Statistics Vocab Notes

Unit 4

Mean

The average value of a data set, found by adding all values and dividing by the number of data points

Example:

5 + 4 + 2 + 6 + 3 = 20

$$\frac{20}{5} = 4$$

The Mean is 4

Median

The middle-most value of a data set; 50% of the data is less than this value, and 50% is greater than it

Example: Put #'s in order

2,3,4,5,6

Median (Middle #) is 4

Range

The difference between the highest and lowest numbers in the data set.

Example: 11, 12, 15, 17, 22, 23, 29

29 – 11 = 18

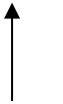
The range is 18

Outlier

A data value that is much greater than or much less than the rest of the data in a data set.

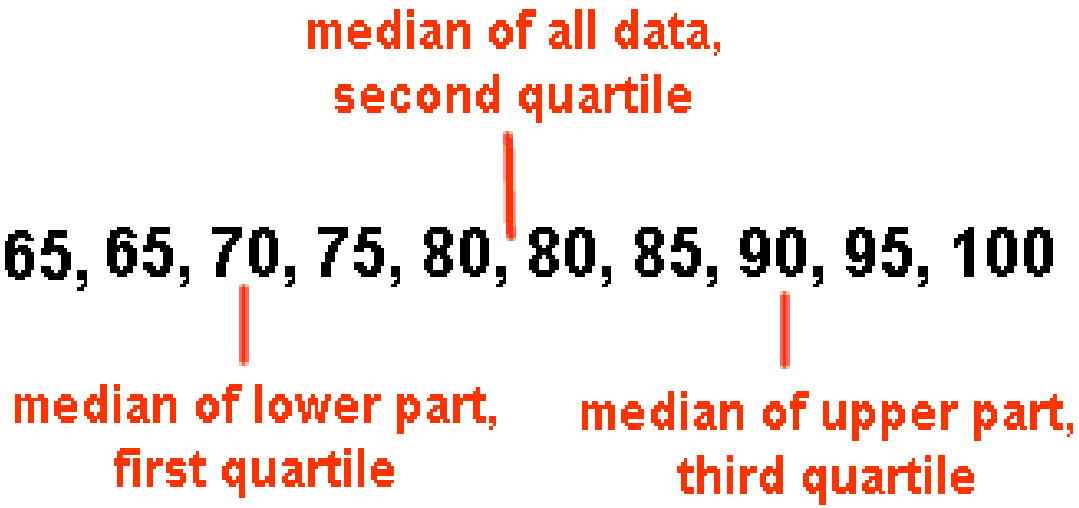
Example:

164, 175, 126, 135, 55, 159, 143



First Quartile

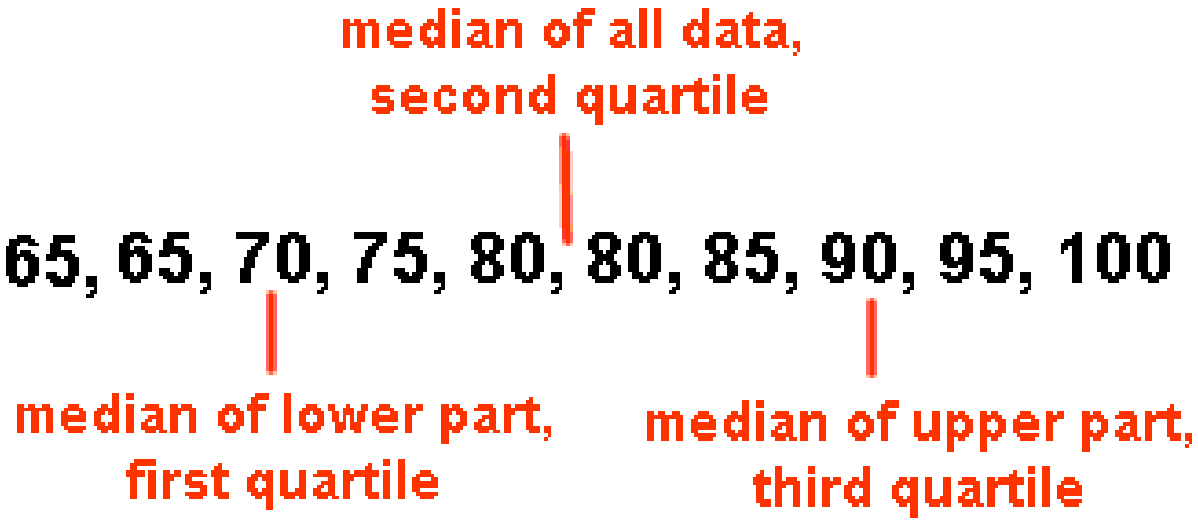
The value that identifies the lower 25% of the data; the median of the lower half of the data set; written as Q_1



Third Quartile

Value that identifies the upper 25% of the data; the median of the upper half of the data set; 75% of all data is less than this value; written as Q_3

Example



Interquartile Range

The difference between the third and first quartiles; 50% of the data is contained within this range

Q1 Q3

Example: 11, 12, 15, 17, 22, 23, 29

Subtract Third Quartile - First Quartile = IQR

23 – 12 = 11

IQR = 11

Interquartile Range

The numbers below represent the number of homeruns hit by players of the North Paulding baseball team.

2, 3, 5, 7, 8, 10,	14, 18, 19, 21, 25, 28
$Q_1 = 6$	$Q_3 = 20$

$$IQR = Q_3 - Q_1$$

Interquartile Range: 20 – 6 = 14

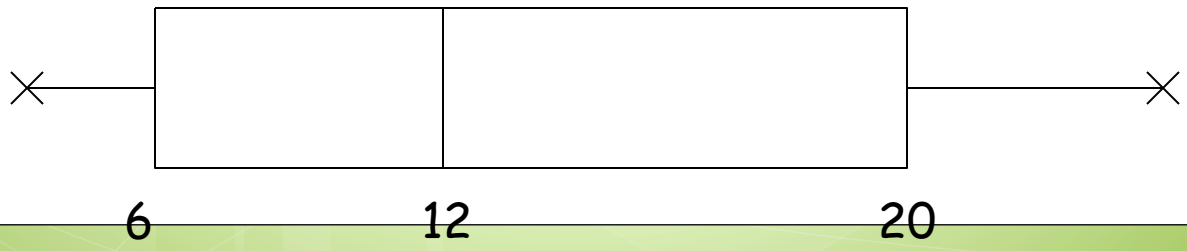
Box and Whisker Plot

The numbers below represent the number of homeruns hit by players of the Hillgrove baseball team.

2, 3, 5, 7, 8, 10, | 14, 18, 19, 21, 25, 28

$Q_1 = 6$ $Q_3 = 20$

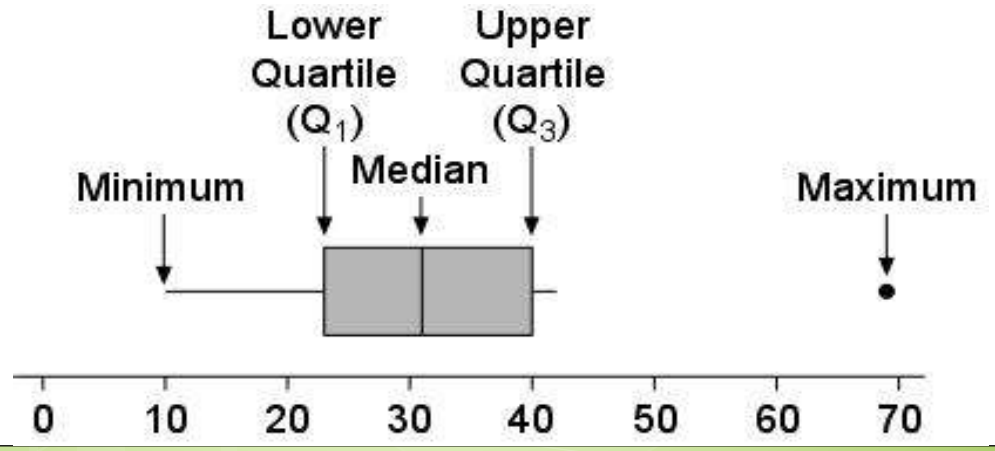
Interquartile Range: 20 – 6 = 14



Box Plot

A plot showing the minimum, maximum, first quartile, median, and third quartile of a data set; the middle 50% of the data is indicated by a box.

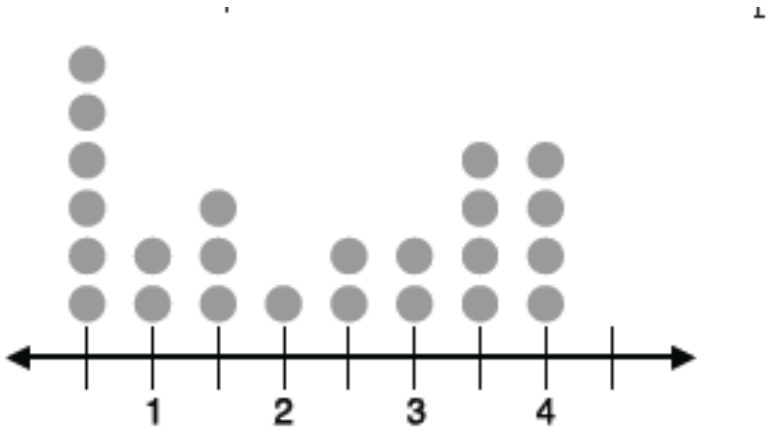
Example:



Dot Plot

A frequency plot that shows the number of times a response occurred in a data set, where each data value is represented by a dot.

Example:



Histogram

A frequency plot that shows the number of times a response or range of responses occurred in a data set.

Example: