# Data Displays

Box Plots
a graphical representation of the five number summary

A **box plot** (also called box and whisker plot) is used to show how data values are distributed. They are created using five important numbers that show the minimum, maximum, median, lower quartile, and upper quartile.

In a box plot, a rectangle is drawn starting at the first quartile and ending at the third quartile. The rectangle shows the middle 50% of the data set. The median is represented by a line. Whiskers are drawn from the rectangle to the minimum and maximum data values. An example of a box plot is below:


Test Scores (out of 100)
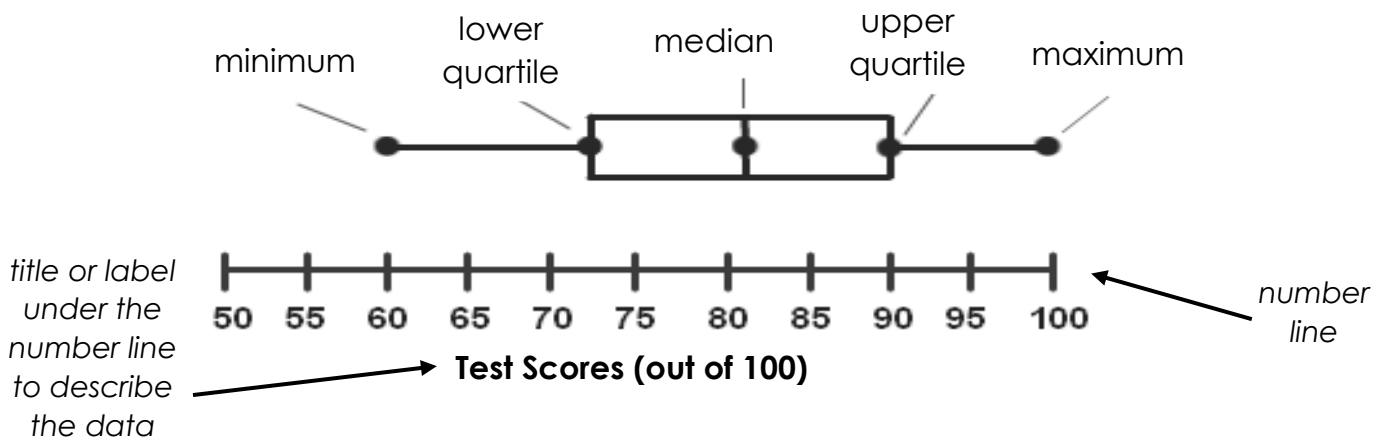
Shapes of Box Plot Distributions (Types of Distributions)

SHAPE/TYPE	DESCRIPTION	PICTURE
SYMMETRIC	When graphed, a vertical line drawn at the center will form mirror images. This shape is referred to as the bell shaped curve or normal curve. The median and mean will be approximately equal.	
UNIFORM	The data is spread equally (or very close to equally) across the range. Uniform distributions are a type of symmetric distributions. The median and mean will be approximately equal.	
SKEWED LEFT (NEGATIVE SKEW)	Fewer data points are found to the left of the graph. The "tail" of the graph is to the left. The mean is less than the median.	
SKEWED RIGHT (POSITIVE SKEW)	Fewer data points are found to the right of the graph. The "tail" of the graph is to the right. The mean is greater than the median.	

Outliers: A data value that lies on the outside of all the other data values. It is denoted by an asterisk (*) or dot.

Data sets can have no outliers, one outlier, or multiple outliers. Outliers can be found on the low end of the data, the high end of the data, or both.

Ages of Oscar Winning Actors from 1975 to 2004
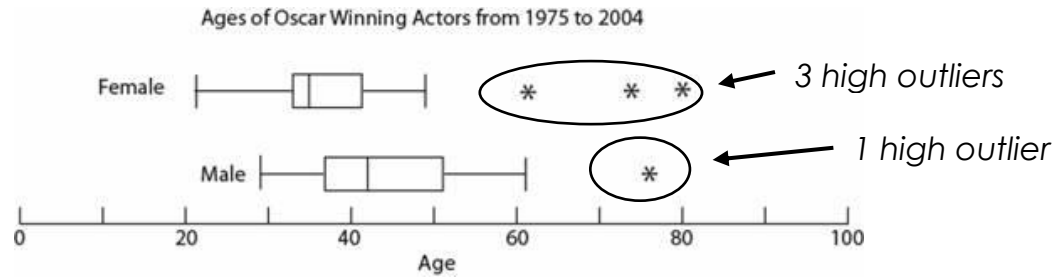
Picking the Best Measure of Center

Mean: _____

Median: _____

Mode: _____

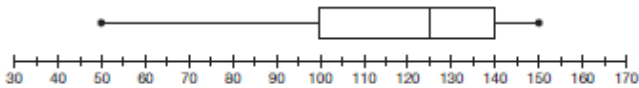## Picking the Best Measure of Spread

Interquartile Range: _____
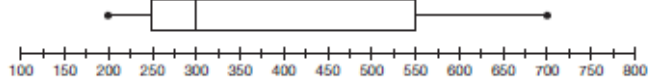
Mean Absolute Deviation: _____

Identifying Distributions

Identify the type of distribution of the following box plots. Then identify the best measure of center and spread.
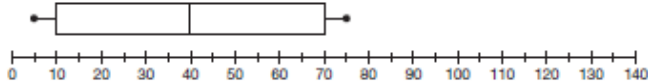
a.

b.

c.

Creating Box Plots

The data below represents the age at which various students obtained their first job. Calculate the five number summary (by hand or with your calculator) and create a box plot to represent this data.

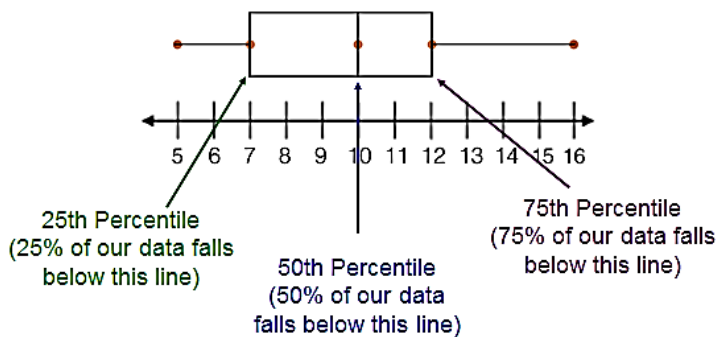{15, 11, 12, 18, 10, 21, 11, 17, 13}
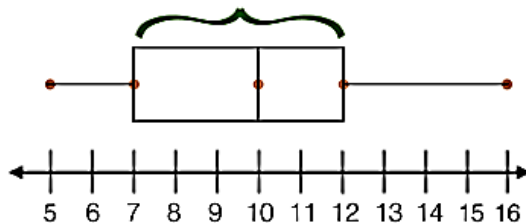
Min: _____   Q1: _____   Med: _____   Q3: _____   Max: _____

Interpreting Box Plots

25th Percentile
(25% of our data falls below this line)

50th Percentile
(50% of our data falls below this line)

75th Percentile
(75% of our data falls below this line)

Interquartile Range (IQR)
Middle 50% of data
Q1 - Q3

1) 25% of data values fall at or below _____.

2) 25% of data values fall at or above _____.

3) 50% of data values fall at or above _____.

4) 75% of data values fall at or above _____.

5) 75% of data values fall at or below _____.

6) The middle 50% of data values fall between _____ and _____.

7) 100% of data values fall between _____ and _____.

Practice with Box Plots

Example 1: Analyze the box plot below about the cost, in dollars, of 12 CD's. Answer the questions.


Cost of CD's (in dollars)
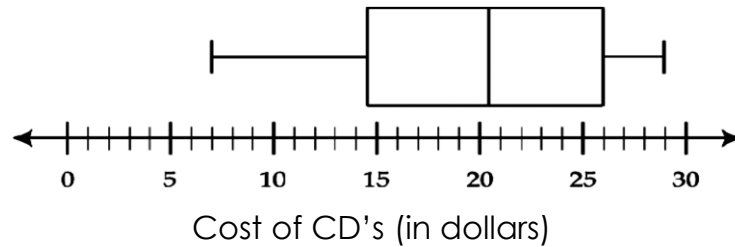
A. Which cost is the upper quartile?

B. What is the range?

C. What is the median?

D. Which cost represents the 100th percentile?

E. How many CD's cost between $14.50 and $26.00?

F. How many CD's cost less than $14.50?

Example 2: Analyze the box plot below and answer the following questions:

Heights of 40 Randomly Chosen Adults
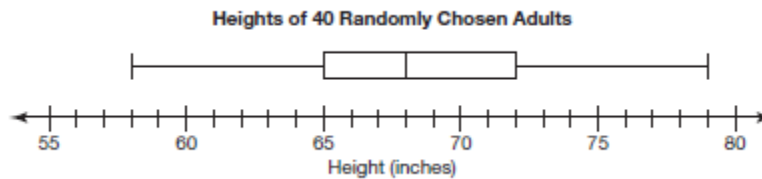
Height (inches)

A. What is the height range of the middle 50 percent of the surveyed adults?

B. How many of the surveyed adults are between 72 and 79 inches?

C. What percent of the surveyed adults are 72 inches or shorter?

D. What is the height of the tallest adult surveyed?

E. About 10 people have a height below what amount?

F. About 20 people have a height above amount?

G. How many of the surveyed adults are at least 58 inches tall?

H. Describe the distribution. Is the median or mean best describe the data?

Example 3: Jamie has organized the amount of sugar, per serving, in 60 different cereals and created a box plot of his data below:

Sugar in Breakfast Cereals

Sugar Amount in One Serving (grams)
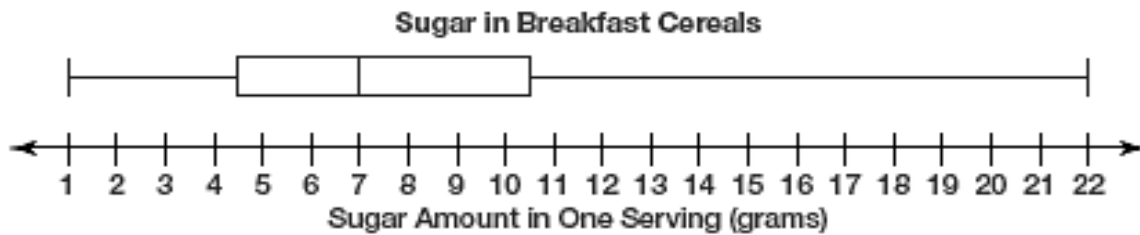
a. State the numbers (including what they represent) for the five number summary.

b. Give three conclusions that can be made about the sugar amount in one serving of breakfast cereal.

▫ _____

_____

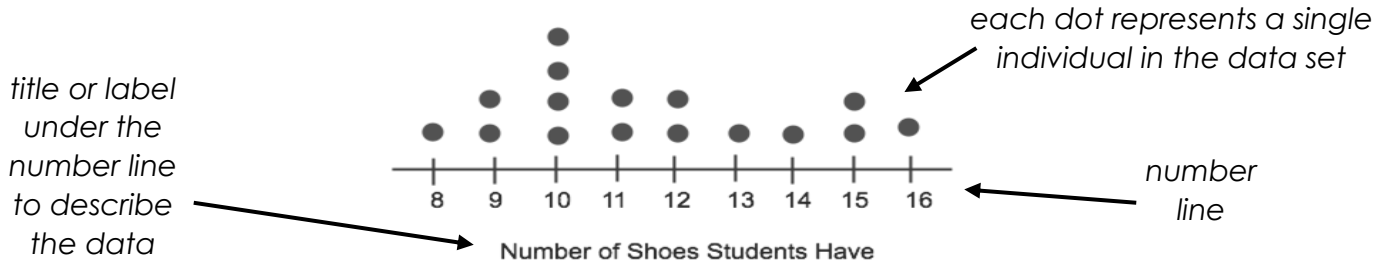▫ _____

_____

▫ _____

_____

c. Describe the distribution and interpret the meaning of the distribution in terms of this problem situation.

d. Jamie says that more breakfast cereals have over 10 grams of sugar per serving than have under 5 grams of sugar per serving because the whisker connecting Q3 to the maximum is longer than the whisher connecting Q1 to the minimum. Is he correct? Explain why or why not.

Dot Plots
shows all data points

A **dot plot** is a data representation that uses a number line and x's, dots, or other symbols to show frequency. The number of times a value is repeated corresponds to the number of dots above that value. A dot plot also shows the size of the data set.

title or label under the number line to describe the data

each dot represents a single individual in the data set

number line

Number of Shoes Students Have

Advantages of Dot Plots: Simple to make
Shows each individual data point

Disadvantages of Dot Plots: Can be time consuming with lots of data points
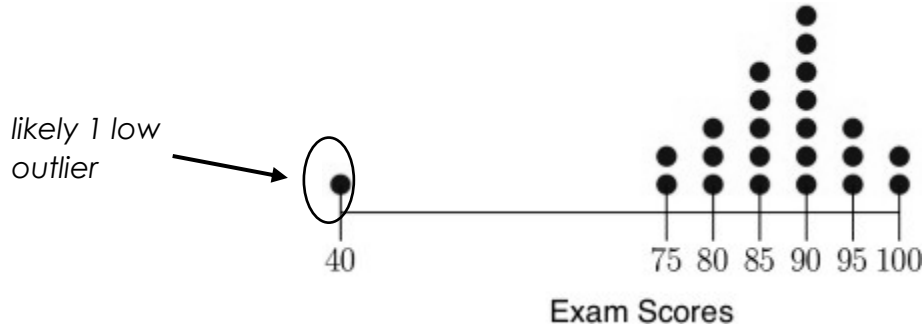Have to count to get exact total
Fractions are hard to display

Shapes/Types of Dot Plot Distributions

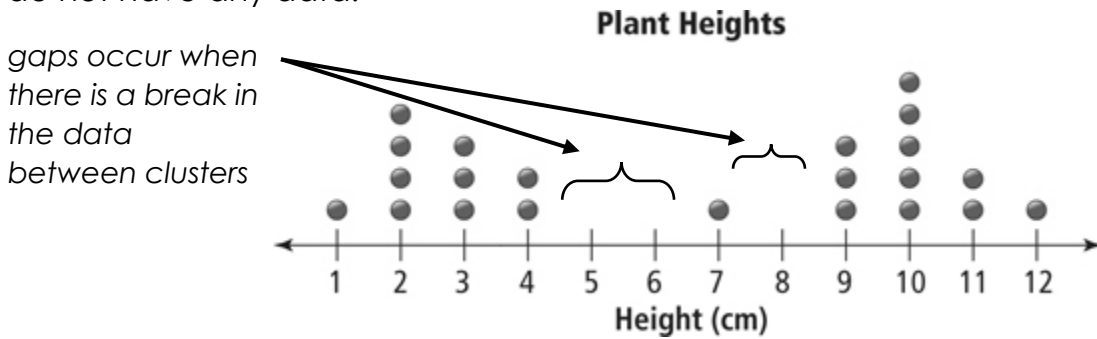| SHAPE/TYPE	DESCRIPTION	PICTURE
SYMMETRIC	When graphed, a vertical line drawn at the center will form mirror images. This shape is referred to as the bell shaped curve or normal curve. The median and mean will be approximately equal.	
UNIFORM	The data is spread equally (or very close to equally) across the range. Uniform distributions are a type of symmetric distributions. The median and mean will be approximately equal.	
SKEWED LEFT (NEGATIVE SKEW)	Fewer data points are found to the left of the graph. The "tail" of the graph is to the left. The mean is less than the median.	
SKEWED RIGHT (POSITIVE SKEW)	Fewer data points are found to the right of the graph. The "tail" of the graph is to the right. The mean is greater than the median.	

Other Characterisitcs

Data distributions modeled by dot plots can still have **outliers** – the outleirs however are not denoted with any special symbols. Outliers will typically be seen as dots that are far from the majority of the data.

Data sets can have no outliers, one outlier, or multiple outliers. Outliers can be found on the low end of the data, the high end of the data, or both.

likely 1 low outlier

Exam Scores

Gaps occur in dot plots when there are values on the number line at the bottom that do not have any data.

Plant Heights

gaps occur when there is a break in the data between clusters

Height (cm)

We can also describe the shapes of dot plots by describing the number of peaks the data has.

Bi-Modal Distributions have _____ peaks.   **Uni-Modal Distributions** have ____ peak.

Picking the Best Measure of Center

Mean: _____

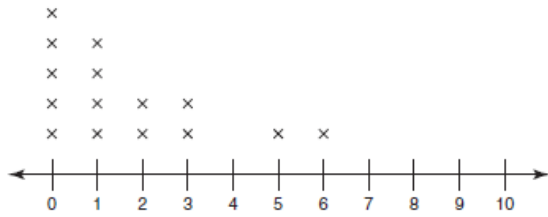Median: _____

Mode: _____

Picking the Best Measure of Spread
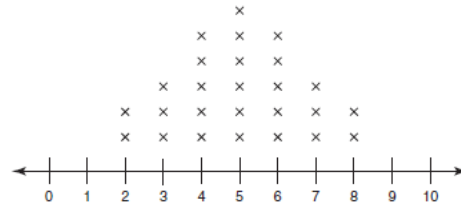
Interquartile Range: _____

Mean Absolute Deviation: _____

Practice 1: Identify the type of distribution of the following dot plots. Then identify the best measure of center and spread.
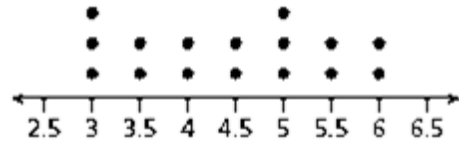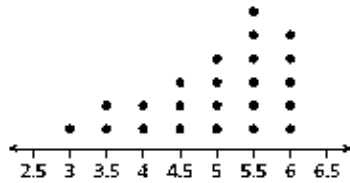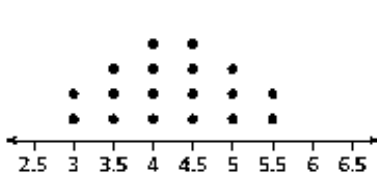
a.

b.

Practice 2: Determine the indicated information for each dot plot below.

Mean:

Median:

Mode:

Range:

Distribution shape:

Mean:

Median:

Mode:

Range:

Distribution shape:

Mean:

Median:

Mode:

Range:

Distribution shape:

Practice 3: The following dot plot represents gold medals won at the Special Olympics:

Number of Gold Medals Won
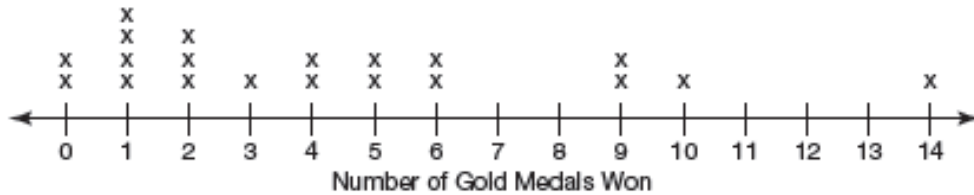
a. How many participants are represented in the dot plot?

b. How many participants won 10 or more medals?

c. How many participants won less than 4 medals?

d. Describe the data distribution and interpret its meaning in terms of this problem situation.

Practice 4: The dot plot below shows the resting heart rate for 16 male students at Harrison. Use this to answer the following questions.

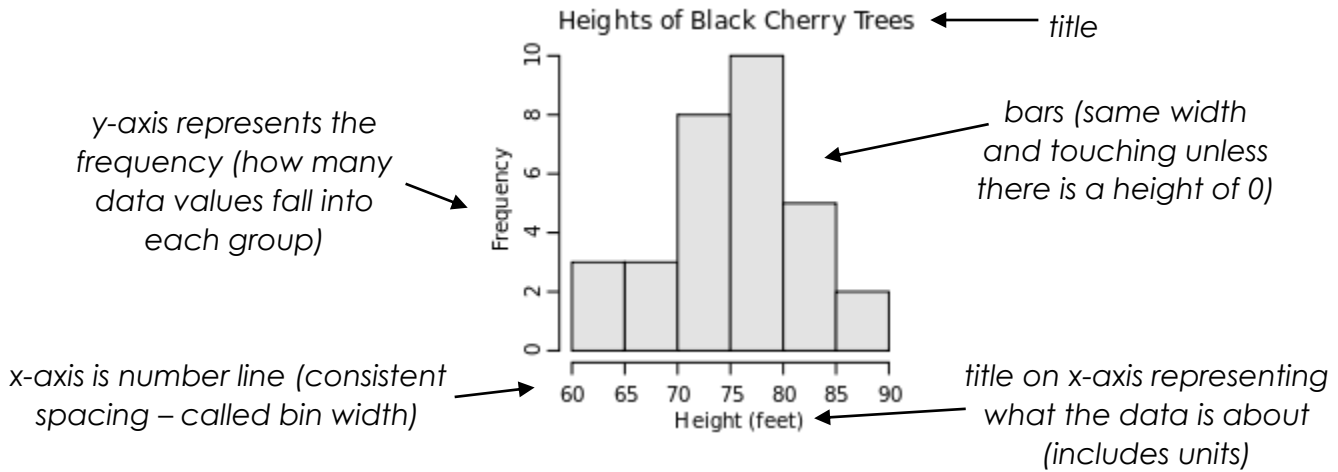Resting Heart Rate in Beats Per Minute (bpm)
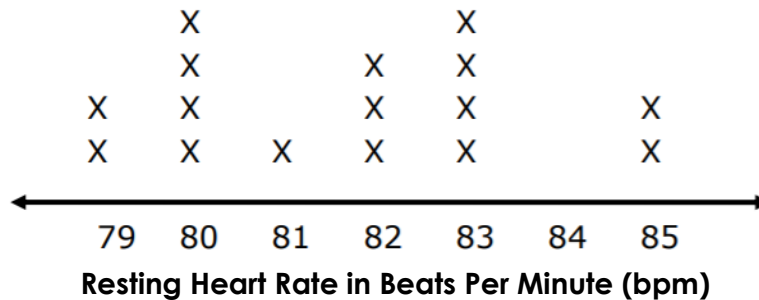
a. How many students had a resting heart rate of at least 83 bpm?

b. How many students had a rest heart rate of at most 81 bpm?

c. Write one sentence drawing a conclusion from the data.

Histograms
displays all the data by grouping

A **histogram** is a bar graph used to display the <u>frequency</u> of data divided into equal intervals, called **bins**. The bars must be of equal width and should touch, but not overlap. The height of each bar gives the frequency of the data.

Heights of Black Cherry Trees ← *title*

y-axis represents the frequency (how many data values fall into each group)

bars (same width and touching unless there is a height of 0)

x-axis is number line (consistent spacing – called bin width)

title on x-axis representing what the data is about (includes units)

Advantages of Histograms: Good for determining the shape of data
                                       Convenient for representing large quantities of data

Disadvantages of Histograms: Cannot read exact values because data is grouped into categories

More difficult to compare two data sets because measures of center and spread cannot be determined
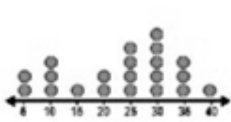
SHAPE/TYPE	DESCRIPTION	PICTURE
SYMMETRIC	When graphed, a vertical line drawn at the center will form mirror images. This shape is referred to as the bell shaped curve or normal curve The median and mean will be approximately equal.	
UNIFORM	The data is spread equally (or very close to equally) across the range. Uniform distributions are a type of symmetric distributions. The median and mean will be approximately equal.	
SKEWED LEFT (NEGATIVE SKEW)	Fewer data points are found to the left of the graph. The "tail" of the graph is to the left. The mean is less than the median.	
SKEWED RIGHT (POSITIVE SKEW)	Fewer data points are found to the right of the graph. The "tail" of the graph is to the right. The mean is greater than the median.	

Other Characterisitcs

We can also describe the shapes of dot plots by describing the number of peaks the data has.

Bi-Modal Distributions have _____ peaks.   **Uni-Modal Distributions** have ____ peak.

Data distributions modeled by dot plots can still have **outliers** – the outleirs however are not denoted with any special symbols. Outliers will typically be seen as dots that are far from the majority of the data.

Data sets can have no outliers, one outlier, or multiple outliers. Outliers can be found on the low end of the data, the high end of the data, or both.
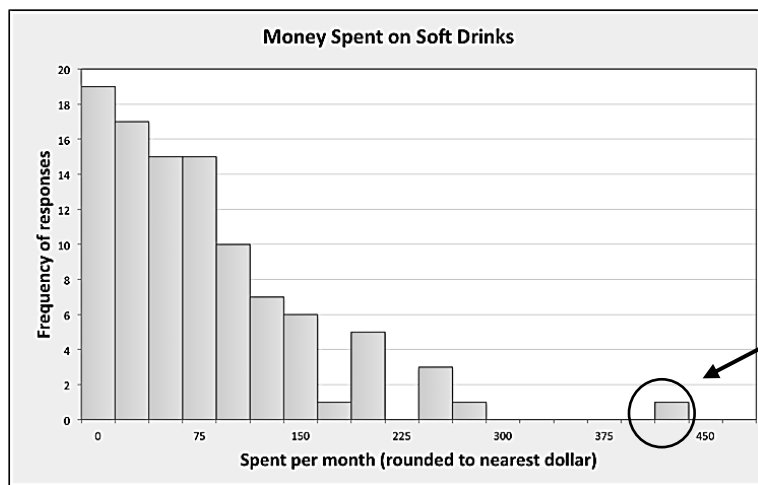
likely 1 high outlier

Gaps occur in dot plots when there are values on the number line at the bottom that do not have any data.
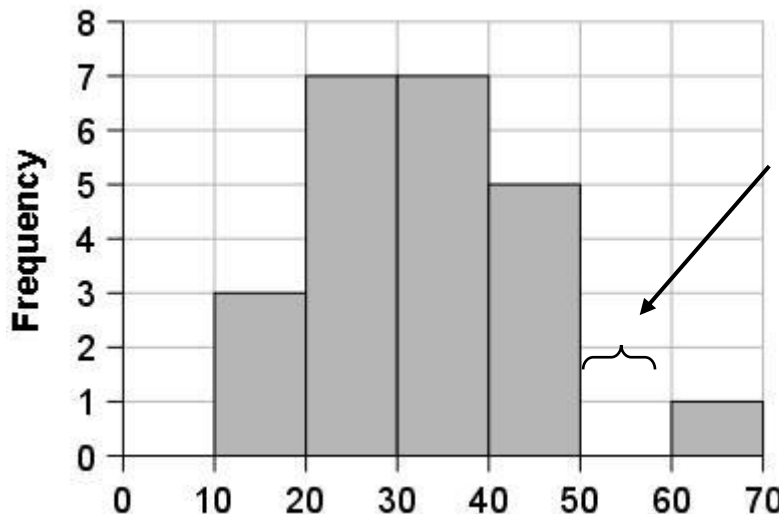
gaps occur when there is a break in the data between bars

Picking the Best Measure of Center

Mean: _____

Median: _____

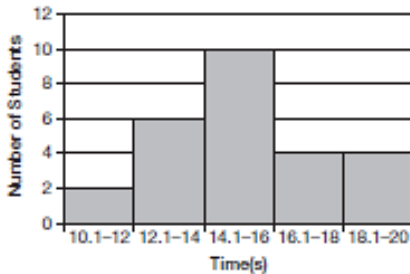## Picking the Best Measure of Spread

Interquartile Range: _____

Mean Absolute Deviation: _____

Histogram Practice

Practice 1: Describe the distribution of each histogram and if the mean is less, greater, or equal to the median. Then describe which would be a better measure of center; the median or mean.
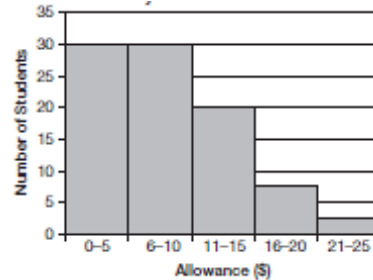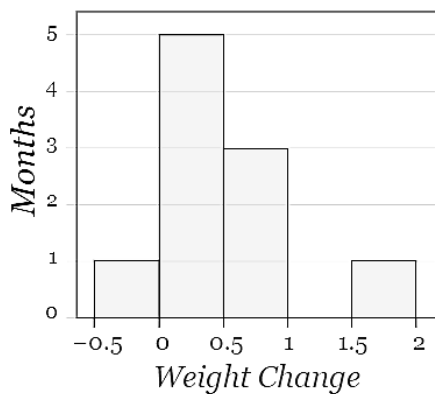
a.

b.

c.

Practice 2: Use the histogram to answer the following questions about how the weight of puppies changes in 5 months.

a. How many puppies are represented by the histogram?

b. How many puppies did not lose weight?

c. Based on the histogram, can you determine how many puppies gained 0.75 pounds?

d. How many puppies gained at least 0.5 pounds?

Practice 3: Analyze the given histogram which displays the ACT composite score of several randomly chosen students.

a. Describe the distribution and explain what it means in terms of the problem situation.

ACT Composite Scores
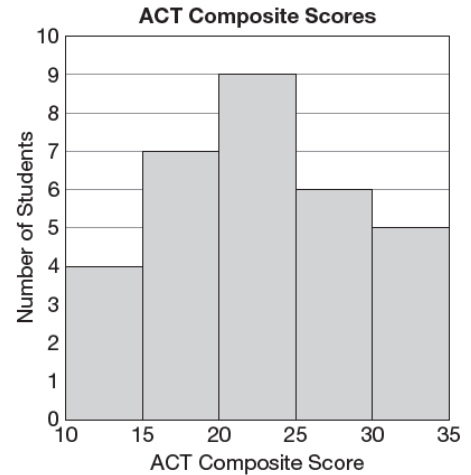
b. How many students had an ACT score of at least 20?

c. How many students had an ACT score less than 30?

d. How many students had an ACT score of exactly 25?

Practice 4: The histogram below shows how much time a sample of freshmen at Harrison High School spend playing video games. Use this graph to answer the following questions.

Hours Spent Playing Video Games on Weekends
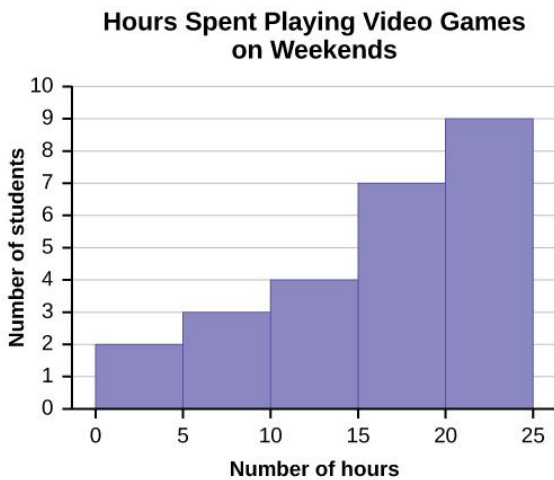
a. What percentage of students played video games for less than 10 hours?

b. How many students played video games for at most 20 hours?

c. How many students played video games for at least 20 hours?

d. How many students played video games between 5 and 15 hours?

e. What is the maximum amount of time spent playing video games by a single student?

f. Would the mean number of hours spent playing video games be greater or less than the median number of hours spent playing video games? Explain how you know. (Hint: you do not need to do any calculations)