

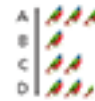
14-1: Data Displays and Measures of Central Tendency

Statistics is the science surrounding _____ most especially _____. There are many ways to display data.

Data Display

What does it do?

Pictograph Shows data using _____



Bar Graph Shows data in specific _____



Circle Graph Shows data as part of a _____



Line Graph Shows how data changes over _____



Histogram Shows frequencies of data values



Stem-and-leaf Plot Orders numerical data and how they are distributed



Box-and-Whisker Plot Shows variability of data using quartiles



Dot Plot Shows the number of times each _____ occurs



Scatter plot Shows the relationship between _____ data sets

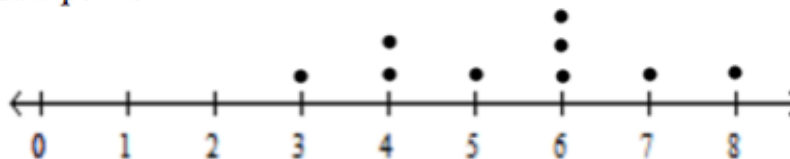


Today we are going to focus on _____ and _____.

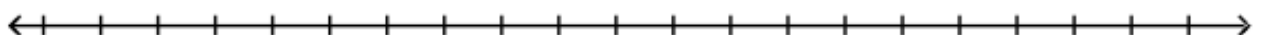
A **Dot Plot** is a graph that shows the _____ (or how often it occurs) of each data

point. Dot Plots are graphed on a number line. The data set 3, 4, 4, 5, 6, 6, 6, 7, 8 would be

graphed as:



Ex. 1 Make a dot plot for the following data set representing test scores for one math class
70, 70, 75, 75, 90, 70, 80, 85, 65, 95, 70, 85, 90, 70, 20, 85, 75, 65, 80, 95



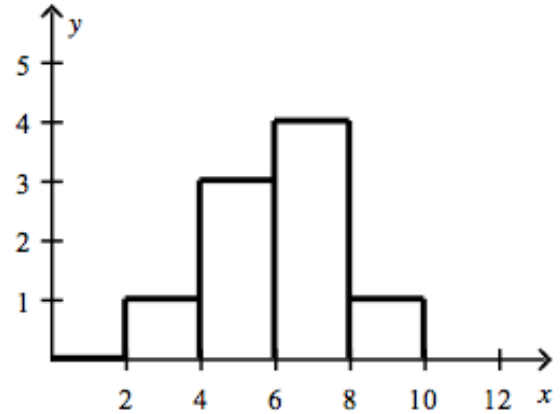
Are there any data points that are very different from all the others? These data points are called _____, because they lie _____ what is normal for the data set.

A **Histogram** is a graph that groups the data points _____, it shows the _____ of the data using intervals of the _____ size.

The data set 3, 4, 4, 5, 6, 6, 6, 7, 8 might be graphed as:

The y-axis tells you how _____ of each there are.

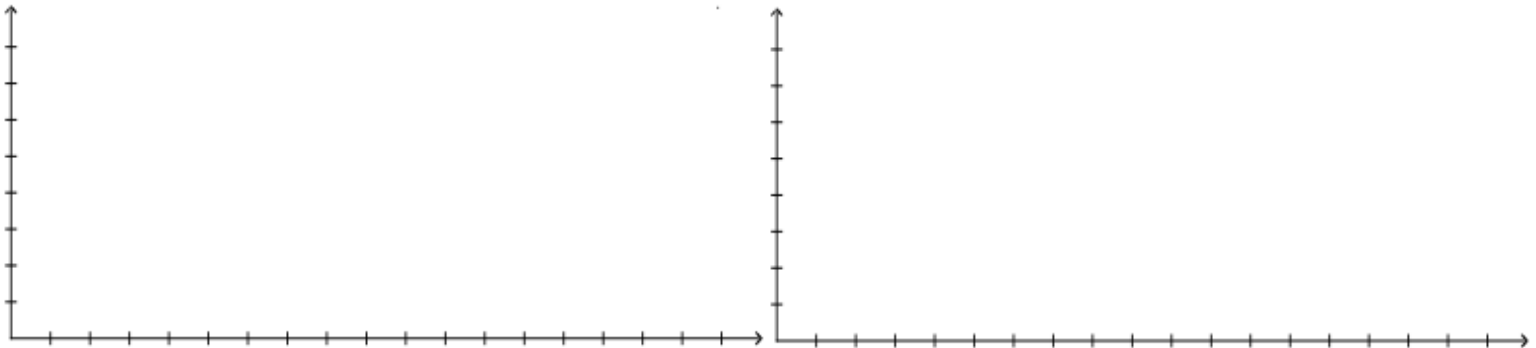
To create a histogram it is important to choose an appropriate scale as the x-axis should be a continuous Number line.



Ex.2 Choose appropriate scaling and make a histogram for the following data sets:

3. The average monthly temperatures in Jacksonville, FL.
52.4, 55.2, 61.1, 67.0, 73.4, 79.1, 81.6, 81.2, 78.1, 69.8, 61.9, 55.1

4. The average monthly temperatures in Austin, TX
48.8, 52.8, 61.5, 69.9, 75.6, 81.3, 84.5, 84.8, 80.2, 71.1, 60.9, 51.6



Do either of the data sets above have outliers? How do the temperatures compare?

A **measure of central tendency** is a _____ that represents the _____ of a data set. The three most common measures are _____, _____, and _____.

Mean: The mean of a data set is the _____ of the data _____ by the number of data values.

Median: To find the median order the data from _____ to _____. For a set with an _____ number of values the median is the _____ value. For a set with an _____ number of values, the median is the mean of the _____ middle values.

Mode: The mode of a data set is the value or values that occur _____ often.

Each one of these is an example of an average. Match which measure goes with which statement:

The average on the test was an 84

The average test score puts you in the middle of the class

The average American student starts college at 18

Ex 3. The salaries of the LA Lakers (who makes more than a million a year) for the 2013-2014 season

Kobe Bryant: \$30,453,805

Pau Gasol: \$19,285,850

Steve Nash: \$9,300,500

Jordan Hill: \$3,563,600

Chris Kaman: \$3,183,000

Jodie Meeks: \$1,550,000

MarShon Brooks: \$1,210,080

Nick Young: \$1,106,942

Jordan Farmar: \$1,106,942

Chris Duhon: \$1,500,000

Mean:

Median:

Mode:

Are there any data points that you would consider outliers? How do they effect the mean?

Take out the outlier and calculate the new mean. How are they different? Is this a better mean?

14-2: Box and Whisker Plots

Today we are going to focus on another type of data display, the Box and Whisker Plot. It is called a box and whisker plot because it has _____ and what look like _____.

In order to use Box-and-Whisker plots we need to learn first a few vocabulary terms.

Quartile: Quartiles mean you have split your data in _____

Range: The spread of your data, found by subtracting _____ - _____ = Range

5-Number Summary: _____ descriptive numbers that represent the data, they are:

Minimum (min): The _____ number in the data set

1st Quartile (Q_1): The median (_____) of the _____ half of the data set

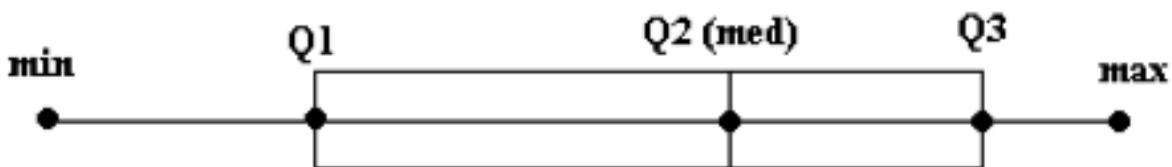
Median (med): The median (_____) of the data set

3rd Quartile (Q_3): The median (_____) of the _____ half of the data set

Maximum(max): The _____ number of the data set

Interquartile Range (IQR): is the range between Q_1 and Q_3 found by subtracting $Q_3 - Q_1$

The five number summary appears on the Box and Whisker Plot like this:



Ex. 1 Find the five number summary for the following data set: Life expectancies in South American males: {59.0, 60.5, 61.5, 66.7, 67.9, 68.5, 69.0, 70.3, 71.4, 71.9, 72.1, 72.6}

Min:

Q_1 :

Median:

Q_3 :

Max:

Find the five number summary for the following data set: Life expectancies in South American females: {66.2, 66.7, 67.7, 72.8, 74.3, 74.4, 74.6, 76.5, 76.6, 78.8, 79.0, 79.4}

Min: Q_1 : Median: Q_3 : Max:

To create a box and whisker plot:

1 - _____ a number line that is appropriately scaled for you data

2- _____ the five-number summary, {min, Q_1 , Median, Q_3 , Max}

3- _____ a rectangular box from Q_1 to Q_3 , with a _____ line at the median

4- _____ line segments (_____) from the max and min to their respective quartiles

Draw a box plot for each of the five number summaries above:

Males:

Females:

Find the range and interquartile range of the above data sets:

Males

Females

Range:

Range:

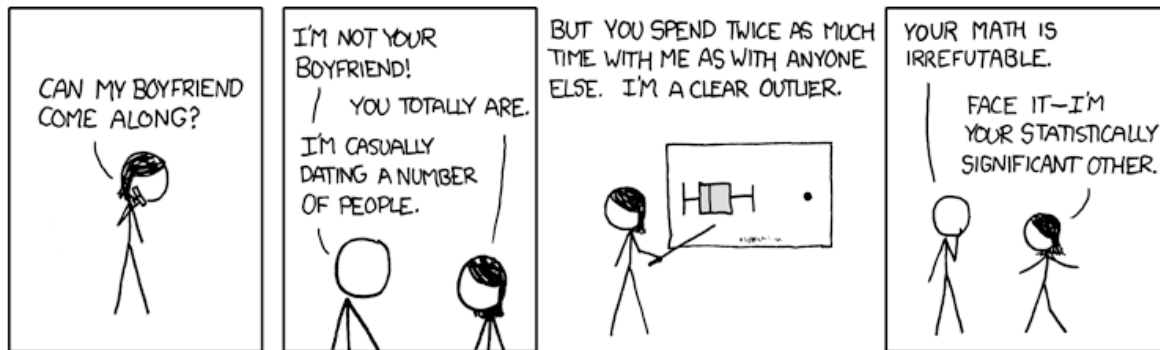
IQR:

IQR:

Is the range heavily affected by a single point? Is the IQR heavily affected by a single point?

Which would be more likely to be affected by an outlier?

Box and whisker plots allow us to get a good visual of outliers: a number that makes one of the _____ noticeably longer than the other. Rule for outliers: A number is considered an **outlier** if it is more than $1.5 \cdot IQR$ _____ Q_1 or _____ Q_3 .



Ex. 2 Below is the five number summary for Roger Maris' home run data. Is 61 an outlier?
 {5, 11, 19.5, 30.5, 61}

What does that mean about his home runs?

The last two things we will talk about are distribution and spread. There are three types of distribution.

Skewed left: Most of the data is on the _____, the left whisker is _____ than the right

Skewed right: Most of the data is on the _____, the right whisker is _____ than the left

Normal/Symmetric: Median is in the _____ middle of the data, whiskers are the _____ length

Spread: Spread has to do with the _____ and _____, is one box bigger than the other?

Label the following according to their distribution and spread:



All types of data displays have distribution and spread, it is about looking from the center of the data and determining where most of your data is.

14-3: Two-Way Frequency Tables

Statistics are often used to describe categorical data. Categorical data is data collected regarding _____ instead of specific numbers. A **Two-way frequency** table displays _____ categories of data collected from the _____ source. Each entry in the table is called a _____.

When we talk about two-way tables we often talk about two ideas being **correlated**. Correlation is the relationship between _____ variables. If it seems like as one variable _____ the other _____ then the two variables are _____ correlated. If it seems like as one variable _____ the other _____ then there is a _____ correlation. If there seems to be no relationship at all then they are _____ correlated.

Reading a two-way table:

		Student		Total
		Studied	Did Not Study	
Grade	Passed	21	2	23
	Failed	1	6	7
Total		22	8	30

Annotations:

- 22 students studied.
- 8 students did not study.
- 23 students passed.
- 7 students failed.
- 30 students were surveyed.

The _____ of the rows and columns are called _____ frequencies.

Ex. 1

You randomly survey students in a cafeteria about their plans for a football game and a school dance. The two-way table shows your results.

- How many students will attend the dance but not the football game?
- Find and interpret the marginal frequencies for the survey.

		Football Game	
		Attend	Not Attend
Dance	Attend	35	5
	Not Attend	16	20

Making a two way table: Is there a relationship between preferred hand and preferred eye? Are people who consider themselves “left handed” more likely to be “left eyed”?

				Total
	Total			

Which is more likely for our class data, someone to be left-handed or right-handed? How do you know?

Which is more likely for our class data, someone to be left-eyed or right-eyed? How do you know?

Which is more likely for the left-handed people, to be left-eyes or right-eyed? How do you know?

Which is more likely for the right-handed people, to be left-eyes or right-eyed? How do you know?

Do you think there is a correlation between preferred hand and preferred eye?

We can also use two-way frequency tables to answer questions regarding an entire population. Do you think there is a correlation between using conditioner and being female?

		Conditioner use		
		Use conditioner	Don't use conditioner	
Gender	Male	15	23	Total
	Female	14	1	
	Total			

Fill in the totals above. To determine percentages based off of a two way table we will need to decide which to values the question wants us to compare. Then to find a percent we simply divide the two.

1- What percent of the group is male?

Keywords: group and male, which items in our table represent the amount of males and the amount in the group?

Percent:

2- What percent of the group is male that used conditioner?

Keywords: Percent:

3- What percent of the group uses conditioner?

Keywords: Percent:

4- What percent of the females use conditioner?

Keywords: Percent:

5- Which group has a higher percentage using conditioner?

Keywords: Percent:

6- Are there more males or more females that use conditioner?

7- Do you think there is a correlation between conditioner use and being female?

14-4: Standard Deviation

The **standard deviation** of a data set is the average distance each point is from the mean. This tells you how spread your data is.

We find standard deviation on our calculators:

- 1- Hit the STAT button, under the edit menu select 1:Edit... (should be what the menu opens you to)
- 2- Clear L1 by putting your cursor on the L1 and selecting CLEAR, then enter your data into L1
- 3- Click on the STAT button again, this time use your cursor to go to the right to the CALC menu, then hit ENTER (this selects 1-Var STATS)
- 4- This will take you back to the main screen with 1-Var Stats on your screen. Then hit ENTER, Standard deviation is the $\sigma_x =$

Find the standard deviation of 70, 70, 75, 75, 90, 70, 80, 85, 65, 95, 70, 85, 90, 70, 20

$\sigma_x =$

Secondary 1

Name: _____

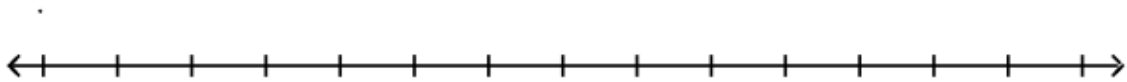
In- Class

14-1 Data displays and Measures of Central Tendency

Make dot plots of the following data sets:

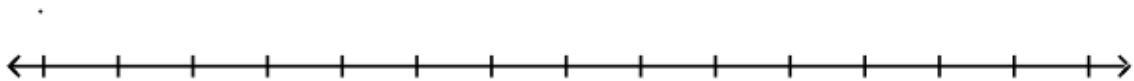
1. Ages of kids at a party:

13, 14, 14, 13, 15, 15, 15, 16, 14, 7, 16, 19, 17, 15, 17, 15, 15, 14, 18



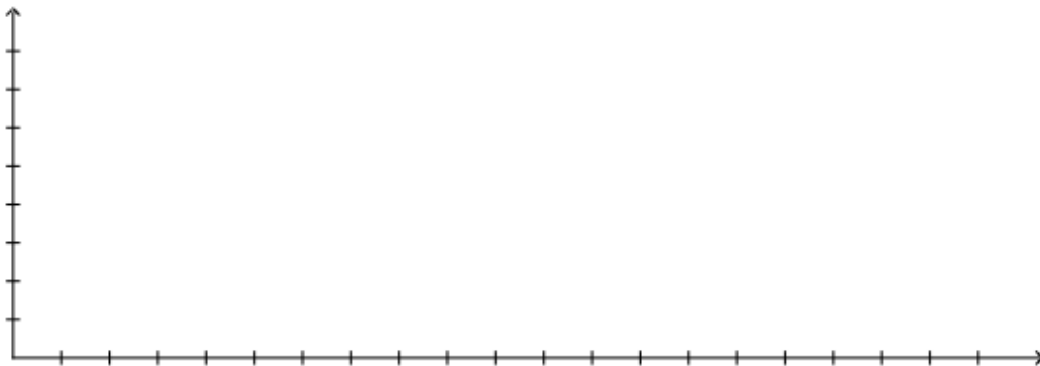
2. Ages of students in a college class:

18, 21, 24, 22, 19, 23, 18, 19, 20, 17, 23, 19, 18, 21, 22, 19, 19, 20, 21



3. Do either of the sets have outliers? Comparing the two sets how to the ages of the kids at the party compare to the ages of the students in the college class?

4. Make a histogram of the following data: Number of degrees over 100 degrees in various states: 12, 10, 7, 16, 20, 0, 18, 12, 8, 13, 27, 17, 14, 10, 20



5. U.S. Presidents have entered into office at many different ages. The table below displays the number of presidents inaugurated within different age groups. Use the data to create a histogram.

Age at Inauguration	40–44	45–49	50–54	55–59	60–64	65–69
U.S. Presidents	2	7	13	12	7	3

6. Label the following as mean, median or mode:

The average test score was a 78.

Her average of her long and short program put her in the middle of the rankings.

The average American Female is 5'6".

Find the mean, median and mode for the following:

7. prices in dollars of smartphones: 311, 309, 312, 314, 399, 312

8. attendance at an event for the last nine years: 68, 99, 73, 65, 67, 62, 80, 81, 83

9. books a student checks out of the library: 17, 9, 10, 17, 18, 5, 2

10. Find the mean of the data, then remove the outlier and find the new mean. How do they differ? Which one is a better measure of the center? Repair costs at 5mph:

Audi A6: 0 BMW 328i: 0 Cadillac Catera: 900 Jaguar X: 1254 Lexus ES300: 234

Lexus IS300: 979 Mercedes C320: 707 Saab 9-5: 670 Volvo S60: 769

Volvo S80:4194

Secondary 1

Name: _____

In- Class

14-2 Box and Whisker Plots

Find the five number summary for the following data sets and then construct a box-plot

1. prices in dollars of smartphones: 311, 309, 312, 314, 399, 312

2. attendance at an event for the last nine years: 68, 99, 73, 65, 67, 62, 80, 81, 83

3. books a student checks out of the library: 17, 9, 10, 17, 18, 5, 2

4. Given the five number summary: {0.3, 1.0, 2.6, 4.5, 17.1} (Salaries of NBA teams in Millions)

Find the range:

Find the IQR:

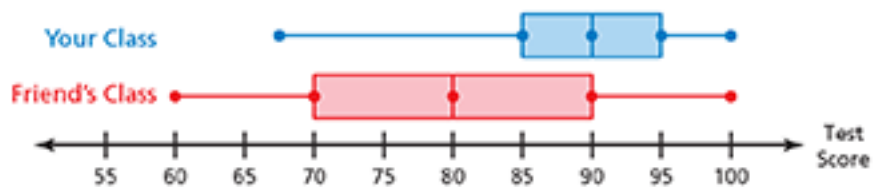
Is 17.1 an outlier?

5. Identify the shape and spread of each distribution

Your Class:

Your friend's class:

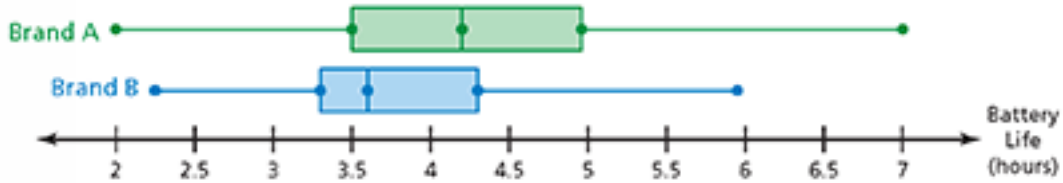
The double box-and-whisker plot represents the test scores for your class and your friend's class.



6.

CELL PHONES The double box-and-whisker plot compares the battery lives (in hours) of two brands of cell phones.

- Identify the shape of each distribution.
- What is the range of the upper 75% of each brand?
- Compare the interquartile ranges of the two data sets.



7/8.

INCHWORM The table shows the lengths of 12 inchworms.

- Make a box-and-whisker plot for the data.
- Find and interpret the range of the data.
- Describe the distribution of the data.
- Find and interpret the interquartile range of the data.



Length (cm)	2.5	2.4	2.3	2.5	2.7	2.1	2.8	2.6	2.1	2.6	2.9	2.0
-------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Determine if the data is skewed right, skewed left, or symmetric

9.



10.



Mean and standard Deviation

1. Calculate the mean and standard deviation of

(a) 14 15 18 20 23 18 (b) 41 45 34 45 46 47 50

2. The costs of a can of diet coke in 6 different shops are

47p 49p 50p 44p 48p 44p

Calculate the mean and standard deviation of these costs.

3. (a) The prices of a bag of sugar in 6 different shops are

86p 88p 84p 79p 81p 86p

Calculate the mean and standard deviation of these prices.

(b) In 6 different shops the same bag of sugar has a mean price of 87 pence and a standard deviation of 5.2 pence.

Make two comparisons between the prices in the two sets of shops.

4. (a) The marks of 7 pupils in an advanced higher maths exam were

77 67 43 90 66 93 75

Calculate the mean and standard deviation of these marks.

(b) Another group of 7 pupils who sat the same exam had a mean of 78 and a standard deviation of 3.2.

Make two comparisons of the marks of the two groups.

5. A gardener grows tomatoes in his greenhouse.

The temperature of the greenhouse, in degrees Celsius, is recorded every day at noon for one week.

18 21 24 17 23 14 16

(a) Calculate the mean and standard deviation of these temperatures.

For best growth the mean temperature should be $(20 \pm 5)^\circ\text{C}$ and the standard deviation should be less than 5°C .

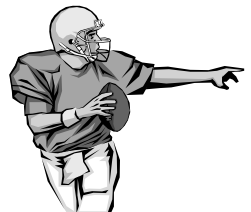
(b) Are the conditions in the greenhouse likely to result in best growth?



6. The number of points scored by an American football team over 7 matches were

34 26 20 23 21 18 26

Calculate the mean and standard deviation of these scores.



7. (a) The number of pupils in 7 third year classes in a secondary school are

25 24 28 22 24 30 22

Calculate the mean and standard deviation of the class sizes.

(b) In the same school the mean and standard deviation of the number of pupils in 7 fourth year classes are 22 and 4.4 respectively.

Make two comparisons between the class sizes in third year and in fourth year.

8. Scientists are studying the differences between crocodiles and alligators.

(a) The lengths of 6 crocodiles are recorded in feet.
The results are shown below.

18.2 23 17.3 22 20.8 18.1

Calculate the mean and standard deviation of these lengths.

(b) The lengths of 6 alligators are recorded. The results give a mean of 16.8 feet and a standard deviation of 1.85 feet.

Make two valid comparisons between the lengths of the crocodiles and the alligators

