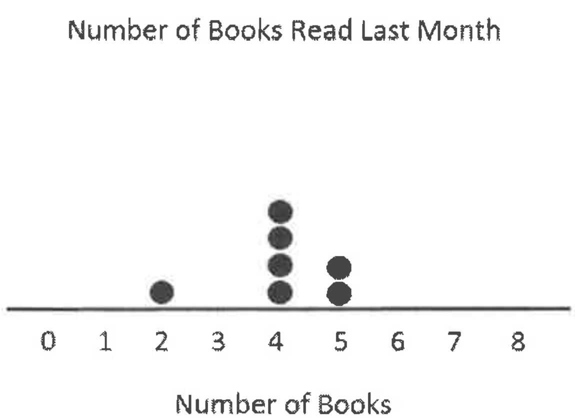


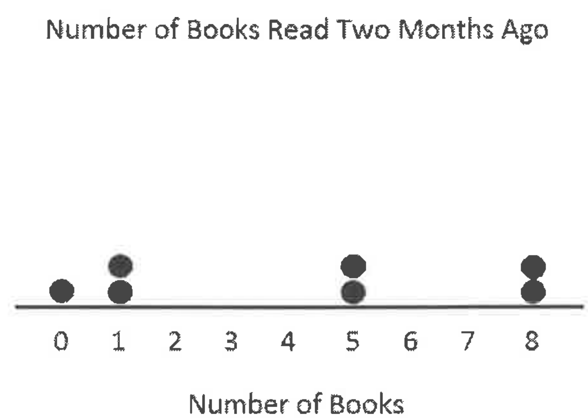
I need to compare the variability in each set to determine if the mean is a good indicator, so I can draw or visualize a dot plot for each to help me. The mean of a set with a smaller spread, where the points are all near the same amount, is considered a better indicator than a set of numbers that have a great amount of variability.

From the dot plots, the mean for Friday is the best indicator of the typical wait time for the five restaurants because there is the least variability in the Friday wait times.

G6-M6-Lesson 9: The Mean Absolute Deviation (MAD)

1. Suppose the dot plot on the left shows the number of books a group of friends read last month. The dot plot on the right shows the number of books the same group of friends read two months ago. The mean for both of these groups is 4.





I notice that on the left, the dots are all close to 4. If I were to calculate the distance from the mean, each one would be 2 or less. On the right, the dots are more spread out.

a. Before doing any calculations, which dot plot has the larger MAD? Explain how you know.

The graph showing the books read two months ago has a larger MAD because the data are more spread out and have the larger distances from the mean.

I know that if the points are more spread out on the dot plot, the values are more varied. MAD measures variability. So the farther the dots are on the number line, the larger the MAD.

b. Use the following tables to find the MAD for each distribution. Round your calculations to the nearest hundredth.

Last Month	
Number of Books	**Absolute Deviation**
2	2
4	0
4	0
4	0
4	0
5	1
5	1
Sum	4

Two Months Ago	
Number of Books	**Absolute Deviation**
0	4
1	3
1	3
5	1
5	1
8	4
8	4
Sum	20

I know that absolute deviation is the distance a value is from the mean. The mean in each case is 4, so I just need to know how far each of the points is from 4.

The MAD for last month is 0.57 *books because* $\frac{4}{7} \approx 0.57$. *The MAD for two months ago is* 2.86 *books because* $\frac{20}{7} \approx 2.86$.

To determine the MAD, I need to find the sum of all of the absolute deviations and then divide by 7 because there are 7 data values.

2. Consider the following data of the number of light bulbs that do not work in a case of light bulbs sampled from each of five companies. Note that the mean of each distribution is 14 broken light bulbs.

Company	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
A	24	10	8	12	6	10
B	18	3	9	20	17	3
C	2	36	14	5	5	8
D	16	16	8	12	8	10
E	5	14	14	2	22	13

a. Complete the following table of the absolute deviations for the six cases of light bulbs for each company.

	Absolute Deviation					
Company	**Case 1**	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**
A	10	4	6	2	8	4
B	4	11	5	6	3	11
C	12	22	0	9	9	6
D	2	2	6	2	6	4
E	9	0	0	12	8	1

Some of the absolute deviations have been determined for me. I just need to fill in the rest. I need to determine the distance from the mean to the values in the first table.

b. For which company is the mean a better indication of a typical number of broken light bulbs in each case? Explain your answer.

I know that for the mean to be a better indication of a typical value, the MAD must be small. I will need to calculate the MADs to support my answer.

	Absolute Deviation						SUM	MAD
Company	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6		
A	10	4	6	2	8	4	34	$\frac{34}{6} \approx 5.67$
B	4	11	5	6	3	11	40	$\frac{40}{6} \approx 6.67$
C	12	22	0	9	9	6	58	$\frac{58}{6} \approx 9.67$
D	2	2	6	2	6	4	22	$\frac{22}{6} \approx 3.67$
E	9	0	0	12	8	1	30	$\frac{30}{6} = 5$

I can extend the absolute deviation table by adding a column for the sum of the absolute deviations and a column for the MAD. This way, I can keep my work organized.

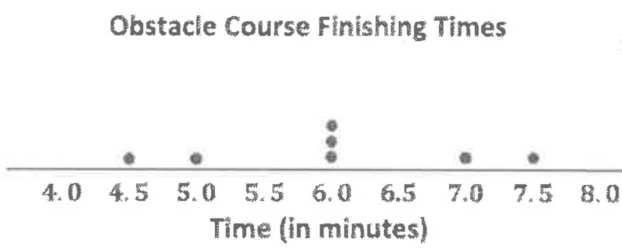
The mean is a good indicator for Company D because the MAD is smaller than for the other companies, showing that there is less variation in the data values.

G6-M6-Lesson 10: Describing Distributions Using the Mean and MAD

Seven people were timed while completing an obstacle course. Their finishing times are shown in the table.

Person	Adam	Bertram	Corrine	Diego	Enrique	Frieda	Gretchen
Time (in minutes)	6	7.5	5	6	7	6	4.5

a. Draw a dot plot for the finishing times for these seven participants.



I remember that I need to label my dot plot. I can use my dot plot to see if my data set is symmetrical or if the data clusters around one value.

b. Find the mean finishing time for the seven participants.

$$\frac{(6+7.5+5+6+7+6+4.5)}{7}=\frac{42}{7}=6$$

The mean of the data is 6 minutes.

c. Corrine said that the MAD for this data set is 0 minutes because the dot plot is balanced around 6. Without doing any calculations, do you agree with Corrine? Why or why not? If not, calculate the MAD, and state what it means.

No, Corrine is wrong. There is variability in the data. Not all points are on 6.

If all the participants finished at the same time, the MAD would be 0. If this were the case, my dot plot would show dots only at 6 and no other finishing times.

In Lesson 9, I learned that I must determine the sum of the absolute deviations and then divide by the number of data values to calculate the MAD.

The sum of the absolute deviations is 5. So, $\frac{5}{8} = 0.625$; therefore the MAD is 0.625 minutes. This means that, on average, the number of minutes these participants took to complete the obstacle course on a typical day differs by 0.625 minutes from the group mean of 6 minutes.

d. Suppose that in the original data set, Adam needs an additional two minutes to complete the course, and Frieda needs two less minutes to complete the obstacle course.

i. Without doing any calculations, does the mean for the new data set stay the same, increase, or decrease as compared to the original mean? Explain your reasoning.

The mean would remain at 6 minutes. One data value moved the same number of units to the right as another data value moved to the left, so, the balance point of the distribution does not change.

ii. Without doing any calculations, does the MAD for the new data set stay the same, increase, or decrease as compared to the original MAD? Explain your reasoning.

Since both scores moved away from the mean, the resulting distribution would be more spread out than the original distribution. Therefore, the MAD would increase.

If I think about the sum of the absolute deviations in part (c), I added in zeros for Adam, Diego, and Frieda because their times were the same as the mean. Now that Adam's and Frieda's time will change, I will have to add more to the sum; therefore, the MAD will increase too.

G6-M6-Lesson 11: Describing Distributions Using the Mean and MAD

1. Two bowling teams wrote down their scores for the last game. Summary measures for the two teams are as follows:

	Mean	MAD
Team A	184	15
Team B	184	4

I know that a larger MAD shows more variability. That means that the scores will possibly range from much lower to much higher than the mean for Team A. The small MAD for Team B means that the scores are clustered closer together.

a. Suppose that Rita, the newest bowler, scored the lowest score on the team. Would she have scored lower on Team A or Team B? Explain your reasoning.

Rita's score would have been lower if she had been on Team A because the means are the same, and the variability, as measured by the MAD, is higher on that team than it is for Team B. This tells me that the lowest score for Team A will be lower than the lowest score for Team B.

b. Suppose that your score was below the mean score. On which team would you prefer to have been? Explain your reasoning.

Because the MAD is larger for Team A, a score below the mean score could be a lot lower than the mean. Because the MAD is smaller for Team B, a score below the mean score would be much closer to the mean.

I would prefer to have been on Team B because then my score would most likely be closer to the mean of 184 *than if I were on Team A. A score below the mean on Team A could be far lower than on Team B because the MAD of Team A is much larger.*

2. A beekeeper keeps 8 hives in each of two of his apiaries, or bee yards. The numbers of pounds of honey produced by each hive is shown:

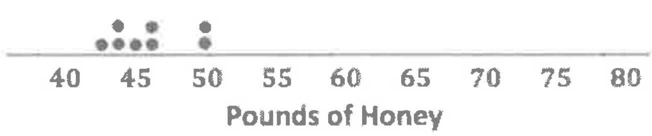
	Hive	1	2	3	4	5	6	7	8
Pounds of Honey	**Yard A**	40	42	57	45	52	73	45	78
	Yard B	43	50	45	46	44	46	44	50

a. Draw dot plots to help you decide which bee yard is more productive. Use the same scale for both of your dot plots (one that covers the span of both distributions).

Hive Productivity in Yard A

I need to determine the smallest and largest numbers in the table to determine the range of numbers to use in my dot plots.

Hive Productivity in Yard B



b. Calculate the mean number of pounds of honey produced for each bee yard. Which one produces more pounds of honey on average?

The mean number of pounds for Yard A is 54 *because* $\frac{40+42+57+45+52+73+45+78}{8} = 54$*, and the mean number of pounds for Yard B is* 46 *because* $\frac{43+50+45+46+44+46+44+50}{8} = 46$*. Yard A produces more pounds of honey on average.*

I can use the values in the table at the beginning of the problem to determine the mean.

c. If you want to be able to accurately predict the number of pounds of honey a bee yard will produce, which yard should you choose—the one with the smaller MAD or the one with the larger MAD? Explain your reasoning.

Yard B produces fewer pounds of honey on average but is far more consistent. Looking at the dot plots, its variability is far less than that of Yard A. Based on these data sets, choosing Yard B should yield numbers in the mid 40's consistently, but the numbers from Yard A could vary wildly from the low 40's to huge yields around 80.

d. Calculate the MAD of each bee yard.

	Hive	1	2	3	4	5	6	7	8
Absolute Deviations (in pounds)	**Yard A**	14	12	3	9	2	19	9	24
	Yard B	3	4	1	0	2	0	2	4

> I can make a table of the absolute deviations. This will help me keep my data organized. I also need to remember that the means are different. In Yard A, the mean is 54 pounds, and in Yard B, the mean is 46 pounds.

The sum of the distances from the mean for Yard A is 92 because $14 + 12 + 3 + 9 + 2 + 19 + 9 + 24 = 92$. Therefore, the MAD for Yard A is 11.5 pounds because $\frac{92}{8} = 11.5$.

The sum of the distances from the mean for Yard B is 16 because $3 + 4 + 1 + 0 + 2 + 0 + 2 + 4 = 16$. Therefore, the MAD for Yard B is 2 pounds because $\frac{16}{8} = 2$.

EUREKA MATH™

G6-M6-Lesson 12: Describing the Center of a Distribution Using the Median

1. Make up a data set such that the following is true:

   a. The data set has 15 different values, and the median is 9.

   Possible Set: 0, 1, 2, 3, 4, 5, 8, 9, 10, 13, 14, 15, 18, 19, 20

   I know that because this question asks for 15 different values, none of the numbers can repeat. Since 9 is the median, there should be 7 values that are lower than 9 and 7 values greater than 9, leaving 9 in the middle when ordered from least to greatest.

   b. The data set has 8 values, and the median is 32.

   Possible Set: 10, 14, 18, 30, 34, 35, 38, 40

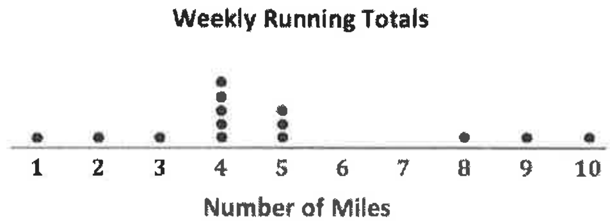
   In this question, I can repeat numbers. There must be an even number of values, so that means that two of the numbers will be in the middle. I could place two 32's in the middle. Or, I could choose two numbers with a mean of 32.

   c. The data set has 5 values, and the median is the same as the greatest value.

   Possible Set: 3, 5, 11, 11, 11

   When I find the median, I start by putting the numbers in order from least to greatest. If they are in correct order and the median and the greatest are the same, then any numbers between the two will also have to be the same.

2. The dot plot shows the number of miles each person in a random sample ran last week.



a. How many people were in the sample?

There are 14 people in the sample.

> Each one of the dots represents a response from a person, so I can just count the dots to see how many people were in the sample.

b. Find the median number of miles run in the sample.

$1, 2, 3, 4, 4, 4, 4, 4, 5, 5, 5, 8, 9, 10$

The median number of miles run this week is 4.

> If I have trouble using the dot plot to determine the median, I could write out all the numbers in order.

c. Do you think the mean or median would be a better description of the typical number of miles run? Explain your thinking.

The mean is approximately 4.86 miles per week, while the median is 4 miles per week. The mean is slightly higher than the median and has a value that is greater than most of the dots on the dot plot. Therefore, the median is a closer representation.

> Calculating the mean is a necessary step in answering this question.

3. The salaries of eleven employees at a local business are given below.

Employee	Salary
President	$\$320{,}000$
Vice President	$\$232{,}000$
Manager	$\$94{,}000$
Employee A	$\$64{,}000$
Employee B	$\$64{,}000$
Employee C	$\$58{,}000$
Employee D	$\$51{,}000$
Employee E	$\$50{,}000$
Employee F	$\$48{,}000$
Employee G	$\$48{,}000$
Employee H	$\$47{,}000$

I notice that these numbers are already in order, so I just need to determine which number is in the middle to determine the median.

a. Find the median salary, and explain what it tells you about the salaries.

The median salary is $\$58{,}000$ *for Employee C. Half of the employees make more than* $\$58{,}000$*, and half of the employees make less than* $\$58{,}000$*.*

b. Find the median of the lower half of the salaries and the median of the upper half of the salaries.

This question is really just asking me to determine two other medians. Instead of using all of the numbers, I will focus on the top half of the numbers and the lower half of the numbers.

$\$48{,}000$ *is the median for the bottom half of the salaries.* $\$94{,}000$ *is the median for the top half of the salaries.*

c. Find the width of each of the following intervals. What do you notice about the size of the interval widths, and what does that tell you about the salaries?

> I can use my answer to parts (a) and (b) to answer these questions. To get the size of the interval, I just need to find the difference between the values given.

i. Minimum salary to the median of the lower half: $\$1,000$

ii. Median of the lower half to the median of the whole data set: $\$10,000$

iii. Median of the whole data set to the median of the upper half: $\$36,000$

iv. Median of the upper half to the highest salary: $\$226,000$

The largest width is from the median of the upper half to the highest salary. The smaller salaries are closer together than the larger ones.

G6-M6-Lesson 13: Describing Variability Using the Interquartile Range (IQR)

1. In each of parts (a)–(c), create a data set with at least 6 values such that it has the following properties:

   a. An IQR equal to 12.

   One example is $\{3, 8, 11, 13, 14, 20, 23\}$ *where the IQR is* 12 *because* $20 - 8 = 12$.

   For the IQR to be equal to 12, $Q3 - Q1$ must also be equal to 12. So Q3 could be 20, and Q1 could be 8, making a difference of 12.

   b. An IQR equal to 0.

   One example is $\{10, 18, 18, 18, 18, 18, 24\}$.

   For the IQR to be equal to 0, Q3 and Q1 must be the same. If the median of the lower half of the data is the same as the median of the top half of the data, all the values between the two must also be the same.

2. A sample of the heights of students in two classes are given, in inches, in the table below.

Mrs. M's Class	44	38	47	46	39	42	40	46	35	46
Mr. V's Class	52	58	42	38	45	40	62	56	45	49

   a. How do you think the data might have been collected?

   Someone at the school may have measured the students' heights. The measuring may have been done by the teacher or the school nurse.

b. Do you think it would be possible for $\frac{1}{4}$ of the heights for Mr. V's class to be 50 inches or above? Why or why not?

Yes, it might be possible. The mean height in Mr. V's class is 48.7 inches, and more than 25% of the sample values are 50 inches or above.

c. Make a prediction about how the values of the IQR for the heights for each class compare. Explain your thinking.

Mr. V's class probably has the larger IQR because those heights seem to vary more than the heights for Mrs. M's class.

I know that the IQR measures variability. I can look at the data and see which set varies more in order to make my prediction.

d. Find the IQR for the heights for each class. How do the results compare to what you predicted?

Mrs. M's class: $35, 38, 39, 40, 42, 44, 46, 46, 46, 47$

$Q1 = 39$, *Median* $= 43$, $Q3 = 46$

Mr. V's class: $38, 40, 42, 45, 45, 49, 52, 56, 58, 62$

$Q1 = 42$, *Median* $= 47$, $Q3 = 56$

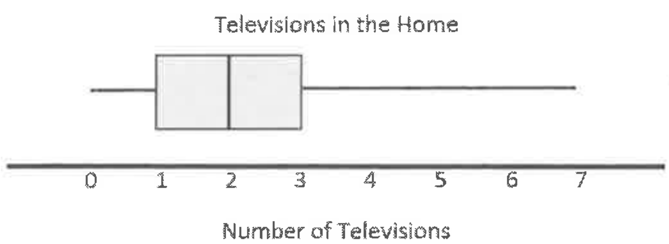
In order to calculate the IQR for each class, I must first place the heights in order from least to greatest. Then, I can determine the key points, like the median, Q1, and Q3.

The IQR for Mrs. M's class is 7 inches because $46 - 39 = 7$. For Mr. V's class, the IQR is 14 inches because $56 - 42 = 14$. This result matches my prediction in part (c).

EUREKA MATH™

G6-M6-Lesson 14: Summarizing a Distribution Using a Box Plot

1. The box plot below summarizes data from a survey of households about the number of televisions they have. Identify each of the following statements as true or false. Explain your reasoning in each case.



I know that a box plot is made using 5 key points: the lowest number, the median of the lower numbers, the median of the entire set, the median of the higher numbers, and the greatest number. The lines and box show me these five values from left to right. This makes up the five- number summary.

a. The maximum number of televisions per house is 3.

False. The line segment at the top reaches 7.

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

True. 2 is the median. This tells us that half the houses have 2 or more televisions. If we use the lower median, we can see that three-fourths of the houses have at least one television, which is more than one-half.

c. All of the houses have televisions.

False. The lower line segment starts at 0, so at least one household does not have a television.

d. Half of the houses surveyed have between 0 and 2 televisions.

True. About 25% of the houses would have between 0 and 1 television, and another 25% of the houses would have between 1 and 2 televisions, making a total of 50%, or half, of the houses.

2. The number of minutes it takes a group of runners to complete a 5K race are as follows:

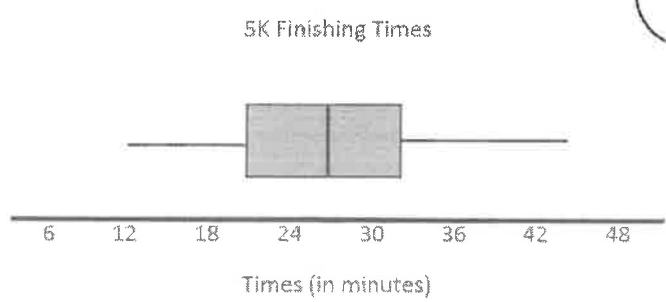
$12, 18, 24, 30, 45, 22, 18, 42, 32, 38, 28, 28, 28, 24, 25, 16, 39, 21$

a. Make a box plot of the finishing times.

$12, 16, 18, 18, 21, 22, 24, 24, 25, 28, 28, 28, 30, 32, 38, 39, 42, 45$

These finishing times are not in order. So the first step is to order the numbers from least to greatest. Then I will be able to determine the medians.

Five-number summary: $12, 21, 26.5, 32, 45$



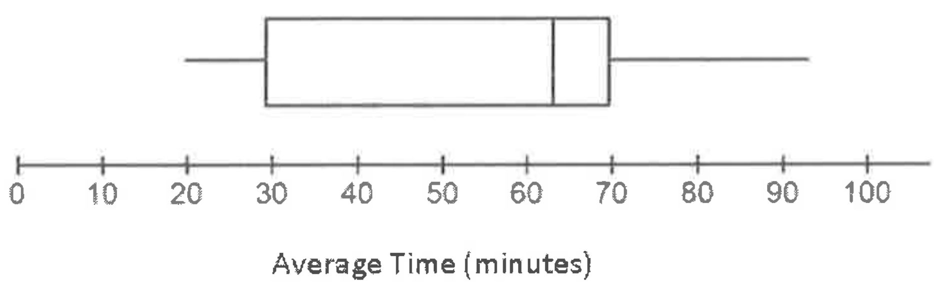
b. Describe the finishing times distribution. Include a description of center and spread.

The IQR is 32 minutes $- 21$ minutes, or 11 minutes. Half of the finishing times were near the middle between 21 minutes and 32 minutes. The median is 26.5 minutes. A quarter of the finishing times are less than 21 minutes but greater than or equal to 12 minutes. A quarter of the finishing times are greater than 32 minutes and less than or equal to 45 minutes.

I know that my box plot has four sections and that each section represents a quarter of the data values. I can use this to help me describe the distribution.

G6-M6-Lesson 15: More Practice with Box Plots

1. The box plot below summarizes the average time students in Ms. Baker's math class spend on homework every night.



a. Estimate the values in the five-number summary from the box plot.

Answers may vary. Minimum $= 20$ min; $Q1 = 30$ min; *Median* $= 63$ min; $Q3 = 70$ min; *Maximum* $= 93$ min

b. The highest average time is 93 minutes, followed by the second highest average time of 77 minutes. What does this tell you about the spread of the average times spent on math homework every night in the top quarter of the box plot?

The Q3 is about 70 minutes, so all but one of the scores in the top quarter are between 70 minutes and 77 minutes.

If approximately $\frac{1}{4}$ of the data values are between Q3 (70) and the maximum (93) and the second highest average is 77, then I know there are no more values between 77 and 93.

c. Use the five-number summary and the IQR to describe the average amount of time students spend on math homework every night.

> $IQR = Q3 - Q1$

The average times vary from 20 minutes to 93 minutes. The IQR is 40 minutes; the middle half of the average times are between 30 minutes and 70 minutes. Half of the times students spent on homework are less than 63 minutes.

> The middle half of the data is between Q1 and Q3.

> The median is 63 and about one half of the data is above/below the median.

2. Suppose the interquartile range for the number of hours students spent playing outside during the summer was 10. What do you think about each of the following statements? Explain your reasoning.

a. About half of the students played outside for 10 hours during the summer.

This may not be correct as you know the width of the interval that contains the middle half of the times was 10, but you do not know where it starts or stops. You do not know the lower or upper quartile.

> I remember the interquartile range describes how spread out the middle 50% of the data are in the data distribution.

b. All of the students played at least 10 hours outside during the summer.

This may not be correct for the same reason as in part (a).

c. About half of the class could have played outside from 10 to 20 hours a week or from 15 to 25 hours.

Either could be correct as the only information given is the interquartile range of 10, and the statement says "could have."

> About half of the class could have also played from 9 to 19 minutes or from 12 to 22 minutes. There are many possibilities.

3. Suppose you know the following for a data set: The minimum value is 65, the lower quartile is 77, the IQR is 21, half of the data are less than 85, and the maximum value is 105.

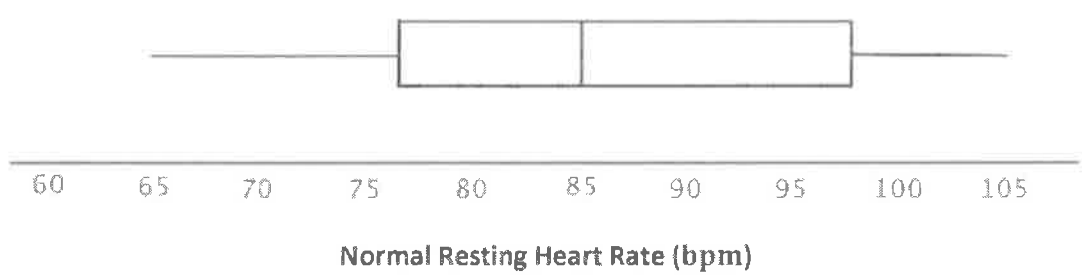
I need to determine the upper quartile to construct the box plot. I know the IQR is calculated by finding the difference of the upper quartile and the lower quartile. Since the IQR and the lower quartile are given, I can write an equation to find the value for the upper quartile.

$$\text{Upper Quartile} - \text{Lower Quartile} = \text{Interquartile Range}$$
$$\text{Upper Quartile} = \text{Interquartile Range} + \text{Lower Quartile}$$
$$\text{Upper Quartile} = 21 + 77$$
$$\text{Upper Quartile} = 98$$

a. Think of a context for which these numbers might make sense.

Answers will vary. One possibility is a healthy person's normal resting heart rate in beats per minute (bpm) since the resting heart rate for a healthy person is between 60 bpm and 100 bpm.

b. Sketch a box plot related to the context in part (a).



c. Are there more data values above or below this median? Explain your reasoning.

The number of data values on either side of the median should be about the same, one half of all of the data.

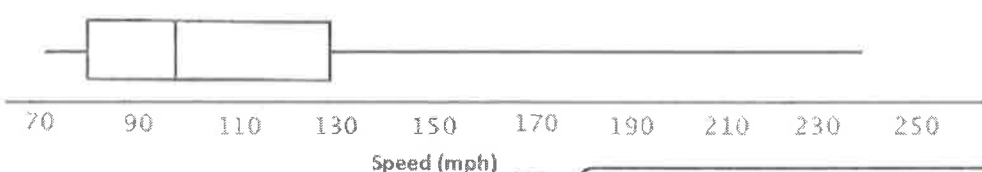
4. The speeds of the fastest birds are given in the table below.

Type of Bird	Speed (mph)
Peregrine Falcon	242
Golden Eagle	199
Gyrfalcon	130
Common Swift	106
Eurasian Hobby	100
Frigatebird	95
Spur Winged Goose	88
Homing Pigeon	87
Red-breasted Merganser	81
White-rumped Swift	77
Canvasback	73

Data Source: http://dinoanimals.com/animals/the-fastest-birds-in-the-world-top-10/, accessed September 28, 2015

a. Find the five-number summary for this data set, and use it to create a box plot of the speeds.

Minimum $= 73$, $Q1 = 81$, *Median* $= 95$, $Q3 = 130$, *Maximum* $= 242$



I remember making box plots in Lesson 14.

b. Why is the median not in the center of the box?

The median is not in the center of the box because about $\frac{1}{4}$ *of the speeds are between* 95 mph *and* 130 mph, *and another* $\frac{1}{4}$ *are closer together, between* 81 mph *and* 95 mph.

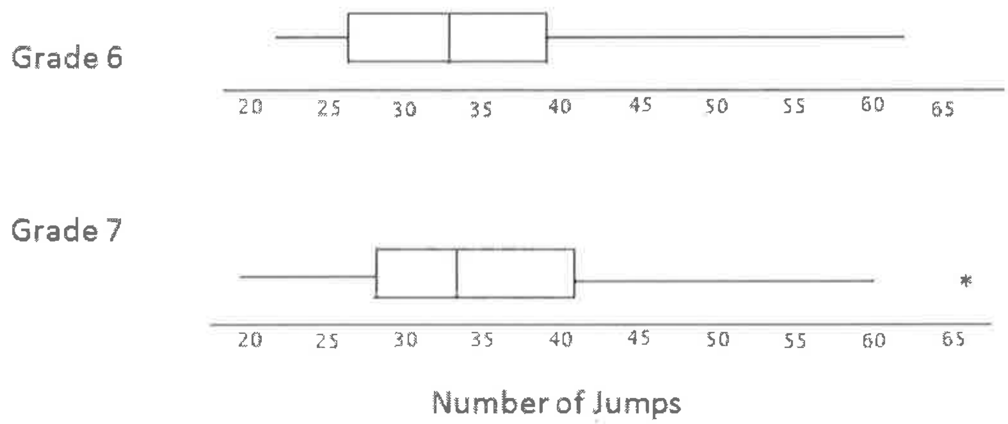
I remember approximately $\frac{1}{4}$ of the data values are found in each section of a box plot.

c. Write a few sentences telling your friend about the speeds of the fastest birds.

Half of the birds fly faster than 95 mph; *the fastest bird in the list is the Peregrine Falcon with a speed of* 242 mph. *The slowest bird in the list is the Canvasback with a speed of* 73 mph. *The middle* 50% *of the speeds are between* 81 mph *and* 130 mph.

G6-M6-Lesson 16: Understanding Box Plots

The results of a jump rope competition between sixth and seventh graders are summarized below. Students recorded how many times they could jump rope in one minute.



a. In which grade did the students do the best? Explain how you can tell.

Students were equally successful in both grades. For both grades, the median, the lower quartile and the upper quartile are about the same although these values for seventh grade are slightly shifted to the right.

I notice that the distribution of the data values in each set of data is very similar.

b. Why do you think one of the data values in Grade 7 is not part of the line segment?

The highest number of jumps was pretty far away from the other number of jumps, so it was marked separately.

c. How do the median number of jumps for the two grades compare? Is this surprising? Why or why not?

The median number of jumps in Grade 7 was about the same, but slightly higher, than the median number of jumps in Grade 6. This makes sense because the sixth and seventh graders should be able to jump approximately the same number of times in one minute. I wouldn't expect the seventh graders to jump many more times in one minute than the sixth graders, so these results did not surprise me.

d. How do the IQRs compare for the two grades?

The middle half of the Grade 7 number of jumps was fairly spread out spanning about 13 jumps from about 28 to 41 jumps with the median around 34 jumps. The middle half of the Grade 6 number of jumps was also fairly spread out spanning about 13 jumps from about 26 to 39 jumps.

e. The sixth grader with the most number of jumps was Max, and the seventh grader with the most number of jumps was Makayla. How many jumps did they do in one minute? How can you tell?

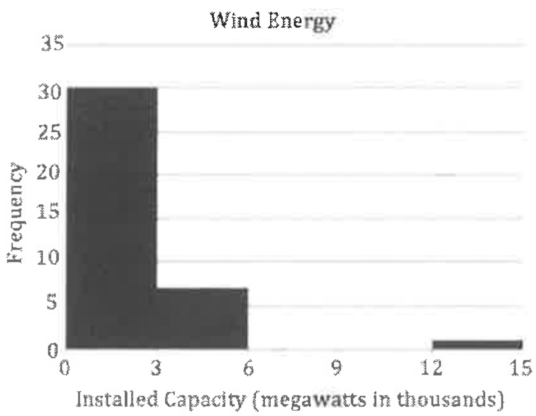
Max jumped about 63 times, and Makayla jumped about 66 times. I know this because I looked at the maximum value for each box plot.

f. Clara, a sixth grader, jumped 47 times in one minute. What can you say about the percent of sixth graders who jumped more times than Clara in one minute?

Clara is in the upper quartile with 47 jumps, so fewer than 25% of the sixth graders were able to jump more times than Clara.

G6-M6-Lesson 18: Connecting Graphical Representations and Numerical Summaries

1. The following histogram shows the amount of wind produced (by state) for the 40 states that had wind facilities by the end of 2014. Many of these states produced less than 3,000 megawatts of wind, but one state produced over 14,000 megawatts (Texas). For the histogram, which *one* of the three sets of summary measures could match the graph? For each choice that you eliminate, give at least one reason for eliminating the choice.



Because a histogram does not show individual values, it is not possible to determine exact values for the 5-number summary. However, I can use my knowledge of what these terms mean to see what makes sense in the context of the situation and the information provided on the histogram.

Data source: http://www.neo.ne.gov/statshtml/205.htm, accessed September 10, 2015

a. Minimum $= 1$, $Q1 = 0.16$, Median $= 0.73$, $Q3 = 2.6$, Maximum $= 13$, Mean $= 1.6$, MAD $= 1.5$

b. Minimum $= 1.5$, $Q1 = 16.6$, Median $= 0.73$, $Q3 = 2.6$, Maximum $= 14$, Mean $= 15.2$, MAD $= 1.67$

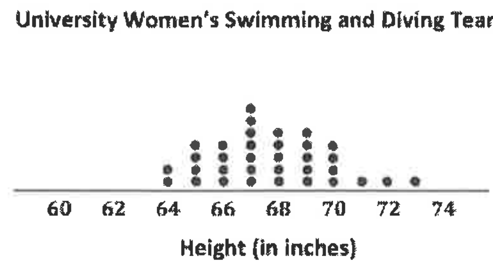
c. Minimum $= 2.7$, $Q1 = 0.16$, Median $= 730$, $Q3 = 6$, Maximum $= 17$, Mean $= 1.6$, MAD $= 1.5$

The correct answer is (a).

Choice (b) would not work because Q1 (median of the lower half of the data) must be less than 0.73*, the median, so a value of* 16.6 *is unreasonable. Also, the mean cannot be larger than the maximum value listed in the graph, so the value of the mean in choice (b) is not reasonable.*

Choice (c) would not work. The value of the median is unreasonable given the scale of the graph. Also, the mean is most likely greater than (not less than) the median given the skewed right nature of the distribution and the large outlier, which is not the case in choice (c). The maximum value is also unreasonable since it is larger than the largest number on the horizontal scale.

2. The heights (rounded to the nearest inch) of the 34 members of the 2015–2016 Brigham Young University Women's Swimming and Diving Team are shown in the dot plot below.



Data source: http://byucougars.com/roster/w-swimming-diving, accessed September 28, 2015.

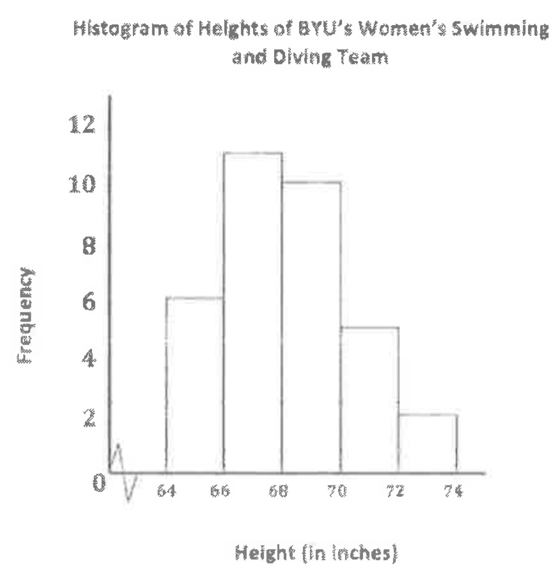
a. Use the dot plot to determine the 5-number summary (minimum, lower quartile, median, upper quartile, and maximum) for the data set.

Min $= 64$, $Q1 = 66$, *Median* $= 67.5$, $Q3 = 69$, *and Max* $= 73$

> The lower quartile, Q1, is the median of the bottom half of the data.

> The upper quartile, Q3, is the median of the top half of the data.

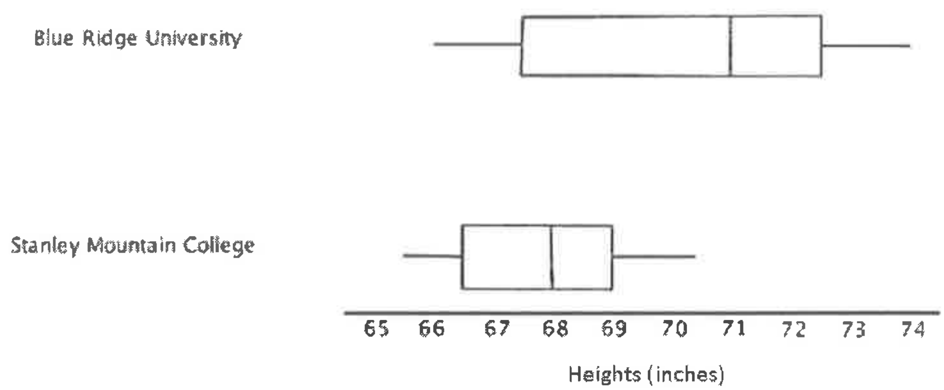
b. Based on this dot plot, make a histogram of the heights using the following intervals: 64 to $<$ 66 inches, 66 to $<$ 68 inches, and so on.



EUREKA MATH™

G6-M6-Lesson 19: Comparing Data Distributions

College athletic programs are separated into divisions based on school size, available athletic scholarships, and other factors. A researcher wondered if members of Division I women's basketball programs (usually large schools that offer athletic scholarships) tend to be taller than members of Division III women's basketball programs (usually smaller schools that do not offer athletic scholarships). To begin the investigation, the researcher creates side-by-side box plots for the heights (in inches) of 35 female basketball players at Blue Ridge University (a Division I program) and the heights (in inches) of 15 female basketball players at Stanley Mountain College (a Division III program).



a. Which data set has the smaller range?

Stanley Mountain College has the smaller range.

From the minimum value to the maximum value, the range is smaller for Stanley Mountain College than Blue Ridge University.

b. True or false: A basketball player who had a height equal to the median for the Blue Ridge University would be taller than the median height of basketball players at Stanley Mountain College.

True

The median height for a player from Blue Ridge University is 71 inches, and the median height for a player from Stanley Mountain College is 68 inches.

c. To be thorough, the researcher will examine many other colleges' sports programs to further investigate the claim that members of Division I women's basketball programs are generally taller than the members of Division III women's basketball programs. But given the graph above, in this initial stage of her research, do you think that the claim might be valid? Carefully support your answer using summary measures or graphical attributes.

Based on just these two teams, it looks like the claim may be correct. A large portion of the Blue Ridge University distribution is higher than the maximum value of the Stanley Mountain College distribution. The median value for the Blue Ridge University distribution appears to be 3 inches higher than the median value of the Stanley Mountain College distribution.

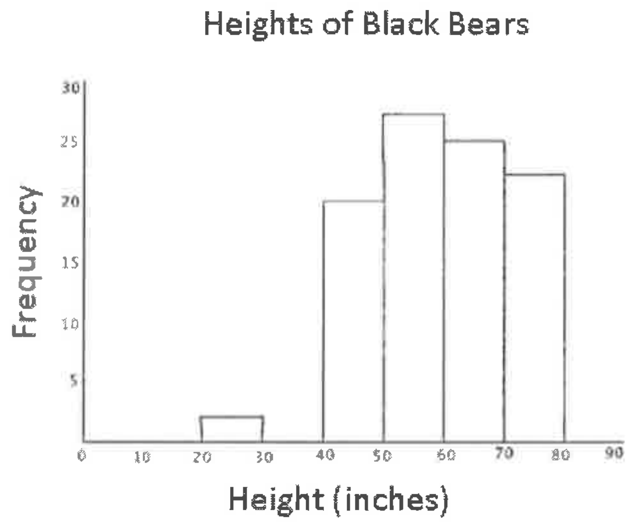
d. True or False: At least one of the basketball players from Blue Ridge University is taller than the tallest basketball player from Stanley Mountain College.

True. The median height for the Blue Ridge University basketball players is greater than the maximum height for the basketball players at Stanley Mountain College, so about 50% of the basketball players from Blue Ridge University are taller than the tallest player from Stanley Mountain College.

EUREKA MATH™

G6-M6-Lesson 20: Describing Center, Variability, and Shape of a Data Distribution from a Graphical Representation

Data was collected on the heights of black bears in a particular forest. A histogram of the heights for the black bear in this sample is shown below.



1. If the height of an average adult black bear is 45 to 75 inches, what can you conclude about this sample? Explain your answer.

   Answers may vary. One response might be: In this sample, there were most likely 2 to 3 baby black bears that were measured since 2 to 3 bears were between 20 and 30 inches tall and much shorter than the average height.

2. Does this histogram represent a data distribution that is skewed or that is nearly symmetrical?

   This distribution is skewed. The tail of this distribution is to the left, or toward the shorter heights.

> In a skewed shape, there are values that are unusual (or not typical) when compared to the rest of the data. In this histogram, there are values much lower than most of the data.

3. What measure of center would you use to describe a typical height of a black bear in this sample? Explain your answer.

   I recommend the median of the data distribution as a description of a typical value of the height of a black bear because this distribution is skewed.

4. Assume the smallest black bear measured was 24 inches tall, and the largest black bear measured was 78 inches tall. Estimate the values in the five-number summary for this sample:

   Minimum (min) $= 24$ *inches*

   $Q1 = 45$ *inches (a value greater than* 40 *but within the interval of* 40 *to* 50 *inches)*

   Median $= 60$ *inches (a value within the interval of* 50 *to* 70 *inches)*

   $Q3 = 70$ *inches (a value within the interval of* 70 *to* 80 *inches)*

   Maximum (max) $= 78$ *inches*

   Since histograms do not show specific values, it is hard to determine values for the five-number summary. The minimum and maximum are given so these are specific values but the values for Q1, median, and Q3 are reasonable estimates.

5. Based on the shape of this data distribution, do you think the mean height of a black bear from this sample would be greater than, less than, or the same as your estimate of the median? Explain your answer.

   An estimate of the mean would be less than the median height because the values in the tail, or to the left of the median, pull the mean in that direction.

   When there are uncharacteristically small or large values in a data set, the mean is sensitive to these values, and the mean may be pulled toward the very small or large values.

6. Estimate the mean value of this data distribution.

   An estimate of the mean would be a value slightly smaller than the median value. For example, a mean of 55 inches would be a reasonable estimate of a balance point.

   Since I estimated the median to be 60 inches and I know the mean will be slightly less because of the tail toward the shorter heights (the left), I can choose a value of 55 inches to represent the mean.

7. What is your estimate of a typical height of a black bear in this sample? Did you use the mean height or median height for this estimate? Explain.

   Since the median was selected as the appropriate estimate of a measure of center because this data distribution is skewed, a value of 60 inches (or whatever students used to estimate the median) would be an estimate of a typical height for a black bear from this sample.

   The measure of center (mean or median) represents a typical value for a data set.

8. Would you use the MAD or the IQR to describe variability in the height of black bears in this sample? Estimate the value of the measure of variability that you selected.

   I would use the IQR to describe the variability because the data distribution is skewed, and the median was used as a typical height for a black bear. An estimate of the IQR based on the above estimates would be as follows: 70 inches − 45 inches = 25 inches.

   $IQR = Q3 - Q1$