VARIABLES

CATEGORICAL

Places individuals in categories

Can not find the average

QUANTITATIVE

Numerical values

Makes sense to find the average

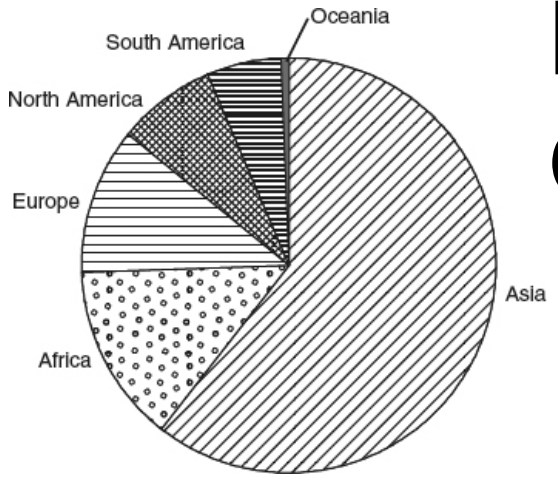
Graphs

For Categorical Variables

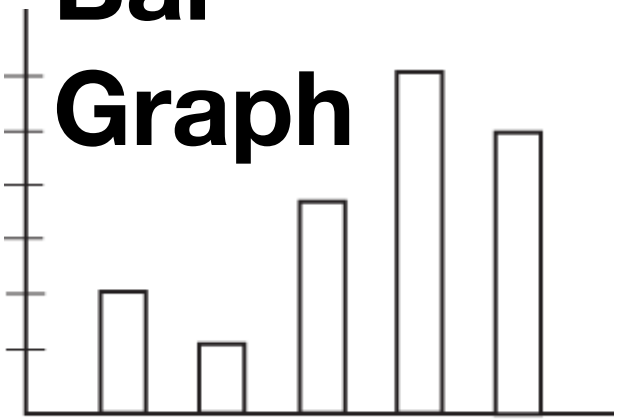
Two-Way Table

What is your favorite sport to watch on television?			
	Football	Basketball	Baseball
Males	40	22	15
Females	12	16	45
Total	52	38	60

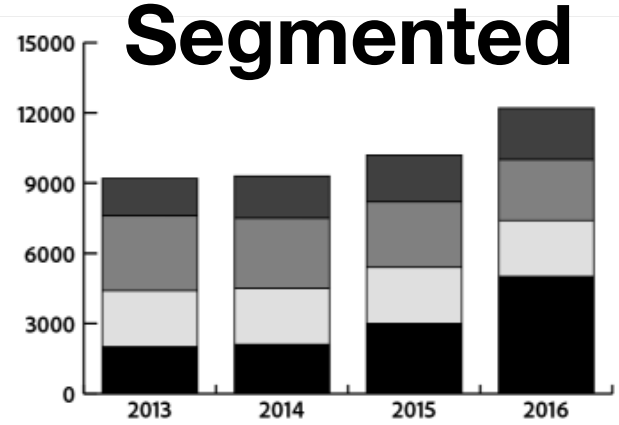
Pie Chart



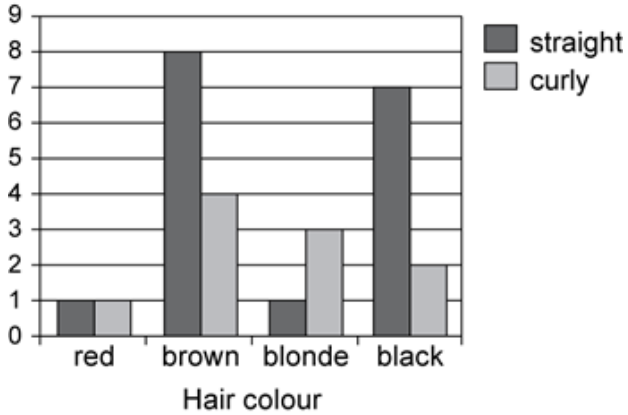
Bar Graph



Segmented



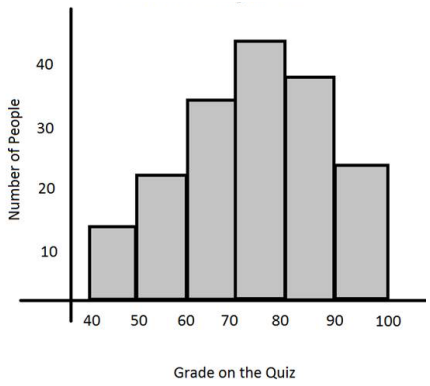
Side-by-side



Graphs

for quantitative variables

Histogram



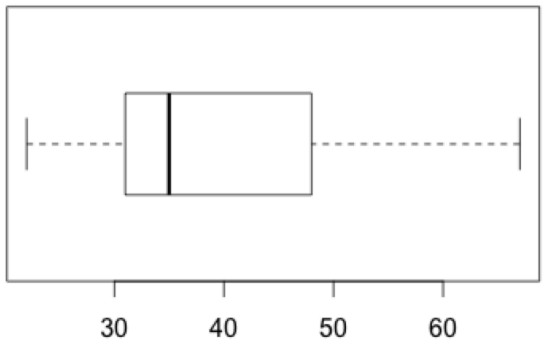
Stemplot

Stem	Leaves
9	0388
8	023355788
7	02222577788
6	05
5	3778
4	5

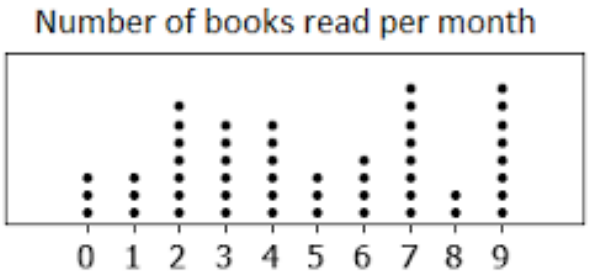
Back-to-Back Stemplot

Male		Female
5, 2, 0	1	5, 8
5, 1	2	1, 6, 9, 9
5, 5, 5, 3, 1	3	
5, 2	4	1, 2, 6, 8
9, 8, 6, 1, 1	5	5
6, 5, 5, 0	6	0, 1
2, 1, 1, 0, 0	7	2

Boxplot

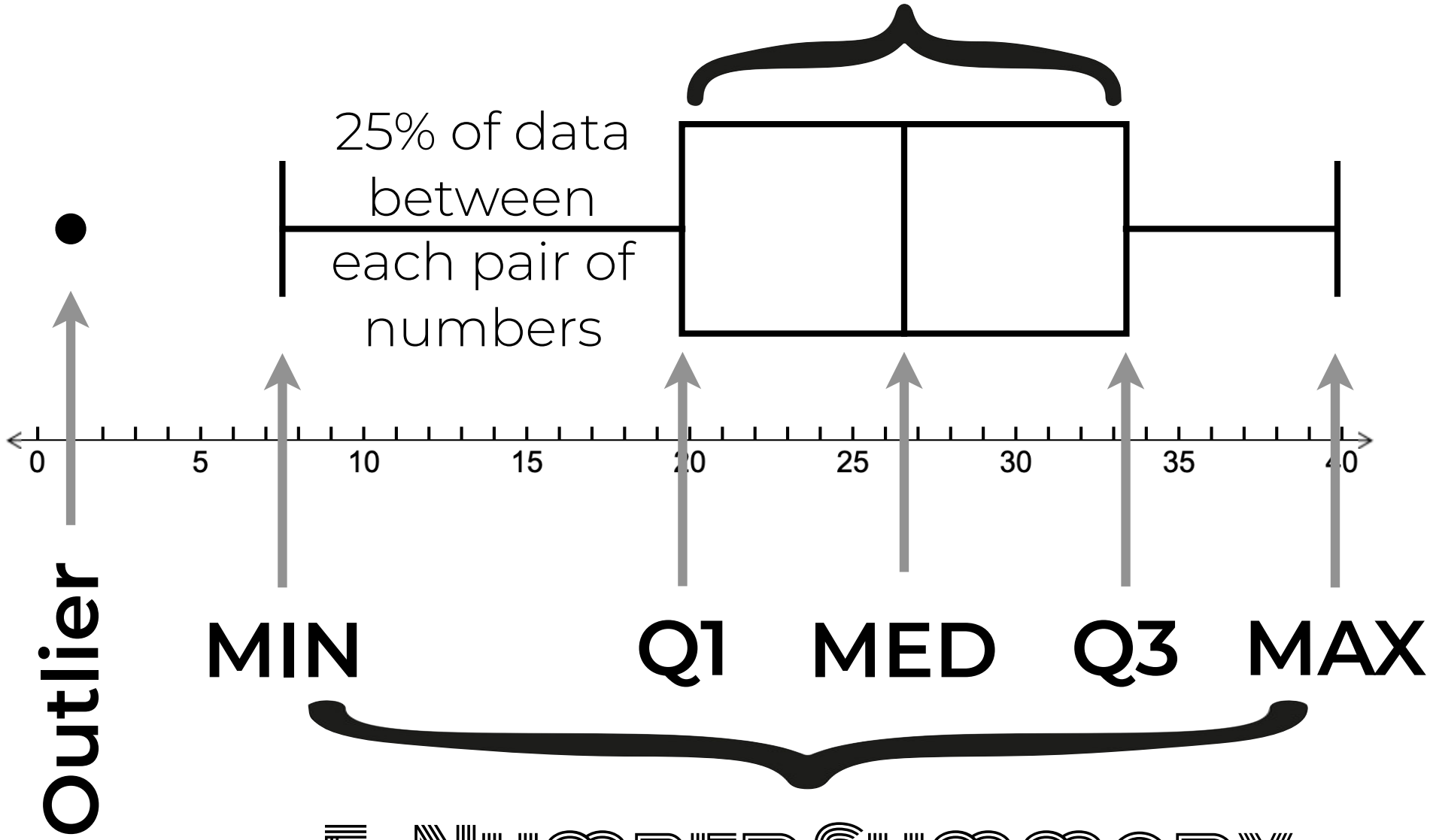


Dotplot



Boxplot

IQR (Middle 50% of data)



5-Number Summary

Two-Way Tables

	Men	Women	**Total**
Dogs	A	A	B
Cats	A	A	B
Total	B	B	**C**

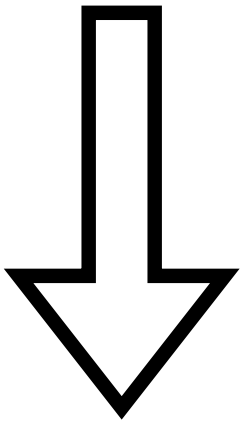
Frequency
= COUNT

Relative Frequency
= %

Conditional Distribution : $\frac{A}{B}$

Marginal Distribution : $\frac{B}{C}$

Independent

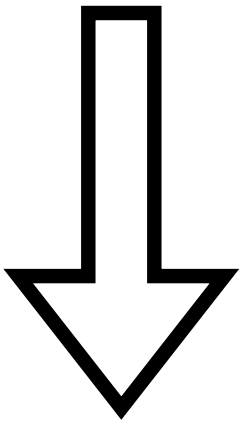


NO

Association

NOT

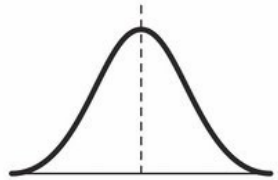
Independent



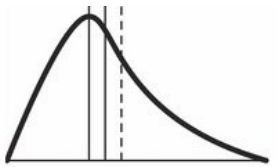
Association

DESCRIBING Distributions

SHAPE

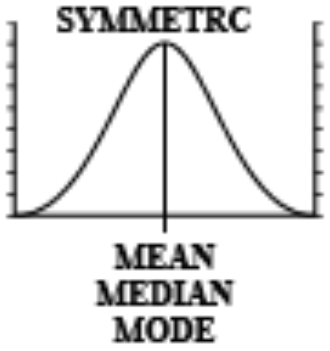


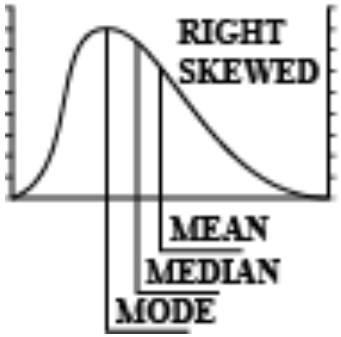
Symmetric



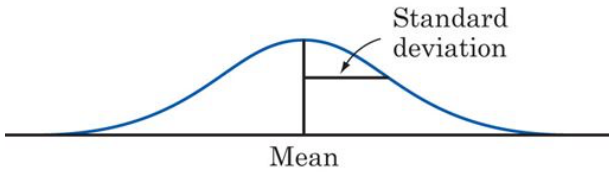
Skewed

CENTER

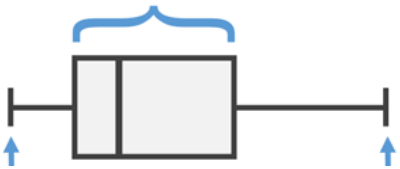




SPREAD



Standard Deviation



Interquartile Range (IQR)

OUTLIERS

Q1 — (IQR x 1.5)

Q3 + (IQR x 1.5)

Don't forget context!!!

COMPARING DISTRIBUTIONS

Use COMPARATIVE words:

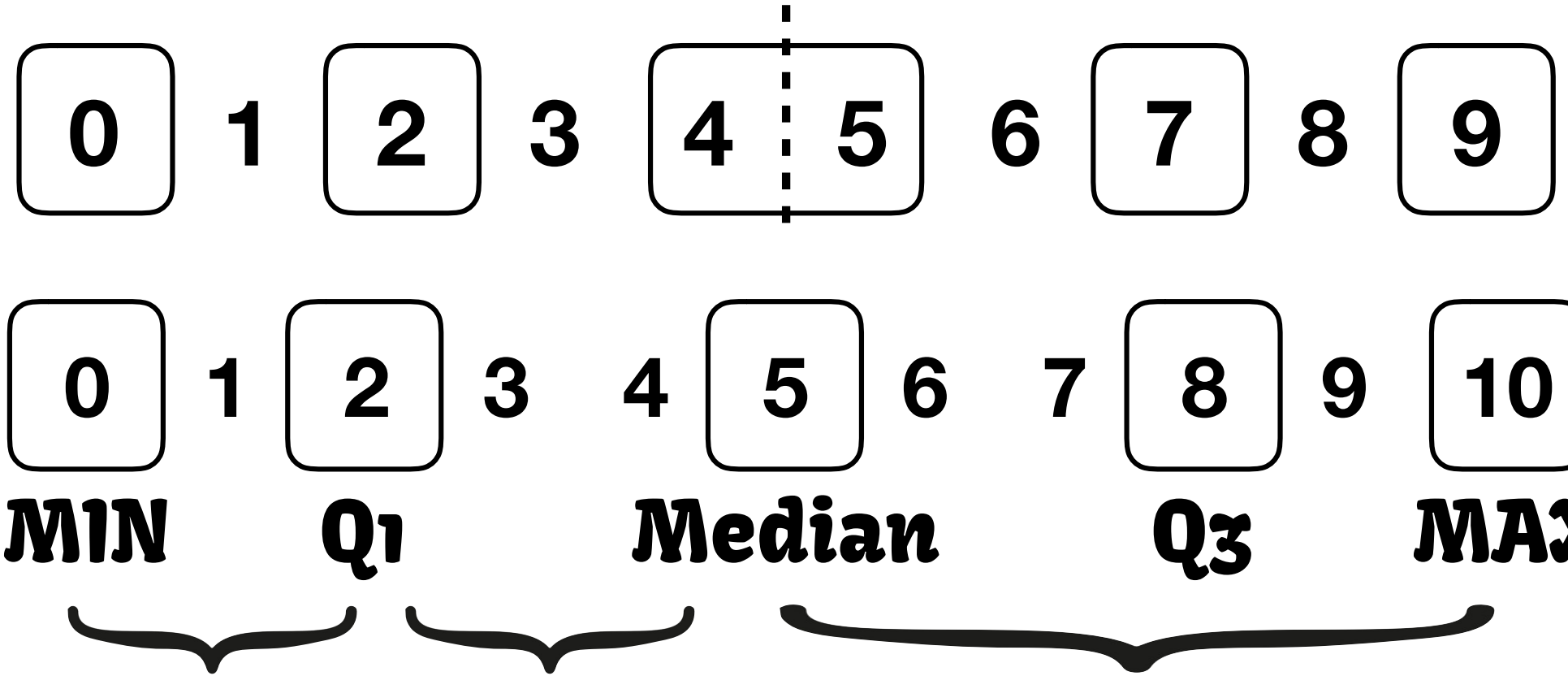
The shape of the two distributions is **SIMILAR**.

IQR is **GREATER THAN** that of the other distribution.

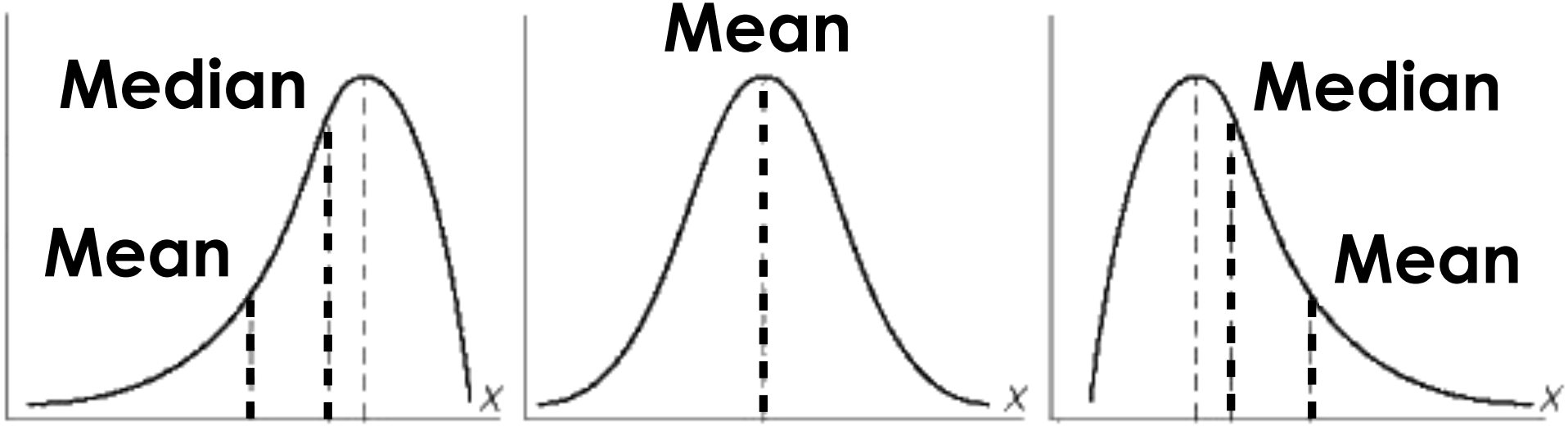
The mean is **LESS THAN** that of the other distribution.

Don't just list! Include context!

Five Number Summary



MEASURES OF CENTER



SKEWED LEFT SYMMETRIC SKEWED RIGHT

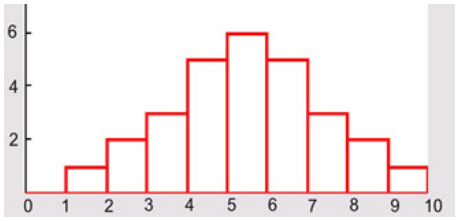
Median is resistant... Mean is NOT.

Mean is "pulled" by a tail MORE THAN the median.

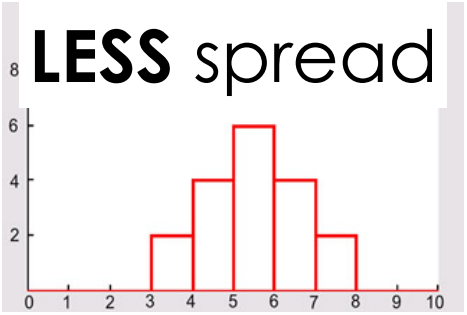
Measures of Spread

(a.k.a. variability)

MORE spread



LESS spread



Standard Deviation

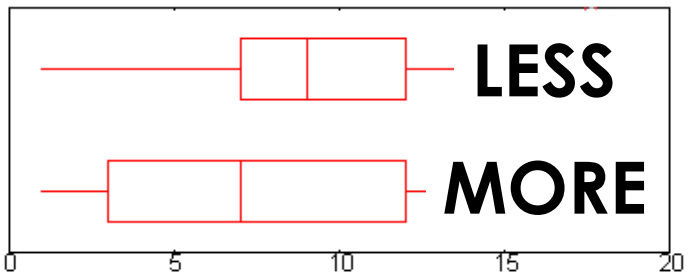
$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

Average distance from the mean

Interquartile Range (IQR)

$$IQR = Q3 - Q1$$

Middle 50% of data



Variance = (St. Dev)^2

St. Dev = $\sqrt{\textbf{Variance}}$

IQR is resistant... Standard deviation is NOT.