Describing Data Patterns in Statistics

The **distribution** of a set of data refers to the pattern with which the observations are arranged.

Center: shown through measures of central tendency.

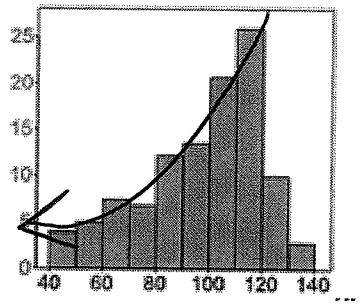
mean
↕ 3
median → resistent to outliers

Spread: The **spread** of a distribution refers to the variability of the data. If the observations cover a wide range, the spread is larger. If the observations are clustered around a single value, the spread is smaller.

range, standard deviation, and IQR

Shape:

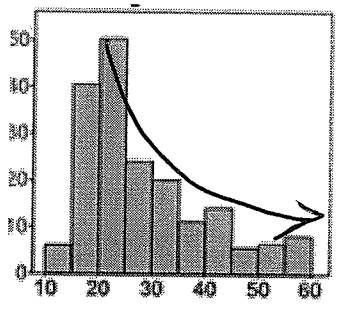
Data can be "distributed" (spread out) in different ways. It can be spread out...



Skewed left

mean < median < mode

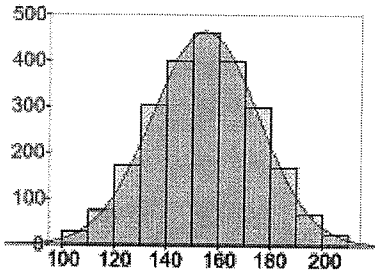
Example:
Age of retirement can be skewed left because most people retire in their mid 60's or older, however there are a few people who retire at a much younger age.



Skewed right

mode < median < mean

Example:
Incomes are skewed to the right because even just a few individuals who earn millions of dollars can greatly affect the mean, and there are no negative incomes.



Or symmetrical

We will talk about this more...

Histograms

Example:

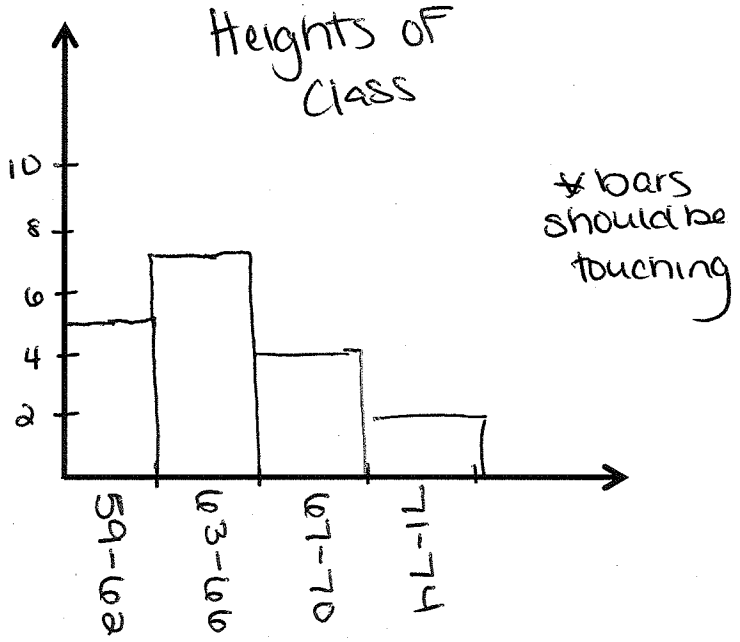
As a class we will determine the height of each student in inches. Make sure you have all of the data written down below.

60, 60, 61, 62, 61, 63, 63, 64, 65, 65, 65, 64, 68, 68, 69
70, 71, 72

Next we need to sort the data into a frequency table:

Interval	Frequency
55-58in	
59-62in	5
63-66in	7
67-70in	4
71-74in	2
75-78in	
79-82in	

Lastly, we will create a histogram with the given data.



Calculate the following statistics:

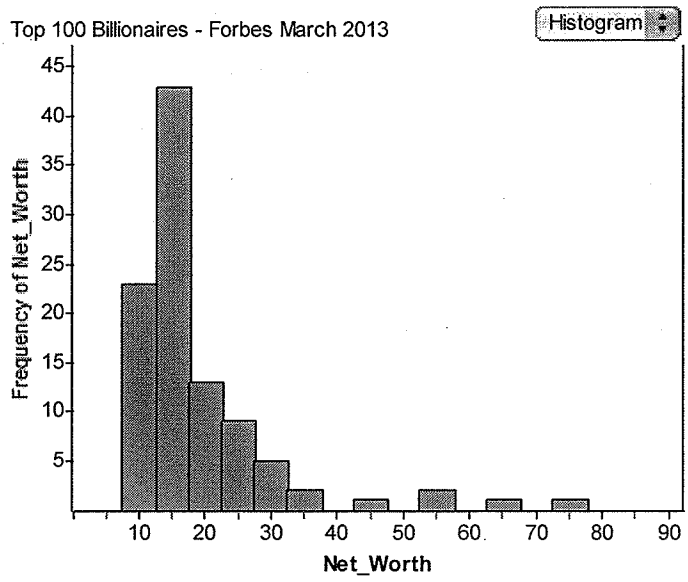
Mean: 65.1 Median: 64.5 Range: 12 Standard Deviation: 3.8

Describe the class's data in terms of shape, center, and spread.

The distribution of heights are slightly skewed to the right with only a few students taller than 70 inches. The center is 64.5 in tall with a range of 12 inches.

Example:

Forbes, on their website www.forbes.com, posted a list of the world's top 100 billionaires based on their net worth. The graph below shows the distribution of net worth in billions of dollars for these 100 billionaires. The mean net worth for these top 100 billionaires is $18.9 billion. The median is $15.3 billion and the standard deviation is $11.0 billion.



a. Describe the distribution of the top 100 billionaires' net worth. Be sure to discuss shape, center and spread.

The shape of net worth is skewed right with a couple of billionaires having a much greater net worth. The center is $15.3 billion with a range of $11 billion.

b. What is the best measure of center for the distribution of net worth for these top 100 billionaires? Explain your reasoning.

median bc data is skewed and contains outliers

c. Carlos Slim Helu & Family from Mexico is listed as the billionaire with the top net worth ($73.0 billion). If Senior Helu's data was removed from this data set, which would be most affected: the mean or the median? Explain.

mean would decrease

d. What impact would removing Senior Helu's data have on the standard deviation? Explain.

standard deviation would decrease