

0-3 Data Distributions

S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

A *measure of central tendency* describes the center of a set of data.

Measures of central tendency include the *mean*, *median*, and *mode*.

MEAN - The Average of the data values.
"X"
EX. 100, 88, 85, 95, 98, 93, 35, 90, 95 $\frac{779}{9} = 86.6$

MEDIAN - The MIDDLE value of the data values. The values MUST be put in numerical order. If there are an even number of values then the median is the mean of the 2 middle numbers.

EX. 100, 88, 85, 95, 98, 93, 35, 90, 95
35, 85, 88, 90, 93, 95, 98, 100

MODE - The values that occur MOST OFTEN. A data set may have one mode or more than one mode. If no value occurs more often than another, the data set has No mode.

EX. 100, 88, 85, 95, 98, 93, 35, 90, 95

RANGE - The DIFFERENCE between the GREATEST and LEAST values in the set.

EX. 100, 88, 85, 95, 98, 93, 35, 90, 95

$$100 - 35 = 65$$

Find the Mean, Median, Mode, and Range of a Data Set

The numbers of hours Isaac did homework on six days are 3, 8, 4, 6, 5, and 4. Find the mean, median, mode, and range of the data set.

Mean: $\frac{30}{6} = 5$

Median: 3, 4, 4, 5, 6, 8 $\frac{4+5}{2} = 4.5$

Mode: 4

Range: $8 - 3 = 5$

Determining the effects of outliers

Identify the outlier in the data set {7, 10, 54, 9, 12, 8, 5}, and determine how the outlier affects the mean, median, mode, and range of the data.

With OUTLIER	Without OUTLIER
Outlier: 54	Outlier: 54 (Pull this out)
Mean: $\frac{105}{7} = 15$	Mean: $\frac{105-54}{6} = \frac{51}{6} = 8.5$
Median: 5, 7, 8, 9, 10, 12, 54	Median: 5, 7, 8, 9, 10, 12
Mode: N 😞	Mode: N 😞
Range: $54 - 5 = 49$	Range: $12 - 5 = 7$

Stu Dent scored 70, 74, 72, 71, 73 and 96 on his six geometry tests. Find the mean, median, or mode.

1. Which measure gives Stu's test average?

Mean

2. Which measure best describes Stu's typical score? Explain.

Median. The \bar{x} is affected by the outlier

3. What is the mode?

N/A

Mean: $\frac{456}{6} = 76$

Median: 70, 71, 72, 73, 74, 96 $\frac{72+73}{2} = 72.5$

Mode: N 😞

Range: $96 - 70 = 26$

Measures of central tendency describe how data cluster around one value. Another way to describe a data set is by its spread -- how the ~~data~~ *data* values are spread out from the center.

QUARTILES - divide a data set into 4 equal parts. Each quartile contains $\frac{1}{4}$ of the values in the set. A quartile is a NUMBER NOT a range.

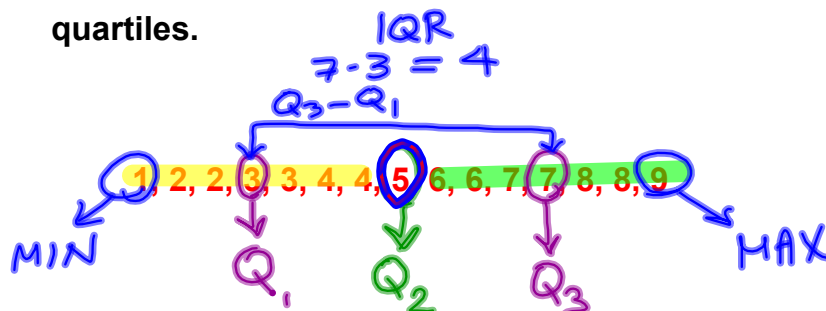
1st QUARTILE - The MEDIAN of the LOWER HALF of the data set.

2nd QUARTILE - The MEDIAN of the data set.

3rd QUARTILE - The MEDIAN if the UPPER HALF of the data set.

INTERQUARTILE

RANGE (IQR) - The DIFFERENCE between the 3RD and 1ST quartiles.



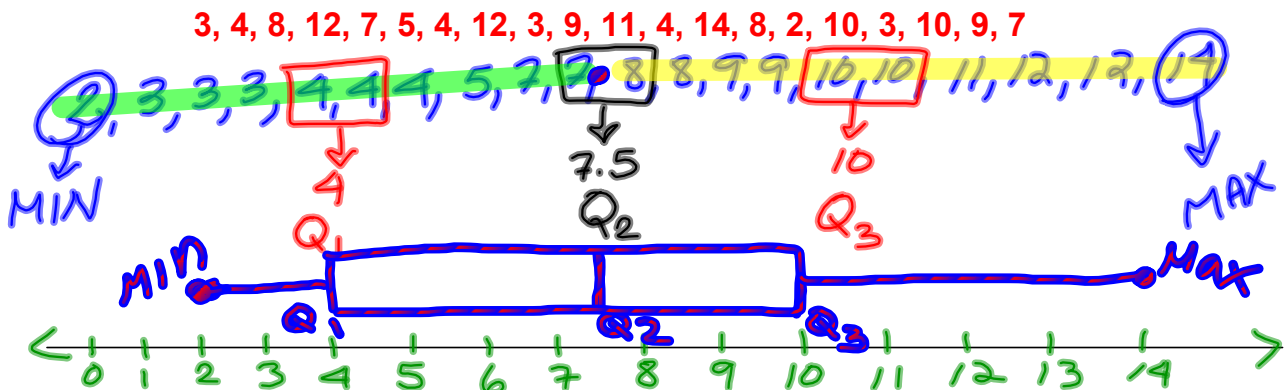
BOX-AND-WHISKER PLOT - Shows how the values in a data set are "Box Plot" DISTRIBUTED.

To create a Box-and-Whisker plot you will need the **5 STATISTICAL SUMMARIES**:

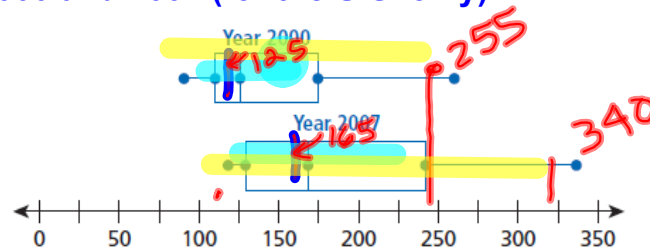
1. Minimum
2. 1st Quartile
3. Median = Q_2
4. 3rd Quartile
5. Maximum

The numbers of runs scored by a softball team in 20 games are given.

Use the data to make a box-and-whisker plot.



The box-and-whisker plots show the ticket sales, in millions of dollars, of the top 25 movies in 2000 and 2007 (for the U.S. only)



1. Which data set has a greater median? Explain.

2007 The Q_2 is to the right of 2000's Q_2

2. Which data set has a greater interquartile range? Explain.

2007 Box is bigger

3. About how much more were the ticket sales for the top movie than for the top movie in 2000?

$$340 - 255 = 85$$

4. Which data set has a smaller range? Explain.

2000: Max & Min are closer

5. About how much more was the median ticket sales for the top 25 movies in 2007 than in 2000?

$$165 - 125 = 40$$