Measures of Center (part 2)

Gaston Sanchez

1 Frequency Tables

2 Charts & Graphics

3 Numeric Summaries

Summary Measures

measures of center

looking for a "middle" value

measures of spread

looking for a value that reflects the amount of variation

Measures of Center

Centers and shapes of Distributions

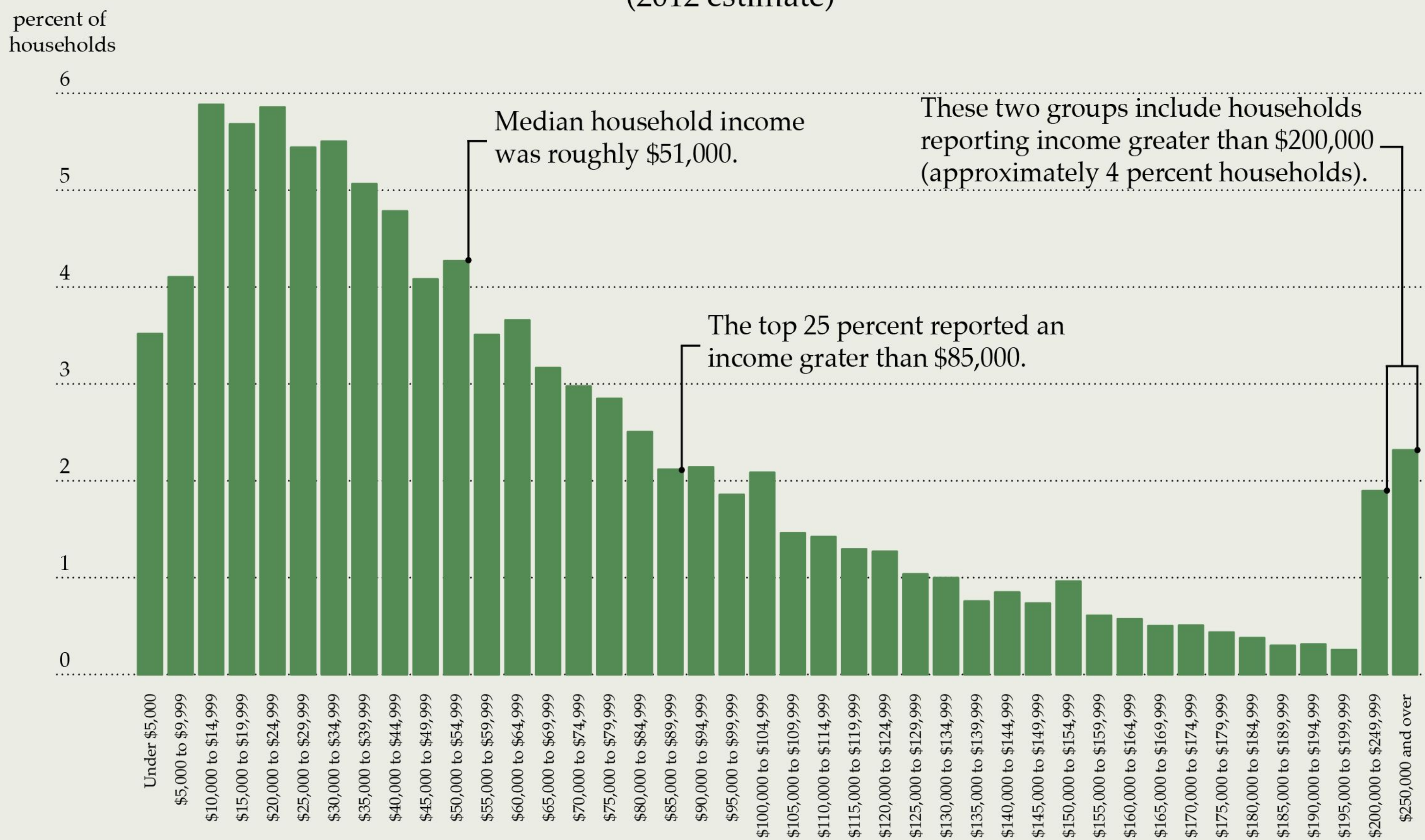
Measures of center

median

average

When do you use one or the other?

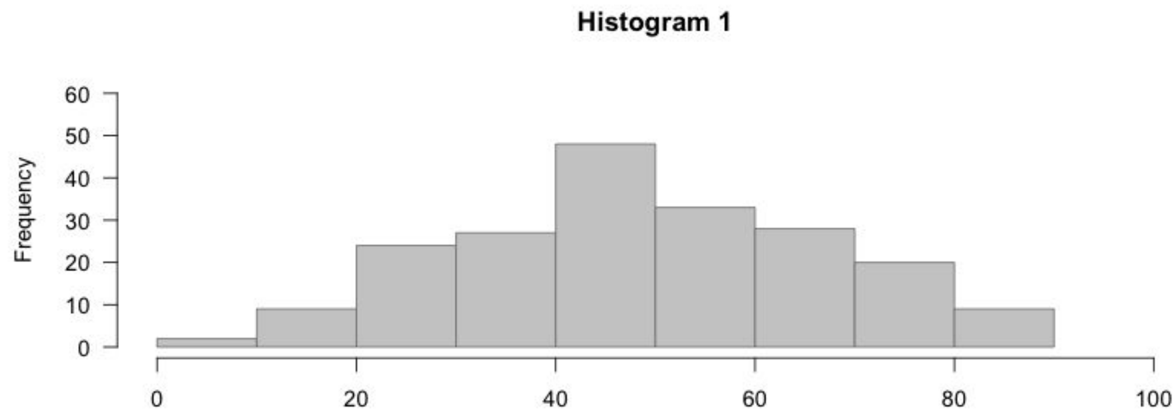
Distribution of annual household income in the United States
(2012 estimate)



Source: U.S. Census Bureau, Current Population Survey, 2012 Annual Social and Economic Supplement

Author: vikjam

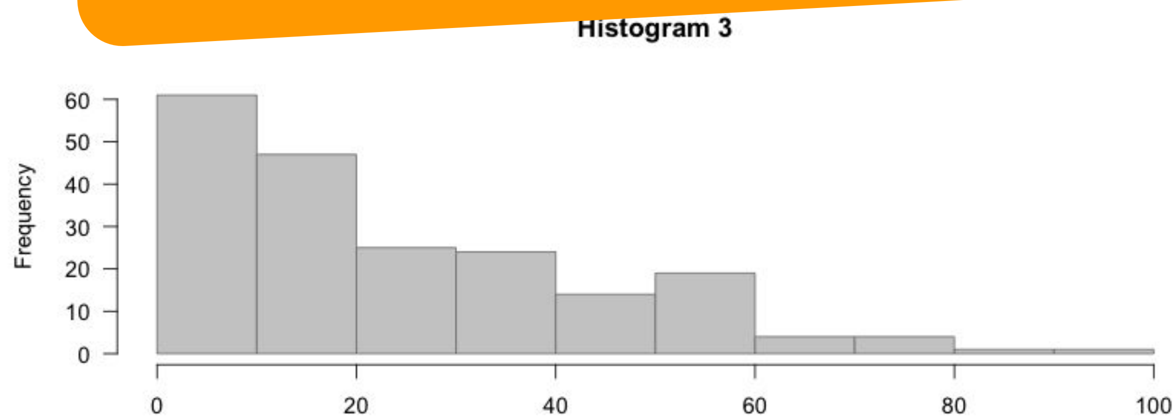
symmetric



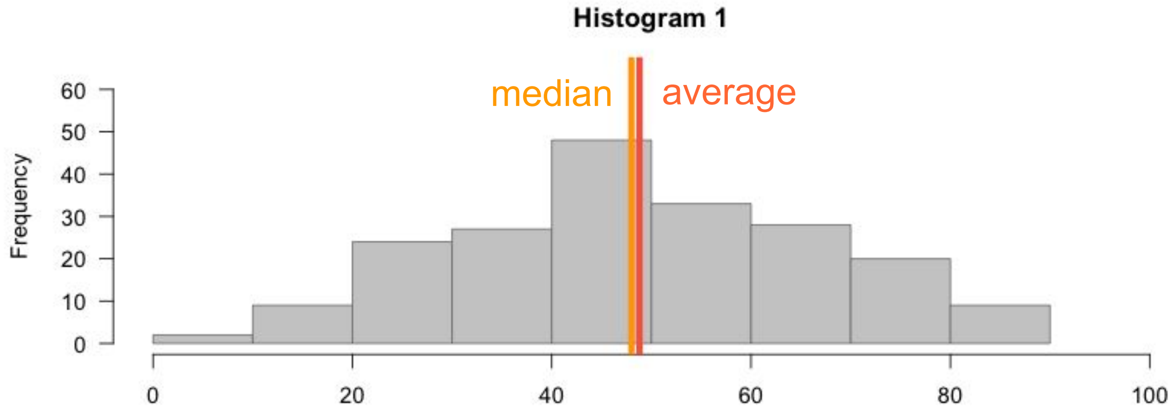
skewed
left

where are the median and average?

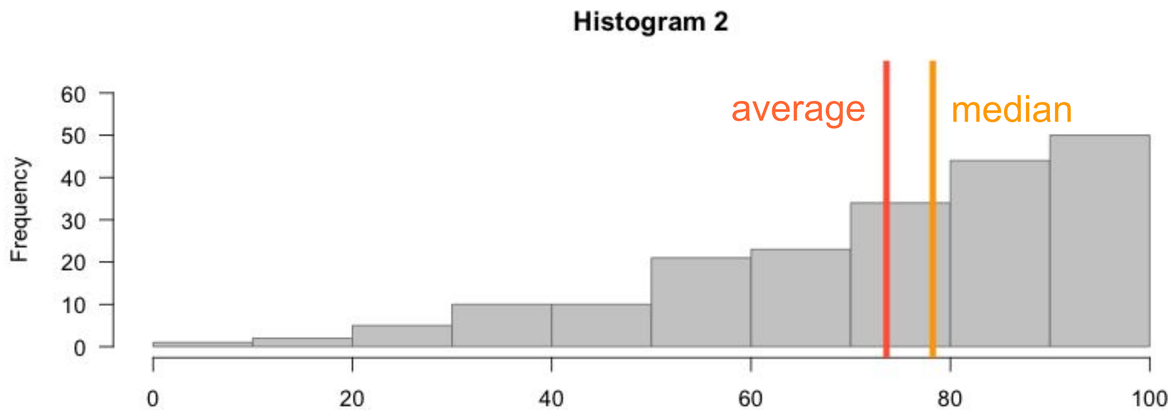
skewed
right



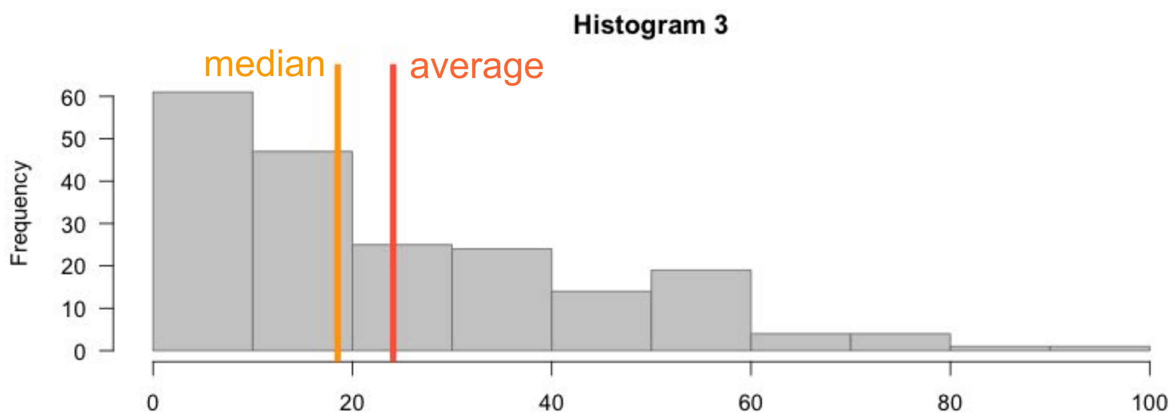
symmetric



skewed
left



skewed
right



Average -vs- Median

When the shape of the distribution is symmetric, the average and the median are very similar

When the distribution is skewed, the average is further out in the tail than the median

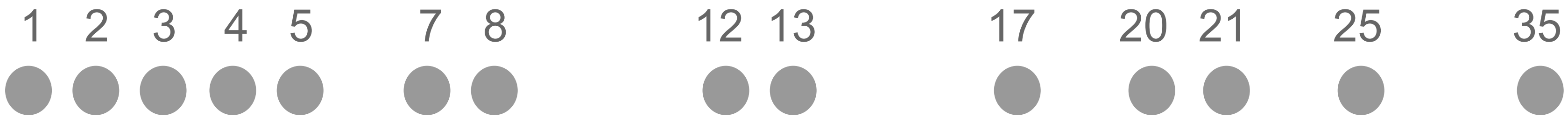
The median is much less sensitive to extreme values

Average

-vs-

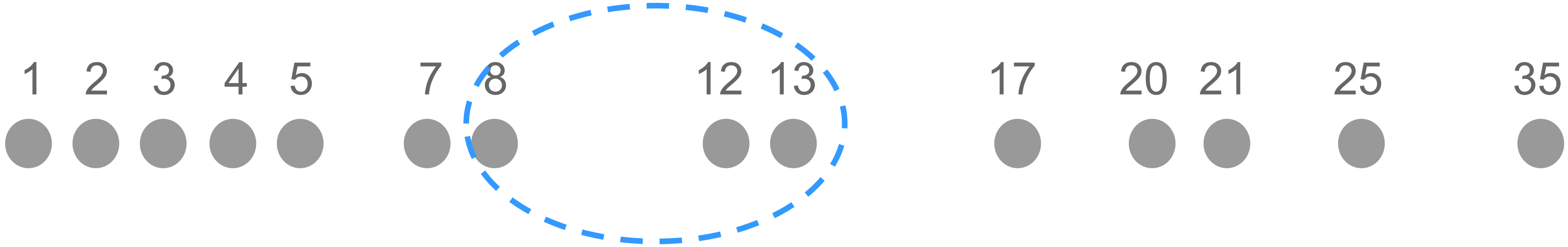
Median

Consider the following list of numbers



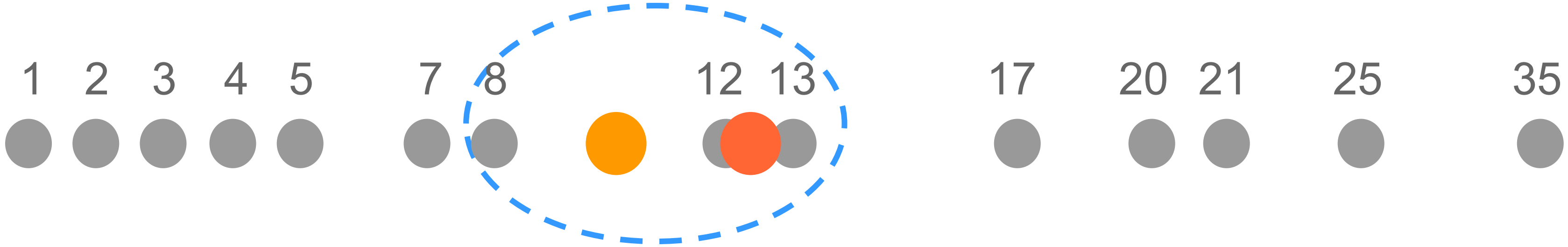
Median -vs- Average

Where's the center?



somewhere here

Median -vs- Average

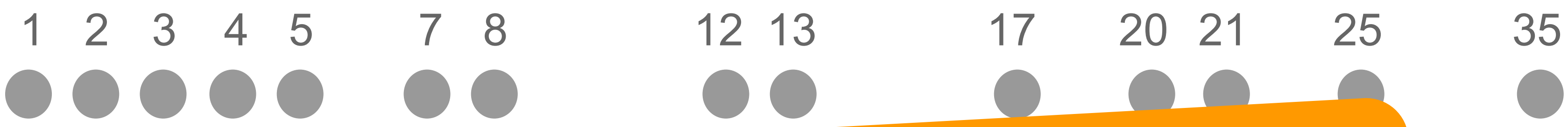


median = (8 + 12) / 2 = 10

average = 12.35

What does the **median** tell us?

Median -vs- Average



let's "ignore" the magnitudes for a minute

Median -vs- Average

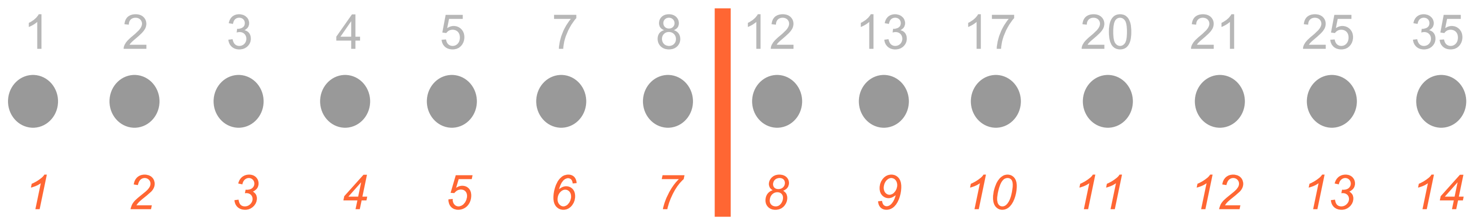
order values from smallest to largest

1 2 3 4 5 7 8 12 13 17 20 21 25 35

let’s focus on the positions

Median -vs- Average

data values are secondary



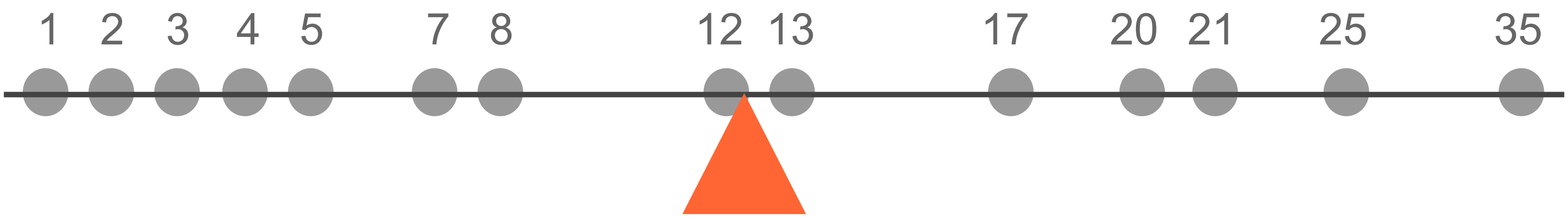
Median

What does the **average** tell us?

Median -vs- Average

let's "focus" on the magnitudes

Median -vs- Average

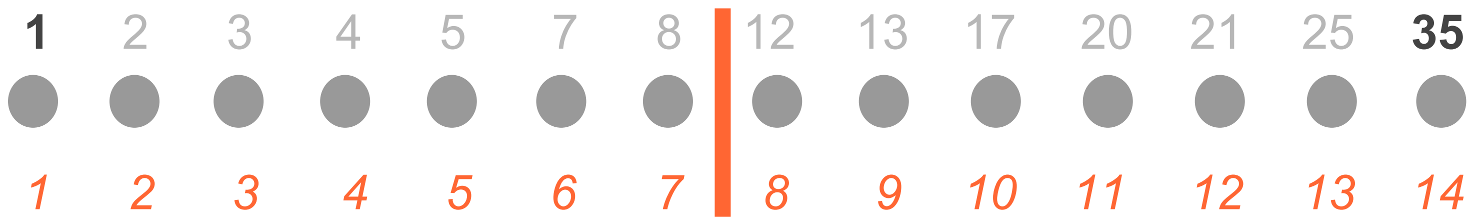


average is the balancing point

Implications?

About the median

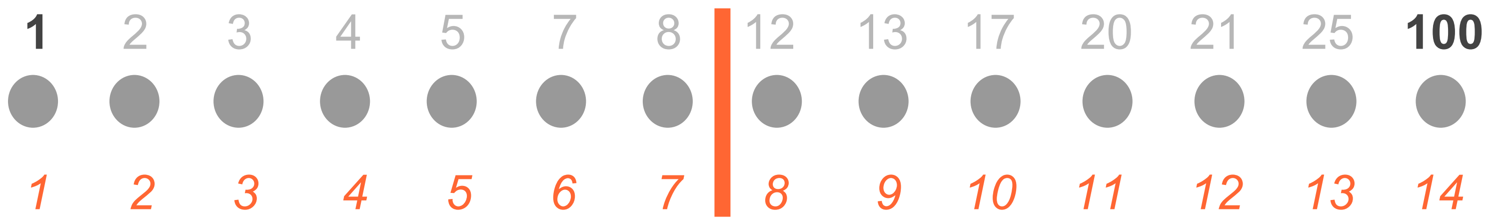
extreme values are secondary



Median

About the median

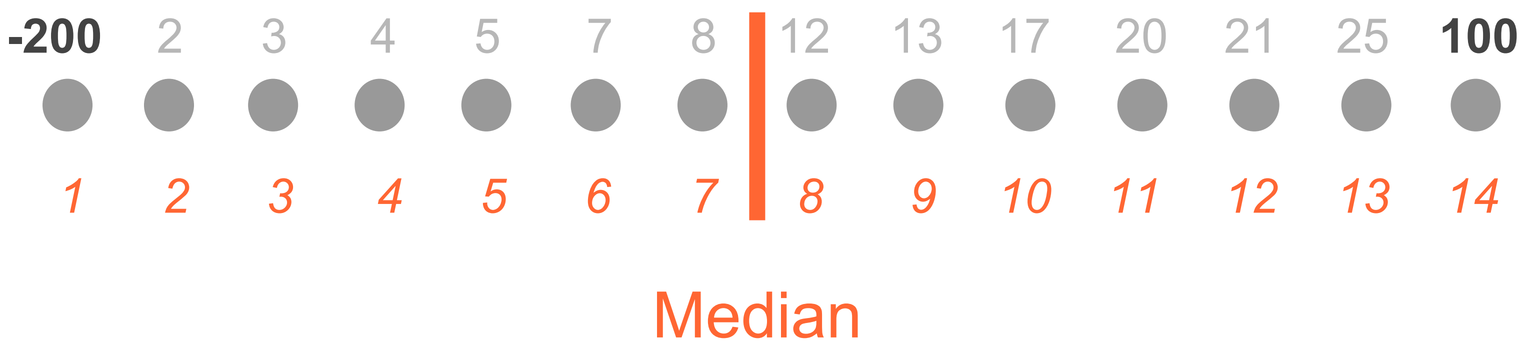
extreme values are secondary



Median

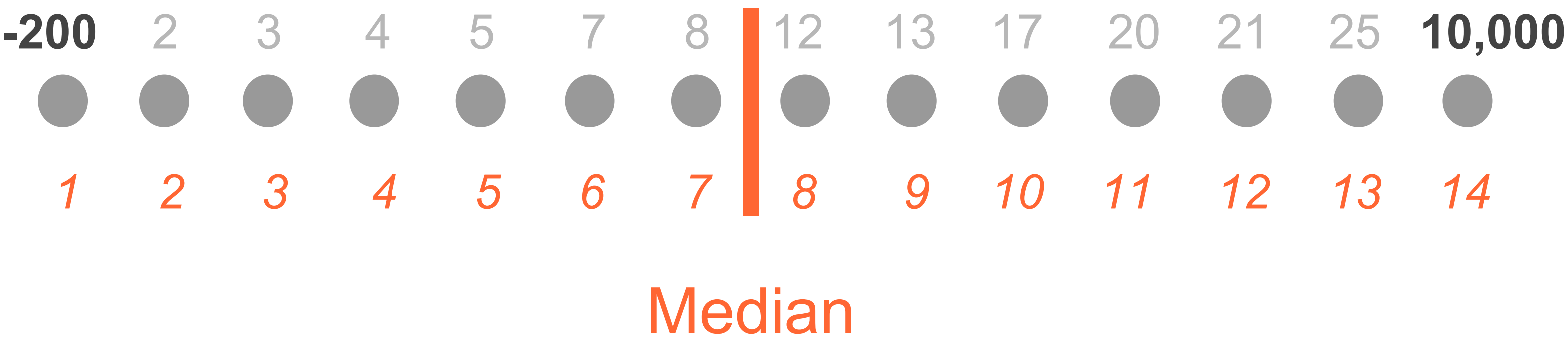
About the median

extreme values are secondary



About the median

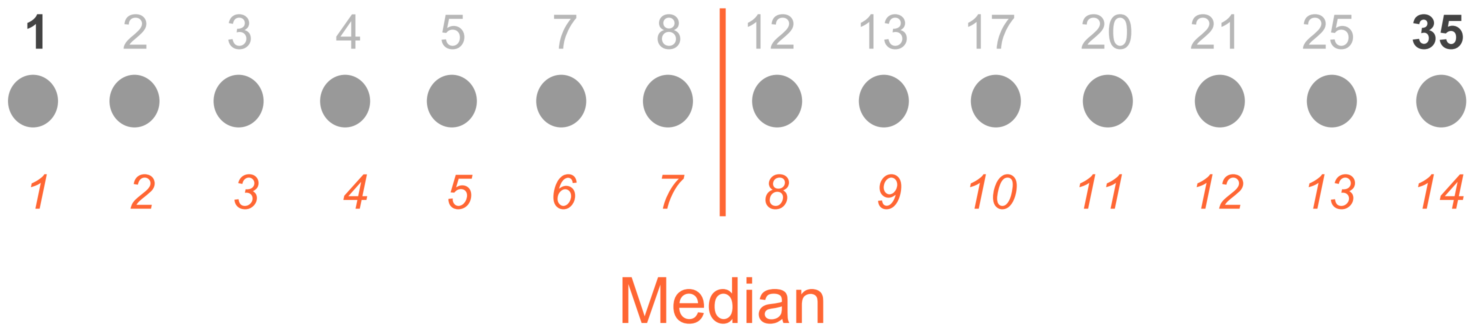
extreme values are secondary



How do extreme values affect the average?

About the average

extreme values affect the average



Average = 12.35

About the average

extreme values affect the average

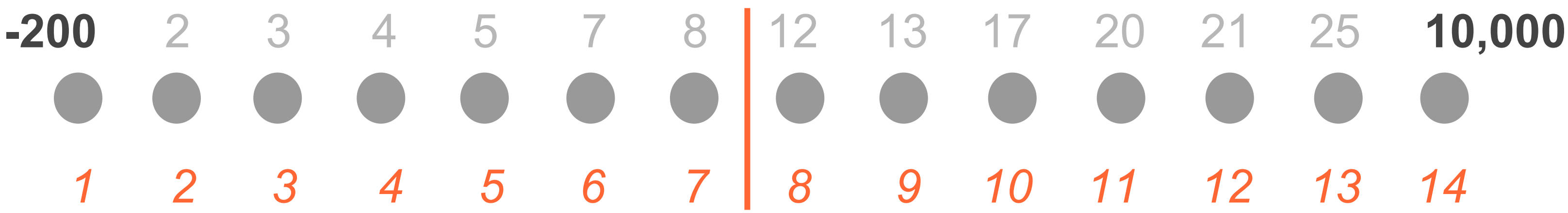
1	2	3	4	5	7	8	12	13	17	20	21	25	**100**
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*

Median

Average = 17

About the average

extreme values affect the average



Median

Average = 709.78

More about the Average


Should we scare the opposition by announcing our mean height or lull them by announcing our median height?

Harvard Salary Survey

In 1998, Harvard conducted a salary survey of entering class of 1973

They were interested in determining a typical salary 25 years after first entering Harvard

They found:

- Average: $750,000
- Median: $175,000

"Despite our efforts to improve literacy in the past 3 years, half of our children in America are still reading below the median reading level"

More Examples

Example: frequency of use

Frequency of usage:

- Never
- Sometimes
- Always

Example: frequency of use

Frequency of usage:

- Never = -1
- Sometimes = 0
- Always = 1

Average on qualitative variables

-1	1	-1	1	1	-1
never	always	never	always	always	never

Avg = 0

Average usage frequency: ***sometimes***

Assigning numbers to qualities

Frequency of usage

never = 0

rarely = 1

sometimes = 2

often = 3

always = 4

Assigning numbers to qualities

0	2	1	4	4	3
never	some	rarely	always	always	often

Avg = 2.3

What does it mean?

Transformations: change of scale

Linear Transformation

Consider a list of values:

2, 3, 5, 7, 8 average = 5

What happens if you **add** a constant to the data?

e.g. add 2 to all values

4, 5, 7, 9, 10 average = 7

Adding a constant to the data

$$\text{Avg} = \frac{1}{n} \sum_{i=1}^{n} (x_i + a)$$

$$\text{Avg} = a + \sum_{i=1}^{n} \frac{x_i}{n}$$

Linear Transformation

Consider a list of values:

2, 3, 5, 7, 8 average = 5

What happens if you **multiply** the data with a constant?

e.g. multiply by 2 to all values

4, 6, 10, 14, 16 average = 10

Multiplying the data by a constant

$$\text{Avg} = \frac{1}{n} \sum_{i=1}^{n} (b x_i)$$

$$\text{Avg} = \frac{b}{n} \sum_{i=1}^{n} x_i$$

Linear Transformation

Consider a list of values:

2, 3, 5, 7, 8 average = 5

What happens if you **add** a constant and **multiply** by a constant?

e.g. add 1 to all values, and multiply by 2

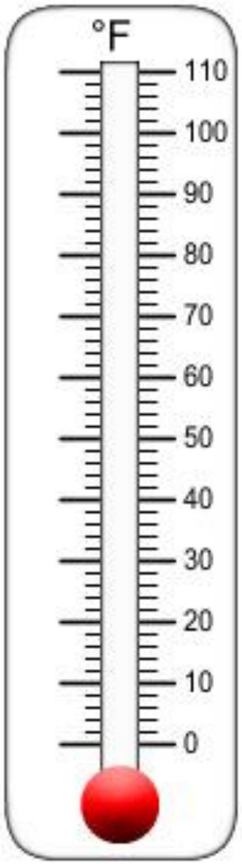
6, 8, 12, 16, 18 average = 12

Adding a constant and multiplying

$$\text{Avg} = \frac{1}{n} \sum_{i=1}^{n} (bx_i + a)$$

$$\text{Avg} = a + \frac{b}{n} \sum_{i=1}^{n} x_i$$

Converting Temperature from ^oC to ^oF



conversion formula

$1\ ^oF = 32 + 9/5\ ^oC$

If avg temp is 30 ^oC, What's the corresponding Avg temp in ^oF?

$\text{Avg} = 32 + 9/5\ (30^oC)$

$\text{Avg} = 86\ ^oF$

Average

Uses all data in the computation

Use with quantitative variables

Can be used for estimating projected totals

Sensitive to outliers

Limiting without information about spread

Should not be used with qualitative variables

The average may be an impossible value

Median

Middle observation

Use with almost any distribution

Not affected by outliers

The median is an actual observed value (or almost)

Not used enough

Not understood

Cannot be used for estimating projected totals