

Name _____

Period _____ Date _____

**UNIT 1
STUDENT PACKET**

MathLinks
GRADE 6



STATISTICS

		Monitor Your Progress	Page
	My Word Bank		0
1.0	Opening Problem: Beach Cleanup		1
1.1	Introduction to Data Analysis <ul style="list-style-type: none"> • Collect and organize numerical data. • Find measures of center and spread for a data set. • Understand how outliers affect measures of center. • Recognize and develop statistical questions. 	3 2 1 0 3 2 1 0 3 2 1 0 3 2 1 0	2
1.2	More Measures of Center and Spread <ul style="list-style-type: none"> • Find the mean and mean absolute deviation (MAD) for a data set. • Interpret the measures of center and spread for a data set. 	3 2 1 0 3 2 1 0	6
1.3	Data Displays <ul style="list-style-type: none"> • Construct dot plots (line plots), histograms, and box plots (box-and-whisker plots). • Describe the distribution of a data set by examining its center, spread, and overall shape. 	3 2 1 0 3 2 1 0	12
	Review		19
	Student Resources		27

Parent (or Guardian) signature _____

MY WORD BANK

Explain the mathematical meaning of each word or phrase, using pictures and examples when possible. See **Student Resources** for mathematical vocabulary.

five-number summary interquartile range (IQR)	mean
mean absolute deviation (MAD)	median
mode	outlier
range (of a data set)	statistical questions

BEACH CLEANUP

Students at Moffett Middle School in San Diego, CA are expected to commit to community service for 120 minutes, on average, per month.

Everette reads that the beaches are littered with cigarette butts, food wrappers, and plastic bottles, so he chooses to serve by cleaning local beaches.

Here is Everette’s beach cleanup log from the start of school through the start of the winter break.

Week #	Week beginning	Number of minutes spent doing cleanup	Week #	Week beginning	Number of minutes spent doing cleanup
1	Aug 17	20	10	Oct 19	10
2	Aug 24	40	11	Oct 26	20
3	Aug 31	30	12	Nov 2	60
4	Sept 7	0	13	Nov 9	50
5	Sep 14	45	14	Nov 16	40
6	Sep 21	55	15	Nov 23	30
7	Sep 28	50	16	Nov 30	40
8	Oct 5	20	17	Dec 7	20
9	Oct 12	30	18	Dec 14	0

Is Everette living up to his volunteering commitment?

INTRODUCTION TO DATA ANALYSIS

We will find our “name scores” using the given values for each letter. Then we will create a human number line to help us organize our data. We will calculate measures of center and spread for the name score data, and understand how outliers may affect measures of center. We will identify characteristics of a statistical question.

[6.SP.1, 6.SP.2, 6.SP.3, 6.SP.5ac, SMP2, 3, 4, 5, 6]

GETTING STARTED

Dr. Dana is a veterinarian. She records the sizes of dogs in their medical records when they come to her clinic.

1. What might Dr. Dana want to measure to describe the size of a dog?

What units of measure might she use?

Here are the weights (in pounds) of Great Danes (a breed of dog) that came to her veterinary clinic in one week. Complete the table below using this data.

Great Danes (Males)		
45	125	140
190	150	

Great Danes (Females)		
130	125	90
100	120	95

	Males	Females
2. How many Great Danes came to the clinic?		
3. Write the weights in order, from least to greatest.		
4. Record the lightest weight, in pounds.		
5. Record the heaviest weight, in pounds.		
6. What do you think is the typical weight? Explain your reasoning.		

NAME SCORES

Follow your teacher's directions.

(1) My first name: _____ My name score: _____

- (2)
- Number of observations (n) in the population: $n =$ _____
 - Make an organized list of the data.

- (3)
- Mode(s): _____
 - Median: _____

- (4)
- Minimum: _____ Maximum _____
 - Range of a data set: $\frac{\text{max}}{\text{max}} - \frac{\text{min}}{\text{min}} =$ _____

- Quartiles: $Q_1 =$ _____ $Q_2 =$ _____ $Q_3 =$ _____
- Five-number summary: $(\frac{\text{min}}{\text{min}}, \frac{Q_1}{Q_1}, \frac{Q_2}{Q_2} \text{ (median)}, \frac{Q_3}{Q_3}, \frac{\text{max}}{\text{max}})$
- Interquartile range: $\frac{Q_3}{Q_3} - \frac{Q_1}{Q_1} =$ _____

- (5)
- (6) Record the meanings of _____, _____, _____, _____, and _____ in **My Word Bank**.

PRACTICE 1

Bobbie likes to play card games with her friends. Listed below are the number of days she played cards each month last year. Let's explore: How often does Bobbie play cards?

Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
5	10	8	7	7	6	5	5	8	0	8	10

Focus on organizing the data.

1. For how many months was data collected? (number of observations) $n =$ _____
2. Rewrite the data in order from least to greatest in the table below.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Focus on measures of center for the data.

3. Write the mode(s) of the data set: _____. What does this tell us about Bobbie's card playing habits?
4. The median of the data set is _____. What does this tell us about her card playing habits?
5. Which of these two measures of center best represents her card playing habits? Explain.

Focus on measures of spread for the data.

6. Range:
7. Five-number summary (_____, _____, _____, _____, _____)
min Q_1 med. Q_3 max
8. Interquartile range (IQR):
9. What does interquartile range represent?
10. Would you say that Bobbie's card playing has a lot of variability? Justify your answer.
11. Use your data analysis to answer: How often does Bobbie play cards?

STATISTICAL QUESTIONS

Follow your teacher’s directions for (1) – (3).

(1)

(2)

These are ... Statistical questions		These aren’t... Statistical questions	I’m not sure

(3)

For each pair of questions, put a check next to the statistical question.

4. What is the average length of hair for 6th graders in your school?
How long is your hair?
5. How far is it from Los Angeles to San Francisco?
How long does it typically take people to drive from Los Angeles to San Francisco?
6. How many points did Giannis Antetokounmpo score last night?
How many points per game can we expect Giannis Antetokounmpo to score?
7. How many blue whales can we expect to see on a whale-watching tour?
How many blue whales did you see on your whale-watching tour?
8. How tall are you?
Are you taller or shorter than the average height of students your age?
9. Do 6th graders today watch more TV compared to 6th graders 10 years ago?
How much TV did you watch yesterday?

Let’s revisit our name score activity again.

10. What was a statistical question that we answered with the name score data?

11. Record the meaning of statistical questions in **My Word Bank**.

MORE MEASURES OF CENTER AND SPREAD

We will revisit our class name score data to calculate the mean score and determine if our data set has potential outliers. We will find the mean absolute deviation (MAD) for data sets. We will then interpret measures of center and spread for data sets.

[6.SP.1, 6.SP.2, 6.SP.3, 6.SP.5abcd; SMP1, 2, 3, 4, 5, 6, 7]

GETTING STARTED

BK likes to talk about sports. She says that two attributes needed to be a good basketball team are the ability to play good defense and to get rebounds.

1. How might you measure the ability to play good defense?
2. How might you measure a team's ability to get rebounds?

Two basketball teams' rebound totals for the last five games are posted below.

The Gremlins				
30	32	45	39	41

The Gladiators				
28	53	45	21	61

3. Put a check next to the better statistical question. Explain.
 - Which team had more rebounds in the fifth game?
 - Which team is better at rebounding?
4. Answer the statistical question from Problem 3. Explain your reasoning using measures of center and/or measures of spread.
5. Sandy thinks that the median number of rebounds for both data sets is 45 because that is the middle number for both. Do you agree with Sandy?

NAME SCORES REVISITED

Follow your teacher’s directions for (1) – (7).

(1)

(2)

(3)

(4)

A →

B →

C →

D →

E →

	Median	Mean	Mode
(5) Measures of Center based on original Name Scores.			
(6) New Name Scores Measures of Center			

(7)

8. Record the meanings of mean and outlier in **My Word Bank**.

PRACTICE 2

Recall in the opening problem that students at Moffett Middle School in San Diego, CA are expected to commit to community service for 120 minutes, on average, per month. Here are the number of volunteer minutes for two students over a 5-week period.

Aaron's number of Minutes				
35	35	35	35	0

John's number of minutes				
27	27	27	27	72

1. For each student, find the mean, median, and mode. Label clearly.

	Mean	Median	Mode
Aaron			
John			

2. Do either of these data sets appear to include potential outliers? Explain.

3. After 5 weeks, do these students appear to be meeting their volunteering commitments? Justify your decisions using the statistical measures you calculated and a written explanation.

4. Aaron volunteered 0 minutes during week 5 because he was sick. What is the lowest number of minutes Aaron could volunteer week 6 to have an average of 30 minutes per week?

Say he reaches that number. Is this 6-week average on pace to reach the goal?

5. Go back to the opening problem, **Beach Cleanup**, and make changes or additions as needed, based on what you've learned about statistical measures.

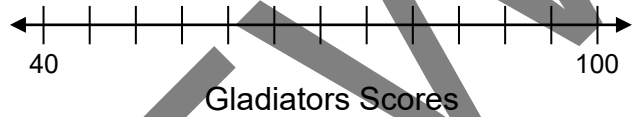
MEAN ABSOLUTE DEVIATION (MAD)

Recall the two basketball teams, the Gremlins and the Gladiators. Here are their scores over the last five games.

The Gremlins				
30	32	45	39	41

The Gladiators				
28	53	45	21	61

1. Graph the data points on the number lines below.



2. Which scores seem to have greater spread (variability)? Justify your answer.

The mean absolute deviation (MAD) is another measure of spread for a data set. It is the arithmetic average of the distance of the data points to the mean.

3. Compute the MAD statistic.

Calculating MAD	Gremlins	Gladiators
<ul style="list-style-type: none"> Find the mean of the data. Mark it on the number line above with an arrow. 		
<ul style="list-style-type: none"> Find the distance between each data point and the mean on the number lines. 		
<ul style="list-style-type: none"> Find the sum of these distances to the mean. 		
<ul style="list-style-type: none"> Find the mean (average) of these five distances. This is the MAD statistic. Circle data points that fall within this distance from the mean. 		

4. What do these MAD scores say about how spread out the scores are for these teams?

PRACTICE 3

This data shows the number of hours of online games played in one week by 13 teens.

13	28	15	10	10	17	4	15	17	10	8	11	6
----	----	----	----	----	----	---	----	----	----	---	----	---

1. Rewrite the data set in order from least to greatest in the table below.

--	--	--	--	--	--	--	--	--	--	--	--	--

2. Calculate measures of center and spread associated with the median.

range _____ median _____

five-number summary _____

interquartile range (IQR) _____

3. What fraction of the observations falls in the IQR?

4. Calculate measures of center and spread associated with the mean.

mean _____ mean absolute deviation (MAD) _____

5. Does 28 appear to be an outlier? _____ How does this value affect measures of center?

6. Use your analysis to answer: How many hours do teens typically spend playing online games?

7. Record the meaning of mean absolute deviation in **My Word Bank**.

PRACTICE 4: EXTEND YOUR THINKING

This table shows heights of players on Ward High School’s Football team.

height	5'5"	5'8"	5'9"	5'10"	5'11"	6'	6'1"	6'2"	6'3"
# of players	1	2	3	9	6	6	6	7	2

1. What characteristic (attribute) is described in the table?
2. How is the characteristic measured?
3. How many data points (observations) are in the table?

What does this number represent?

For the table below, use a calculator as needed. Explore the question “What is the typical height of a player on the football team?” with statistical measures. Hint: changing heights to inches may be easier.

4. Find measures of center and spread. Explain what they mean.

Statistical Measure	Value	Explanation
Range		
Median		
Five-number summary		
Interquartile Range		
Mean		
Mean Absolute Deviation		

5. A new player joins the team and increases the range by 2 inches. What could be the height of the player?

What effect would this player have on the median and the mean?

DATA DISPLAYS

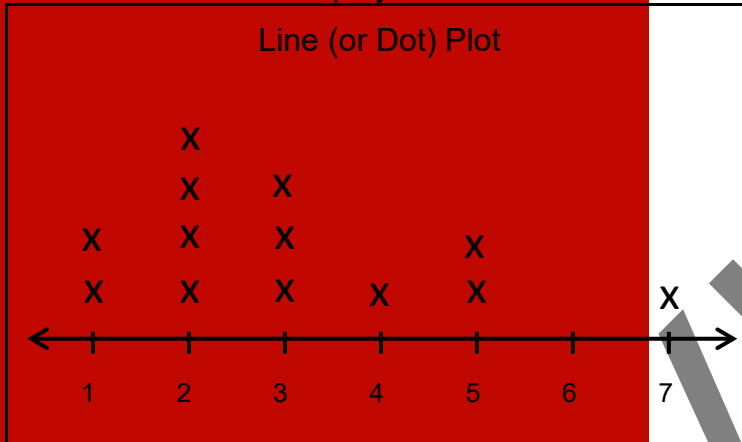
We will display data by constructing line plots (dot plots), histograms, and box plots (box-and-whisker plots). We will describe and analyze the overall shape of the data.

[6.SP.4, 6.SP.2, 6.SP.5abcd; SMP2, 4, 6]

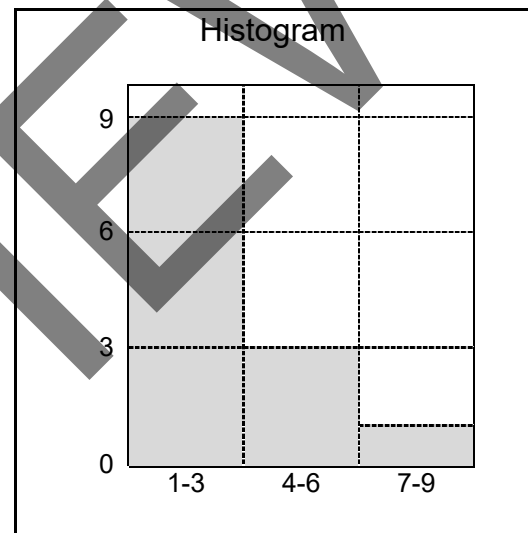
GETTING STARTED

Here are three data displays for the same data set.

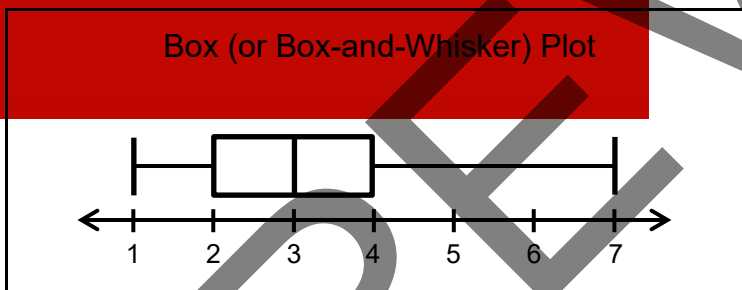
Line (or Dot) Plot



Histogram



Box (or Box-and-Whisker) Plot

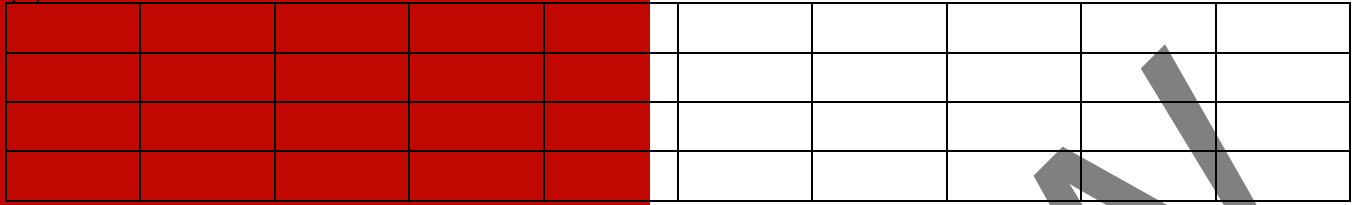


1. Make a list, in order, of the data values in the displays.
2. Write the five-number summary for the data. Circle the median on each display if possible.
3. Find the mean.
4. Which has greater value, the median or the mean? Why?

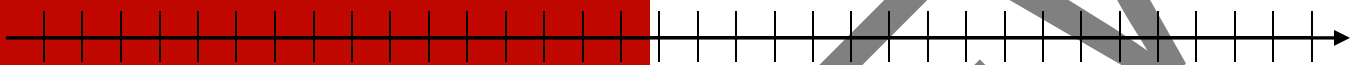
THREE DATA DISPLAYS

Follow your teacher's directions for (1) – (5).

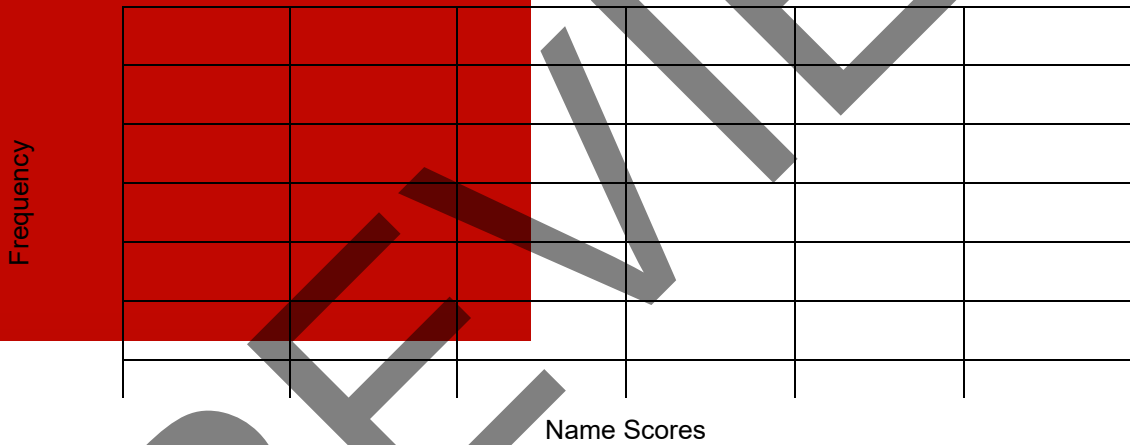
(1)



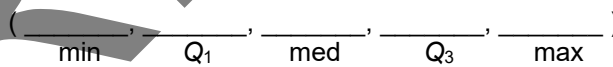
(2)



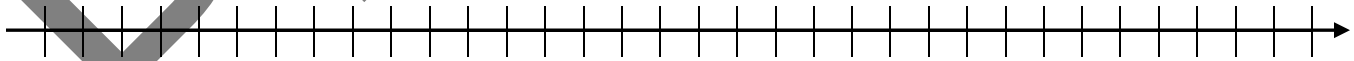
(3)



(4)



(5)



6. Circle the words below that could be used to describe the shape of the data on our name score data displays.

- symmetric gap peak cluster tail outlier

PRACTICE 5

Recall the number of hours of online games reported by teens in one week.

13	28	15	10	10	17	4	15	17	10	8	11	6
----	----	----	----	----	----	---	----	----	----	---	----	---

1. Record the data, in order, from least to greatest, and write the five-number summary.

2. Construct a line plot of the gaming hours and label appropriately.

3. Construct a histogram of the gaming hours and label appropriately.

4. Construct a box plot of the gaming hours data.

5. Describe the shape of the data. Use words such as symmetry, gap, peak, cluster, tail, or outlier if appropriate.

PRACTICE 6: THE FOOTBALL TEAM REVISITED

Recall the heights of players on Ward High School’s Football team.

height	5’5”	5’8”	5’9”	5’10”	5’11”	6’	6’1”	6’2”	6’3”
# of players	1	2	3	9	6	6	6	7	2

1. Make a line (dot) plot of heights of the team. Describe the shape of the data, based on this plot.

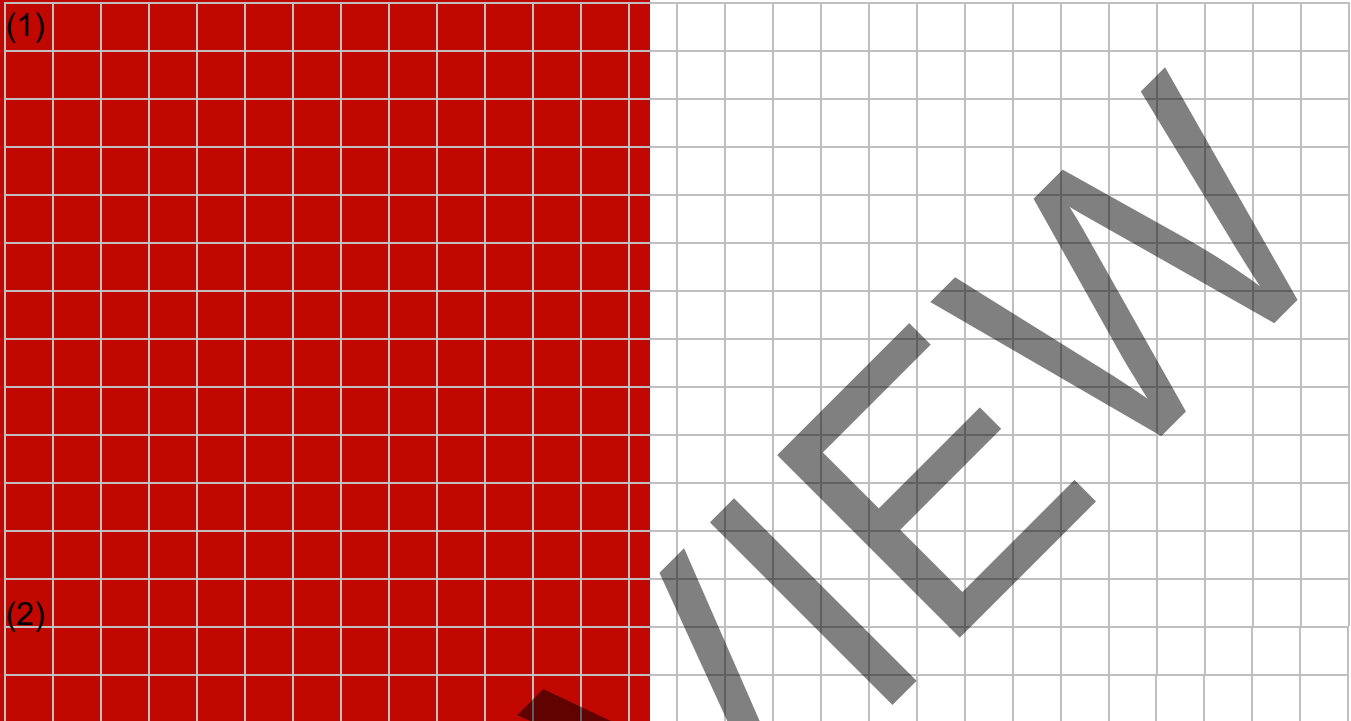
2. Make a box plot of heights of the team. Describe the shape of the data, based on this plot.

3. Which statistical display do you think best helps you to describe the shape of the data? Why?

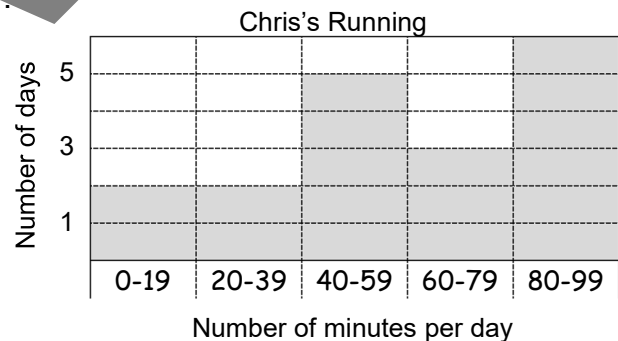
4. Answer this statistical question by referring to displays that you think are most useful: What is the typical height of a player on the Ward High School football team? Why?

INTERPRETING HISTOGRAMS

Follow your teacher's directions for (1) – (2).



Chris is training for a half-marathon in the fall. The histogram below shows the number of minutes he ran each training day over the summer.



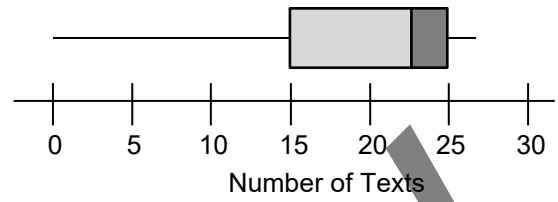
- Can you tell from this display how many minutes he ran each day? If so, how many? If not, why not?
- Can you tell from this display how many days he ran? If so, how many? If not, why not?
- Based on this display, describe Chris's typical training schedule.
- Why is it important to pay attention to histogram intervals?

INTERPRETING BOX PLOTS

Follow your teacher's directions.

(1)

(_____ , _____ , _____ , _____ , _____)



(2)

(3)

(4)

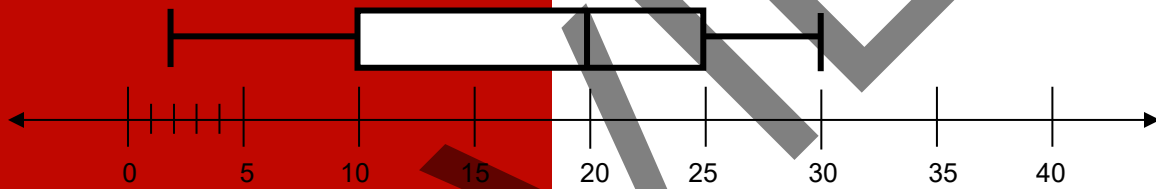
(5)

PRACTICE 7

Let the data displays on **Getting Started** represent data collected about the number of hours teens typically play sports on the weekend.

1. How many teens were surveyed? Explain how you know.
2. What is the typical number of hours played? Explain why you think this.
3. Do you think that anything about these displays is misleading? Explain.

The box plot below contains only whole number values. Use the box plot for problems 4 – 9.



4. Create an appropriate title for this box plot if the data represents prices.
5. What is the range of the data set?
6. List the five-number summary. ($\underline{\hspace{1cm}}$, $\underline{\hspace{1cm}}$, $\underline{\hspace{1cm}}$, $\underline{\hspace{1cm}}$, $\underline{\hspace{1cm}}$)
min, Q₁, med., Q₃, max
7. The median does not split the box into two identical rectangles. Explain what this means.
8. Why can't we determine the mean from the box plot?
9. What conclusions can you draw from the shape of the graph?

REVIEW

POSTER PROBLEMS: STATISTICS

Part 1: Your teacher will divide you into groups.

- Identify members of your group as A, B, C, or D.
- Each group will start at a numbered poster. Our group start poster is _____.
- Each group will have a different colored marker. Our group marker is _____.

Part 2: Do the problems on the posters by following your teacher's directions.

Poster 1 (or 5)	<i>Shop Shoes</i> sold the following sizes during the last hour. 9, 7, 8, 8, 10, 8, 6, 5, 9, 8
Poster 2 (or 6)	Below are the housing prices (in thousands) for the most recent sales in Mathville. \$475, \$470, \$460, \$375, \$500, \$450, \$650, \$480, \$500, \$410
Poster 3 (or 7)	Teens were surveyed on the number of hours per week they spend looking at a screen. 63, 50, 40, 15, 35, 45, 54, 29, 25, 37, 49, 38
Poster 4 (or 8)	The number of pets students own are shown below. 3, 4, 2, 0, 1, 2, 12, 4, 2, 3, 5, 1, 0, 2, 4
<p>A. Copy the data in numerical order and determine the median and mode.</p> <p>B. Determine the mean for the data set and note any potential outliers.</p> <p>C. Find the five-number summary for the data set.</p> <p>D. Make a data display that would be appropriate for the data set. Be sure to label the graph.</p>	

Part 3: Return to your seats with your original poster. Work with your group.

Write a statistical question that can be answered with your group's data set and display. Answer the question and explain your thinking using the measures of center, variability, and/or the data display.

MATCH AND COMPARE SORT: STATISTICS

1. Individually, match words with descriptions. Record results.

Card set \triangle			Card set \circ		
Card number	word	Card letter	Card number	word	Card letter
I			I		
II			II		
III			III		
IV			IV		

2. Partners, choose a pair of numbered matched cards and record the attributes that are the same and those that are different.

3. Partners, choose another pair of numbered matched cards and discuss the attributes that are the same and those that are different.

WHY DOESN'T IT BELONG?: STATISTICS

Use the data set with eight data points below, and the box with four numbers to the right.

2 4 6 8 10 12 14 18

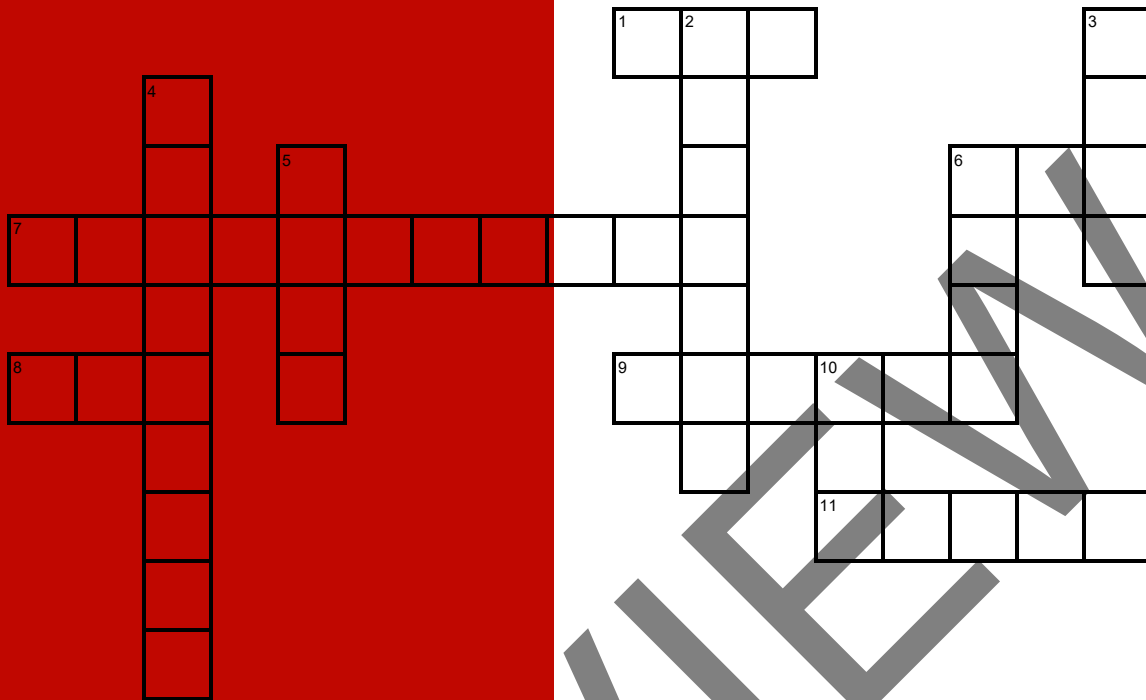
All responses below should include statistical concepts and vocabulary.

A 9.25	B 39
C 12	D 5

1. Choose each of the four numbers in the box and explain why it does not belong with the other three in relation to the given data set.

2. Make any data display using the eight data points given at the top of the page. Create a reasonable context for these data points, and label the display appropriately.

VOCABULARY REVIEW



Across

- 1 plot based on the five-number summary
- 6 average of the distances of data points from the mean (abbreviation)
- 7 a question where data that has potential for variability, and can be collected and analyzed
- 8 plot that shows data frequencies above a number line
- 9 middle number of an ordered data set of data
- 11 difference between greatest and least value in a data set

Down

- 2 value that has striking deviation from others in a data set
- 3 most common value in a data set
- 4 data points that divide an ordered data set into four equally-sized groups
- 5 number of values in a data set summary that includes quartiles
- 6 the arithmetic average
- 10 difference between Q_3 and Q_1 (abbr.)

SPIRAL REVIEW

1. **Computational Fluency Challenge:** This paper and pencil exercise will help you gain fluency with multiplication and division. Try to complete this challenge without any errors. No calculators!
- a. Begin with the number 2. Multiply your number by 6. Multiply the result by 7. Multiply that result by 8. Multiply that result by 9. (You should have a “big number.”)
I began with the number _____. After multiplying, my big number is _____.
- b. Start with your big number from above. Divide it by 6. Divide that result by 7. Divide that result by 8. Divide that result by 9. After dividing, I got _____.
- (Organize and show work below.)

SPIRAL REVIEW

Continued

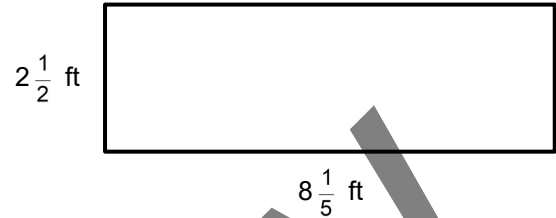
2. Sage and Mason were playing a video game. Sage started with 1,345 points. Sage then lost 329 points, gained 415 points, lost a third of his points and then doubled his points. Mason started with 580, gained 1,002 points, lost half the points, and gained 272 points. Who had the most points at the end?

3. Sookie went to the craft store to buy items for her project. She bought 2 containers for \$2.75 each, three-fourths yard of fabric at \$7.00/yard, one box of buttons for \$3.42 and a dozen foam balls for \$0.78 each.
- Write a numerical expression for the total bill.
 - Simplify the expression.

SPIRAL REVIEW

Continued

4. Kristi is painting a board for a home improvement project, shown on the right.



- a. What is the area of the board?
- b. What is the perimeter of the board?
- c. What is the difference between the larger side and the smaller side?

5. Write 3 equivalent fractions following the directions below.

	Write the fraction in its simplest form.	Write an equivalent fraction with 24 as the denominator.	Write an equivalent fraction with 18 as the numerator.
$\frac{3}{4}$			
$\frac{2}{12}$			
$\frac{3}{6}$			

REFLECTION

1. **Big Ideas.** Shade all circles that describe big ideas in this unit. Draw lines to show connections that you noticed.

<p style="text-align: center;">Extend the number system to include negatives. <input type="radio"/></p> <p style="text-align: center;">Explore relationships between inputs and outputs. <input type="radio"/></p> <p style="text-align: center;">Rewrite and evaluate expressions and solve equations. <input type="radio"/></p>	<p style="text-align: center;"><input type="radio"/> Investigate concepts and solve problems involving length, area, and volume.</p> <p style="text-align: center;"><input type="radio"/> Use statistical measures and displays to describe center and spread.</p> <p style="text-align: center;"><input type="radio"/> Gain computational fluency with positive rational numbers.</p> <p style="text-align: center;"><input type="radio"/> Explore and apply ratio and rate reasoning and representations.</p>
---	---

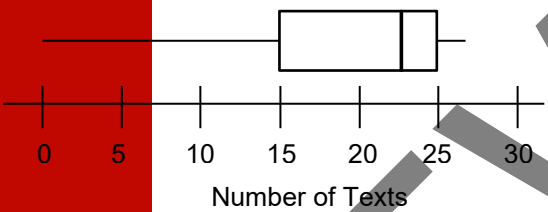
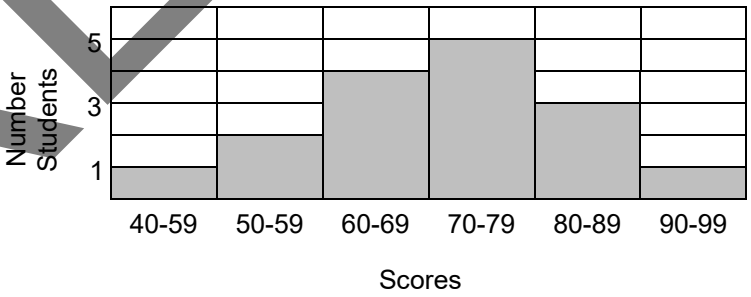
Give an example from this unit of one of the connections above.

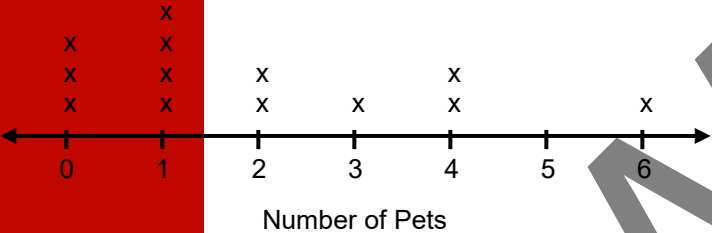
2. **Unit Progress.** Go back to **Monitor Your Progress** on the cover and complete or update your responses. Explain something you understand better now than before or something you would still like to work on.

3. **Mathematical Practice.** What tools did you find useful as you explored relationships of shapes and space [SMP5]? Then circle one more SMP on the back of this packet that you think was addressed in this unit and be prepared to share an example.

4. **More Connections.** Explain how you communicated a story with data.

STUDENT RESOURCES

Word or Phrase	Definition
<p>box plot</p>	<p>A <u>box plot</u>, or <u>box-and-whiskers plot</u>, is a graphical representation of the five-number summary of a data set. See <u>five-number summary</u>.</p> <p style="text-align: center;">Box Plot of Number of Texts Per Day of 6th Graders</p> 
<p>dot plot</p>	<p>A <u>dot plot</u> is a graphical representation of a data set where the data values are represented by dots above a number line. See <u>line plot</u>.</p>
<p>five-number summary</p>	<p>The <u>five-number summary</u> of a data set consists of its minimum value (min), first quartile Q_1, median Q_2, third quartile Q_3, and maximum value (max). The five-number summary is usually written in the form (min, Q_1, med., Q_3, max).</p> <p style="text-align: center;">The five-number summary of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is given by (min, Q_1, med., Q_3, max) = (1, 1, 5, 6.5, 23).</p>
<p>histogram</p>	<p>A <u>histogram</u> is a graphical representation of frequencies of a numerical variable using rectangles. For a histogram, the horizontal axis is divided into intervals. Each interval forms the base of a rectangle whose height corresponds to the frequency of values of the variable in that interval.</p> <p style="text-align: center;">Quiz Scores of a Class of 16 Students</p> 
<p>interquartile range</p>	<p>The <u>interquartile range</u> (IQR) of a numerical data set is the difference between the third quartile and the first quartile of the data set. The interquartile range is a measure of the variation of the data set.</p> <p style="text-align: center;">For the data set 1, 1, 1, 3, 5, 5, 6, 7, 23, $Q_1 = 1$, $Q_3 = 6.5$, and $IQR = 5.5$</p>

Word or Phrase	Definition
<p>line plot</p>	<p>A <u>line plot</u> is a graphical representation of a data set where the data values are represented by marks, such as dots or X's, above a number line. See <u>dot plot</u>.</p> <p style="text-align: center;">Line Plot of Number of Pets for 13 Students</p>  <p style="text-align: center;">Number of Pets</p>
<p>mean</p>	<p>The <u>mean</u> of a data set is a measure of center equal to the average of the values in the data set. The mean is calculated by adding the values in the data set and dividing by the number of data values.</p> <p style="text-align: center;">The mean of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is</p> $\frac{1 + 1 + 1 + 3 + 5 + 5 + 6 + 7 + 23}{9} = 5\frac{7}{9} = 5.77\dots$
<p>mean absolute deviation</p>	<p>The <u>mean absolute deviation</u> (MAD) of a data set is the average of the (positive) differences between the values in the data set from the mean. The MAD is a measure of the variation of the data set.</p> <p style="text-align: center;">For the data set {3, 3, 5, 6, 6}, the mean is 4.6. The distances of the data points to the mean are 1.6, 1.6, 0.4, 1.4, and 1.4. The MAD is $\frac{1.6 + 1.6 + 0.4 + 1.4 + 1.4}{5} = 1.28$</p>
<p>measure of center</p>	<p>A <u>measure of center</u> is a statistic describing the middle of a data set.</p> <p style="text-align: center;">The mean, the median, and the mode are three commonly used measures of center of a numerical data set.</p>
<p>measure of spread</p>	<p>A <u>measure of spread</u> is a statistic describing the variability of a data set. It describes how far the values in a data set are from the mean or median.</p> <p style="text-align: center;">The standard deviation, the mean absolute deviation (MAD), and the interquartile range (IQR) are three measures of spread of a numerical data set.</p>
<p>median</p>	<p>The <u>median</u> of a data set is a measure of center equal to the middle number in the data set, when the values are placed in order from least to greatest. If there is an even number of values in the data set, the median is taken to be the mean (average) of the two middle values.</p> <p style="text-align: center;">The median of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is 5, since the first 5 is the middle value.</p> <p style="text-align: center;">The median of the data set 5, 6, 7, 23 is the mean (average) of the two middle numbers, $(6 + 7) \div 2 = 6.5$, which is the average of 6 and 7.</p>

Word or Phrase	Definition
mode	<p>The <u>mode</u> of a data set is the value(s) that occur(s) most often. A data set may have more than one mode. It may also have no mode if all values occur the same number of times.</p> <p>The mode of the data set 1, 1, 1, 3, 5, 6, 6, 7, 23 is 1, since the data value 1 occurs more frequently than any other data value. If a 6 were added to this data set, 6 would also be a mode.</p>
outlier	<p>An <u>outlier</u> of a data set is a data value that is a striking deviation from the overall pattern of values in the data set.</p> <p>For the data set 1, 1, 1, 3, 5, 6, 6, 7, 23, the data value 23 is a potential outlier. It appears unusually large relative to the other data values.</p>
quartiles	<p>The <u>quartiles</u> of a data set are points that divide the data set into four equally sized groups, when the values are placed in order from least to greatest. The <u>second quartile</u> is the median, denoted by Q_2. The <u>first quartile</u>, denoted by Q_1, is the median of the lower half of the data set (the data values less than the middle data value), and the <u>third quartile</u>, denoted by Q_3, is the median of the upper half of the data set.</p> <p>Given the ordered data set 1, 1, 1, 3, 5, 5, 6, 7, 23,</p> <ul style="list-style-type: none"> • The middle value is the first 5: Median = 5. This is also the second quartile Q_2, • The lower half of the data set is 1, 1, 1, 3. Therefore $Q_1 = 1$. • The upper half of the data set is 5, 6, 7, 23. Therefore, $Q_3 = 6.5$.
range (of a data set)	<p>The <u>range</u> of a numerical data set is the difference between the greatest and least values in the data set.</p> <p>The range of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is 22, since $22 = 23 - 1$.</p>
statistical question	<p>A <u>statistical question</u> is a question where numerical data that has potential for variability can be collected and analyzed for the purpose of answering the question.</p> <p>A statistical question: "How much TV do students in my class watch on average?" NOT a statistical question: "How many hours of TV did you watch last week?"</p>

Finding Measures of Center

Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To find the median, order the value from least to greatest and find the middle number. If there is an even number of values in the data set, the median is the mean (average) of the two middle numbers.

The median for the siblings data set: 1, 1, 2, 2, 2, 2, (3), 3, 3, 4, 5, 5, 7

To find the mode, find the value(s) that occur(s) most often.

The mode for the siblings data set: the value of 2 occurs most often.

To find the mean (average) of a data set, add all the values in the data set and divide it by the number of values (number of observations, n).

Number of observations: $n = 13$

The mean for the siblings data set: $\frac{3 + 4 + 5 + 2 + 2 + 3 + 3 + 2 + 2 + 5 + 7 + 1 + 1}{13} = 3.08$

Finding the Range and the Quartiles

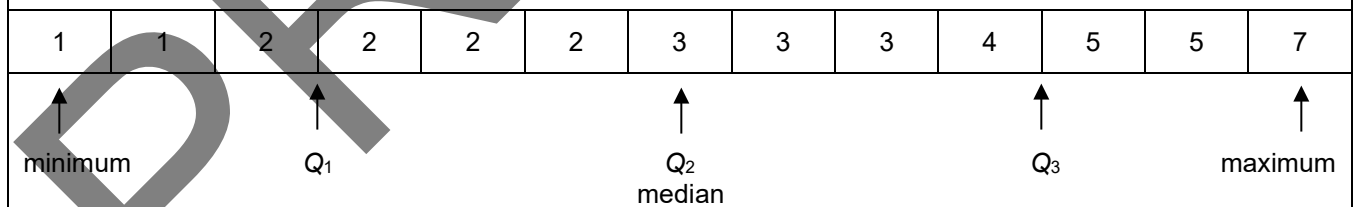
Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To find the range of a data set, find the difference between the greatest and least values in the data set.

For the siblings data set, the range is 6, since $7 - 1 = 6$

To find quartiles, first put the numbers in numerical order. Then locate the points that divide the set into four equal parts.



For the siblings data set: $Q_1 = 2$ (the 1st quartile)
 $Q_2 = 3$ (the 2nd quartile)
 $Q_3 = 4.5$ (the 3rd quartile)

Note that Q_1 is the median of the first half of the data set and Q_3 is the median of the second half.

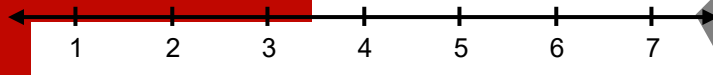
How to Construct a Dot Plot

A dot plot (also called a line plot) displays data on a number line with a dot (•) or an X to show the frequency of data values.

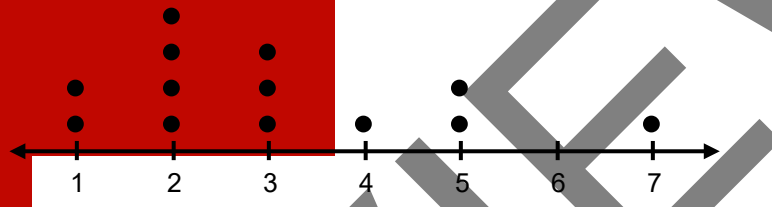
Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

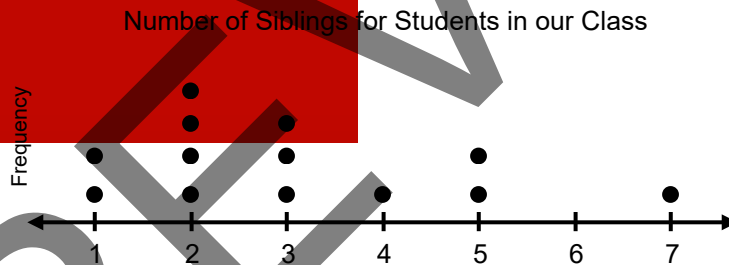
1. Make a number line that extends from the minimum data value to the maximum data value.



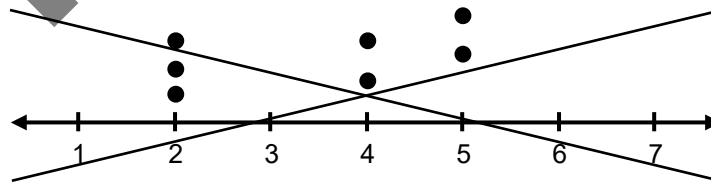
2. Mark a dot or an X for every data value.



3. Write a title and labels.



Heads Up! Be sure to line up dots or X's properly. The graph below is visually misleading in a few places. The number of dots at 4 and 5 are the same, but one set is higher than the other, possibly implying there are more. The number of dots at 2 and 4 are different, but they peak at the same height, possibly implying there are the same number of dots.



How to Construct a Histogram

A **histogram** is a data display that uses adjacent rectangles to show the frequency of data values in intervals. The height of a given rectangle shows the frequency of data values in the interval shown at the base of the rectangle.

Nancy asks each of her 21 classmates how many coins they have in their backpacks. Then she puts the data set in order.

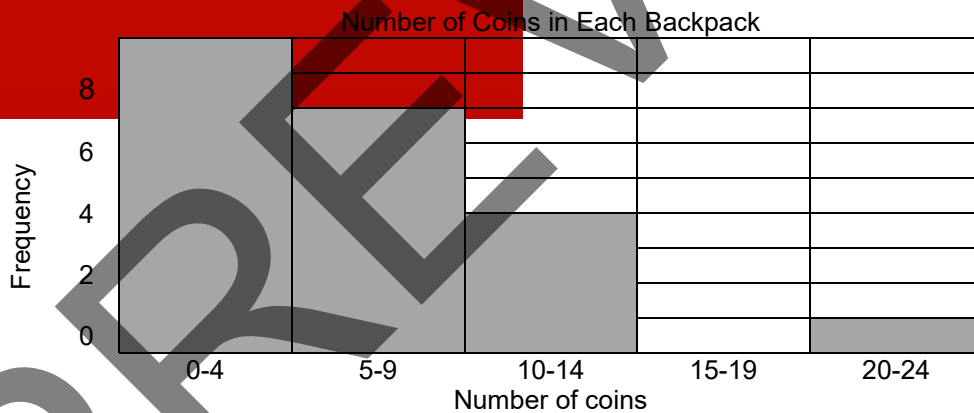
0, 0, 1, 2, 2, 2, 2, 3, 3, 5, 5, 7, 7, 7, 7, 7, 10, 10, 10, 12, 21

To construct the histogram:

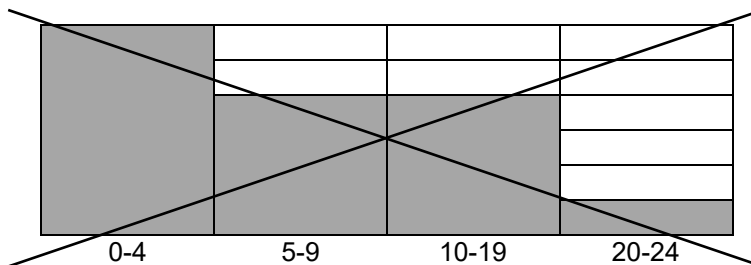
1. Divide the number of coins into equally spaced intervals and make a frequency table:
(Here we choose intervals of five.)

Intervals (number of coins)	Frequency
0-4	9
5-9	7
10-14	4
15-19	0
20-24	1

2. Record frequencies as rectangles on a data display. Add a title and label the axes.



Heads Up! Be sure to make equally spaced intervals. The graph below is visually misleading. The third column has an interval that is twice the others, but the same number of data points as the column to the left of it



How to Construct a Box Plot

A box plot (or box-and-whisker plot) is a visual representation of the center and spread of a data set. The display is based on the five-number summary.

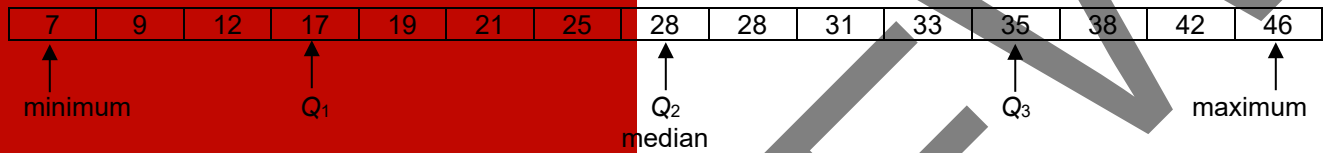
Here are the ages of 15 people:

21, 12, 28, 17, 46, 35, 7, 38, 42, 33, 19, 9, 31, 25, 28

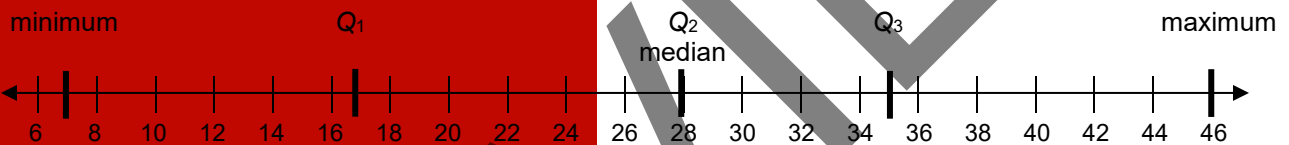
1. Write the values of the data set from least to greatest.

7, 9, 12, 17, 19, 21, 25, 28, 28, 31, 33, 35, 38, 42, 46

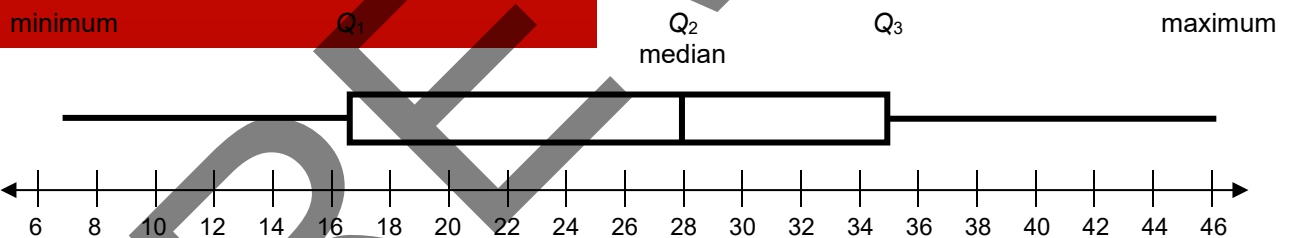
2. Find the five-number summary.



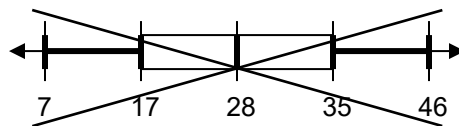
3. Locate the five-number summary values on a number line, and indicate with vertical segments.



4. Create a "box" to highlight the interval from the first to the third quartile, and draw "whiskers" that extend to the minimum and maximum.



Heads Up! Be sure to scale the box and whisker plot properly. This plot is **WRONG**:



COMMON CORE STATE STANDARDS

STANDARDS FOR MATHEMATICAL CONTENT	
6.SP.A	Develop understanding of statistical variability.
6.SP.1	Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. <i>For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students’ ages.</i>
6.SP.2	Understand that a set of data collected to answer a statistical question has a distribution that can be described by its center, spread, and overall shape.
6.SP.3	Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
6.SP.B	Summarize and describe distributions.
6.SP.4	Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
6.SP.5	Summarize numerical data sets in relation to their context, such as by: <ul style="list-style-type: none"> a. reporting the number of observations. b. describing the nature of the attribute under investigation, including how it was measured and its units of measurement. c. giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered. d. relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

STANDARDS FOR MATHEMATICAL PRACTICE	
SMP1	Make sense of problems and persevere in solving them.
SMP2	Reason abstractly and quantitatively.
SMP3	Construct viable arguments and critique the reasoning of others.
SMP4	Model with mathematics.
SMP5	Use appropriate tools strategically.
SMP6	Attend to precision.
SMP7	Look for and make use of structure.

