

# Measures of Central Tendency

---

Mike DeWitt  
Institutional Research



WAKE FOREST  
UNIVERSITY

Winston Salem, NC | Dec, 2017

So why do we need measures of central tendency?

Measures of central tendency allow us to summarize data sets with one, convenient number

---

So why do we need measures of central tendency?

When asked how much  
does gas cost today  
what numbers do you  
report?



## So why do we need measures of central tendency?

When asked “how much is gas” what numbers do you report?

- Every different price you saw?

---

## So why do we need measures of central tendency?

When asked “how much is gas” what numbers do you report?

- Every different price you saw?
- The lowest or highest?

---

## So why do we need measures of central tendency?

When asked “how much is gas” what numbers do you report?

- Every different price you saw?
  - The lowest or highest?
  - Or a measure of centrality?
    - Mean (average)
    - Median
    - Mode
-

## Measures of Central Tendency—Mean

The **mean** or **average** is the most common measure of central tendency and is the sum of all numbers in a data set divided by the count

---

## Measures of Central Tendency—Mean

The mean or average is the most common measure of central tendency and is the sum of all numbers in a data set divided by the count

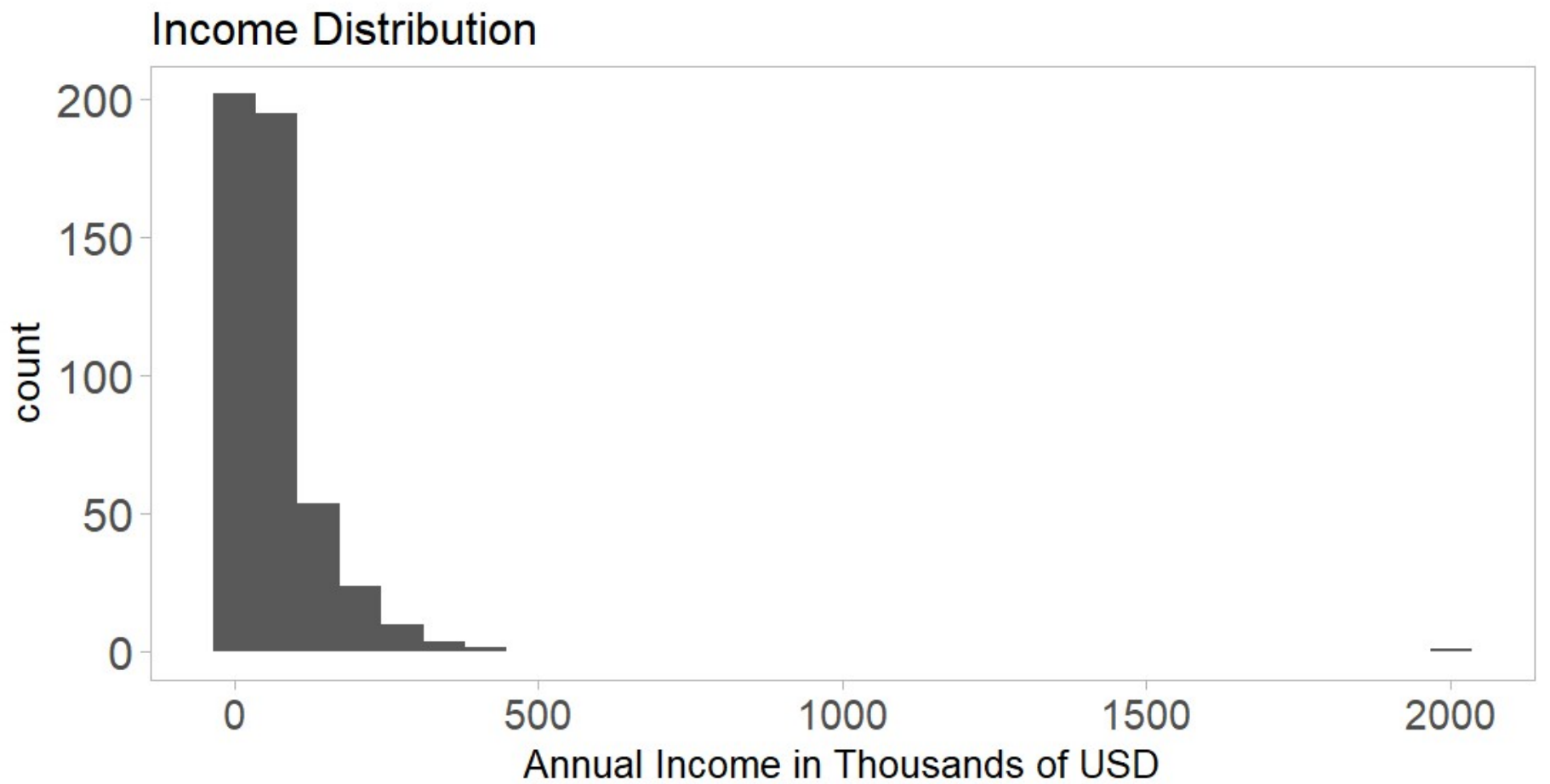
We'll let Excel do the calculating for us though...

---



## The mean is not a robust measure

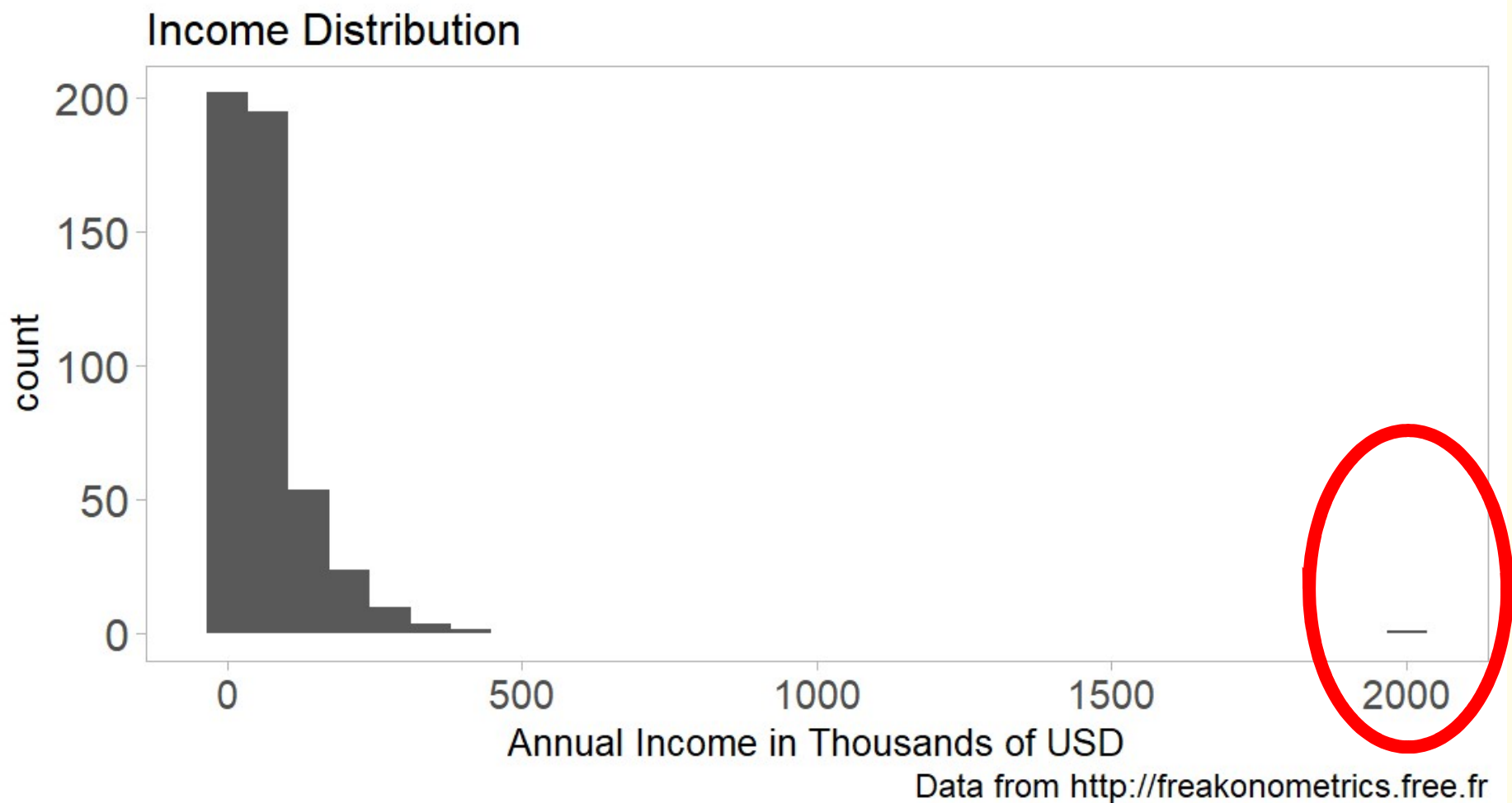
An issue with the mean is that it is sensitive to outliers



Data from <http://freakonometrics.free.fr>

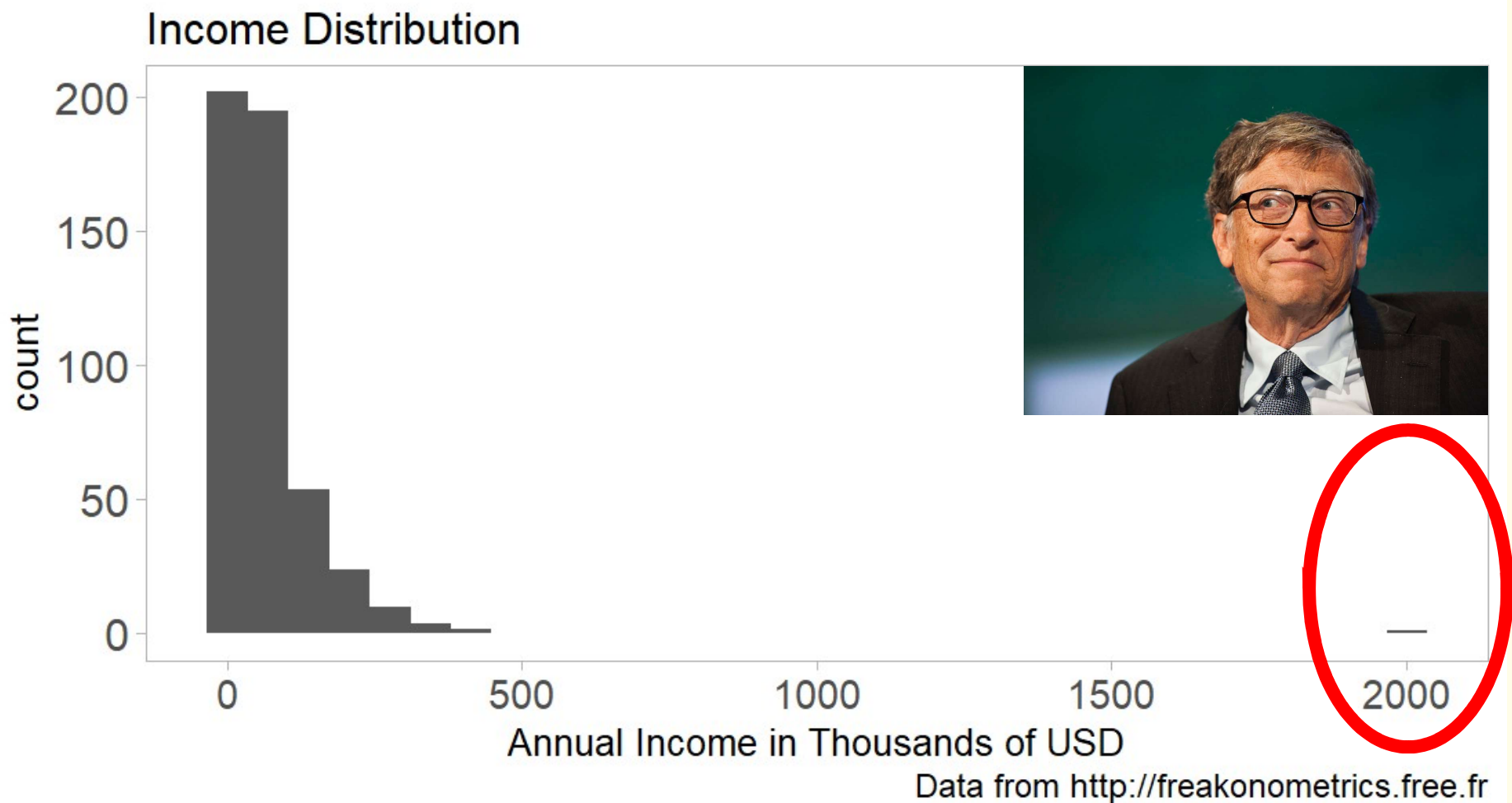
## The mean is not a robust measure

An issue with the mean is that it is sensitive to outliers



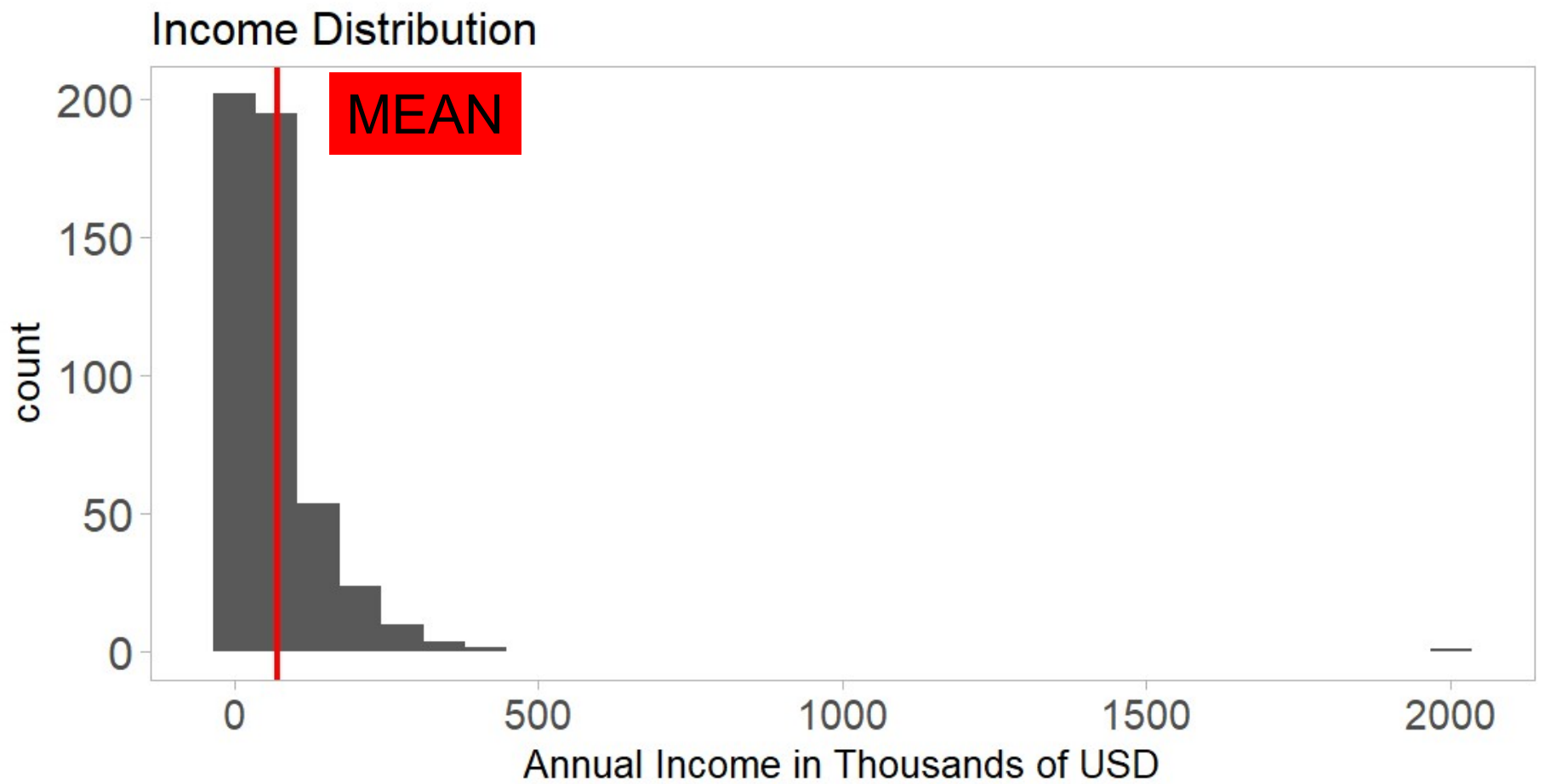
# The mean is not a robust measure

An issue with the mean is that it is sensitive to outliers



# The mean is not a robust measure

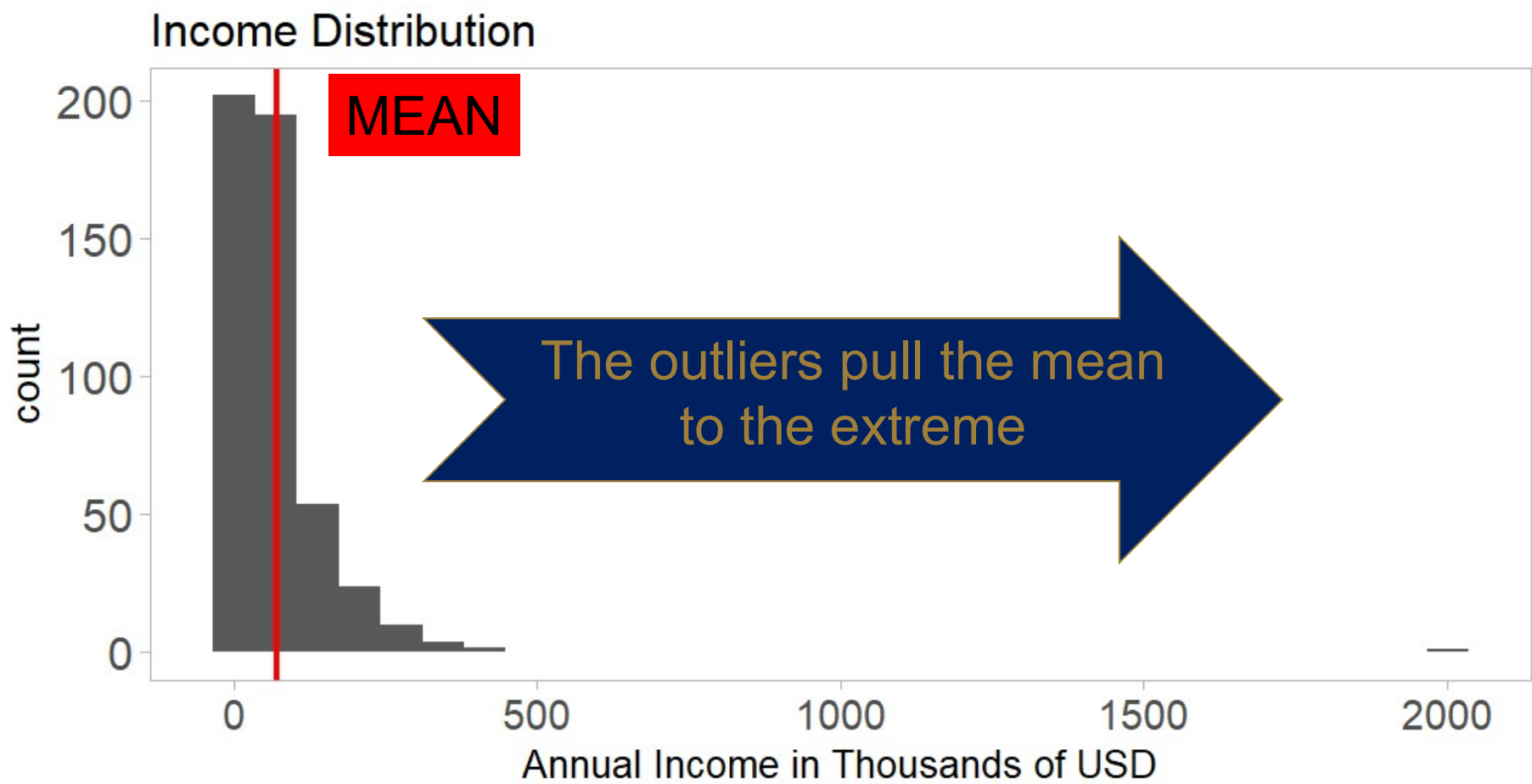
An issue with the mean is that it is sensitive to outliers



Data from <http://freakonometrics.free.fr>

## The mean is not a robust measure

An issue with the mean is that it is sensitive to outliers



Data from <http://freakonometrics.free.fr>

## Measures of Central Tendency—Median

The **median** is the “middle” number in the dataset

If you were to order the values in a dataset from largest to smallest, the **middle value** would be your median

---

## Measures of Central Tendency—Median

The **median** is the “middle” number in the dataset

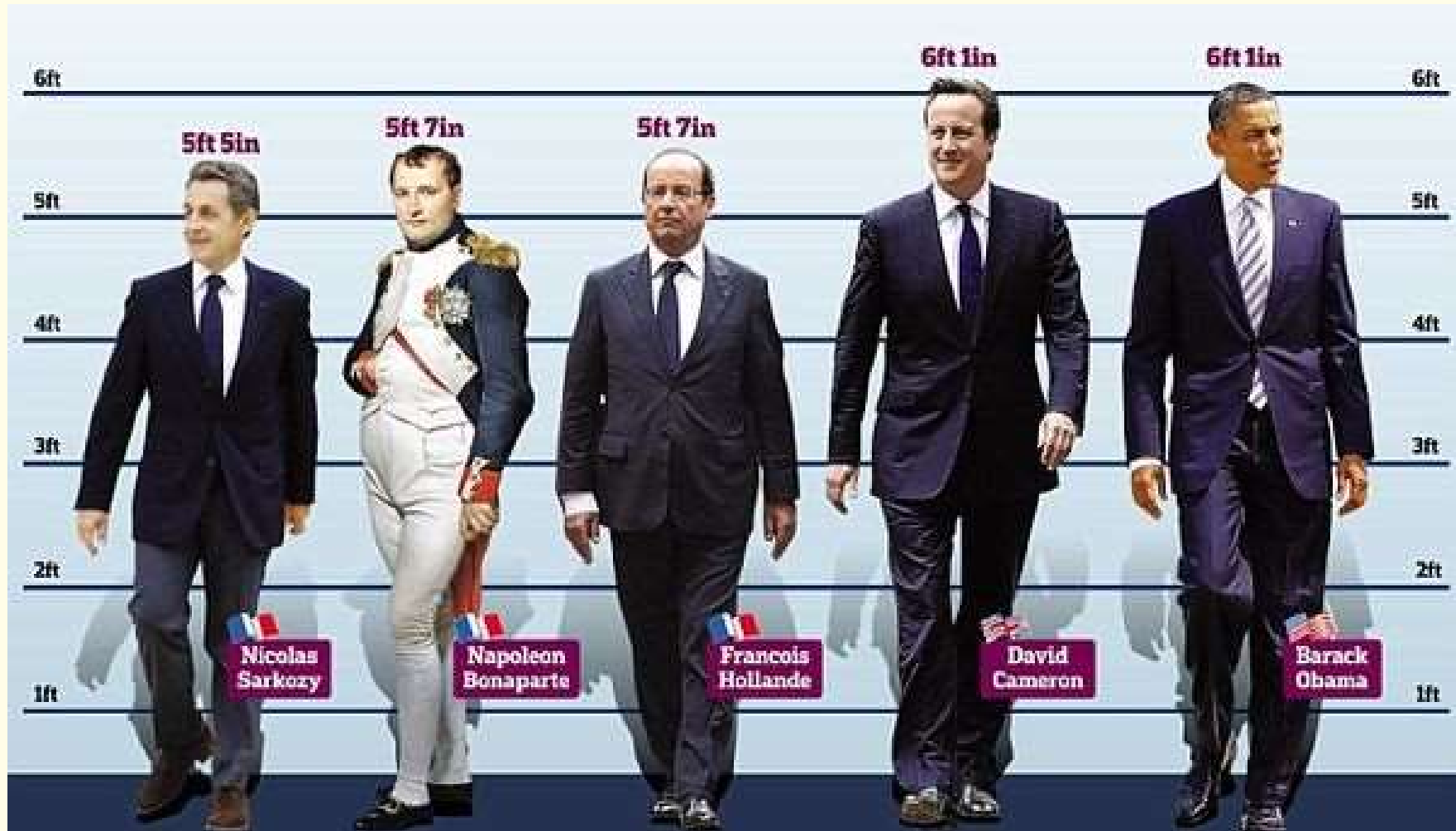
If you were to order the values in a dataset from largest to smallest, the **middle value** would be your median

Half of the population is above this value and half of the population is below this value

---

# Measures of Central Tendency—Median

The **median** is the “middle” number in the dataset





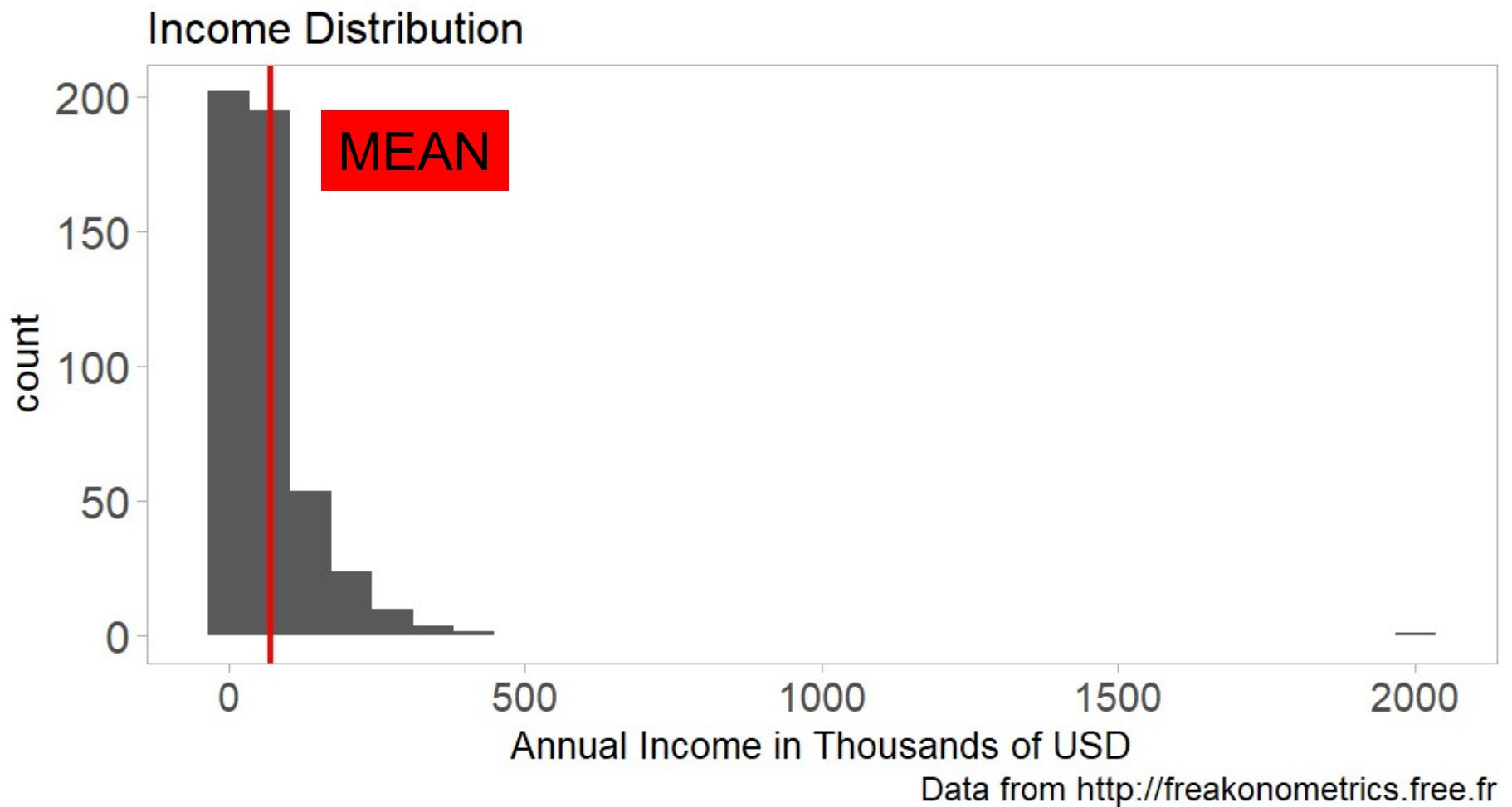
# Measures of Central Tendency—Median

The **median** is the “middle” number in the dataset



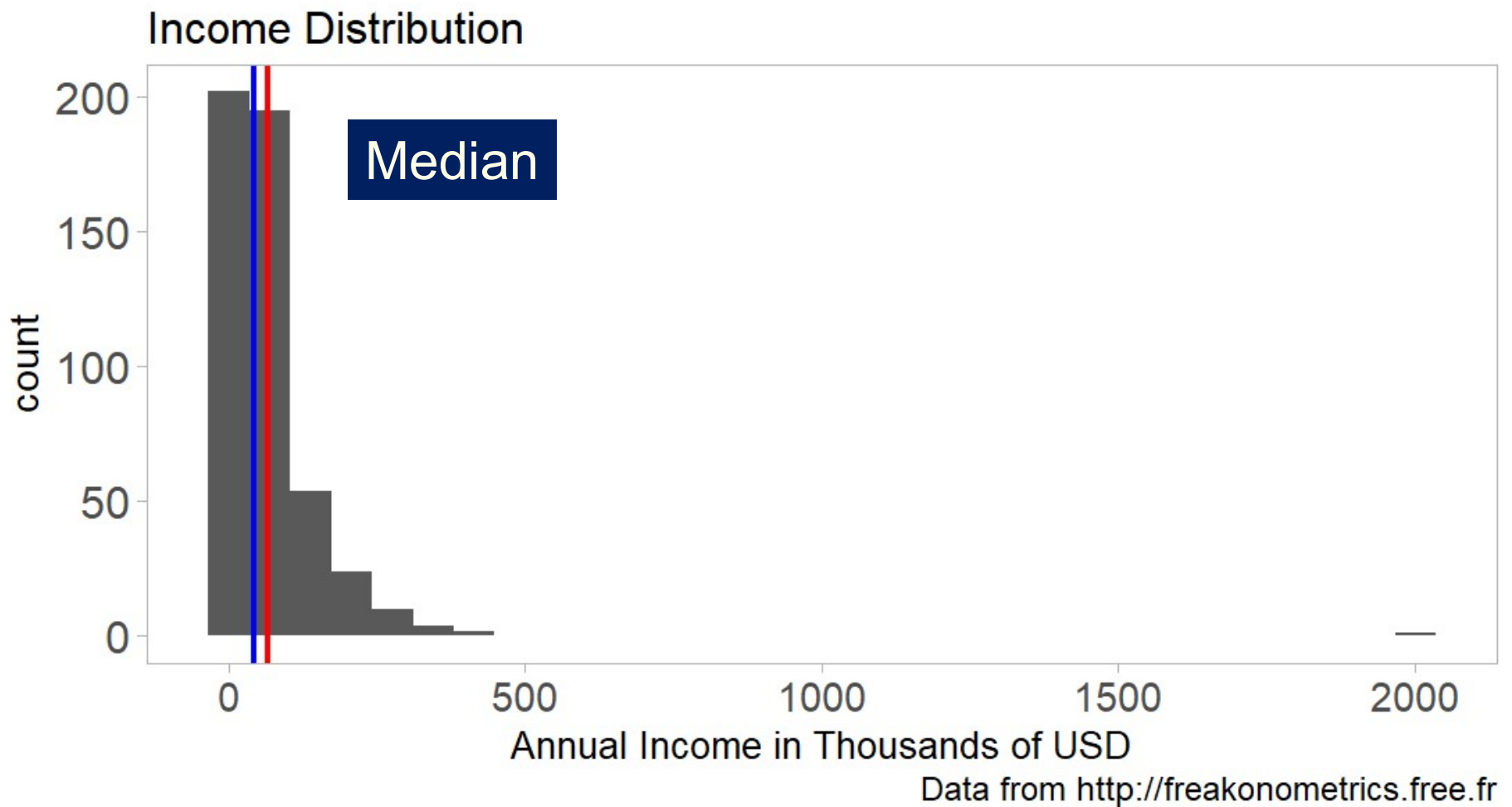
The median is a robust measure of central tendency

## The median is more robust against **outliers**



The median is a robust measure of central tendency

## The median is more robust against **outliers**



## Measures of Central Tendency—Mode

**Mode** is the most often occurring value that appears in a dataset

---

## Measures of Central Tendency—Mode

**Mode** is the most often occurring value that appears in a dataset

This is a useful measure to understand how often values appear

---

## Measures of Central Tendency—Mode

**Mode** is the most often occurring value that appears in a dataset

This is a useful measure to understand how often values appear

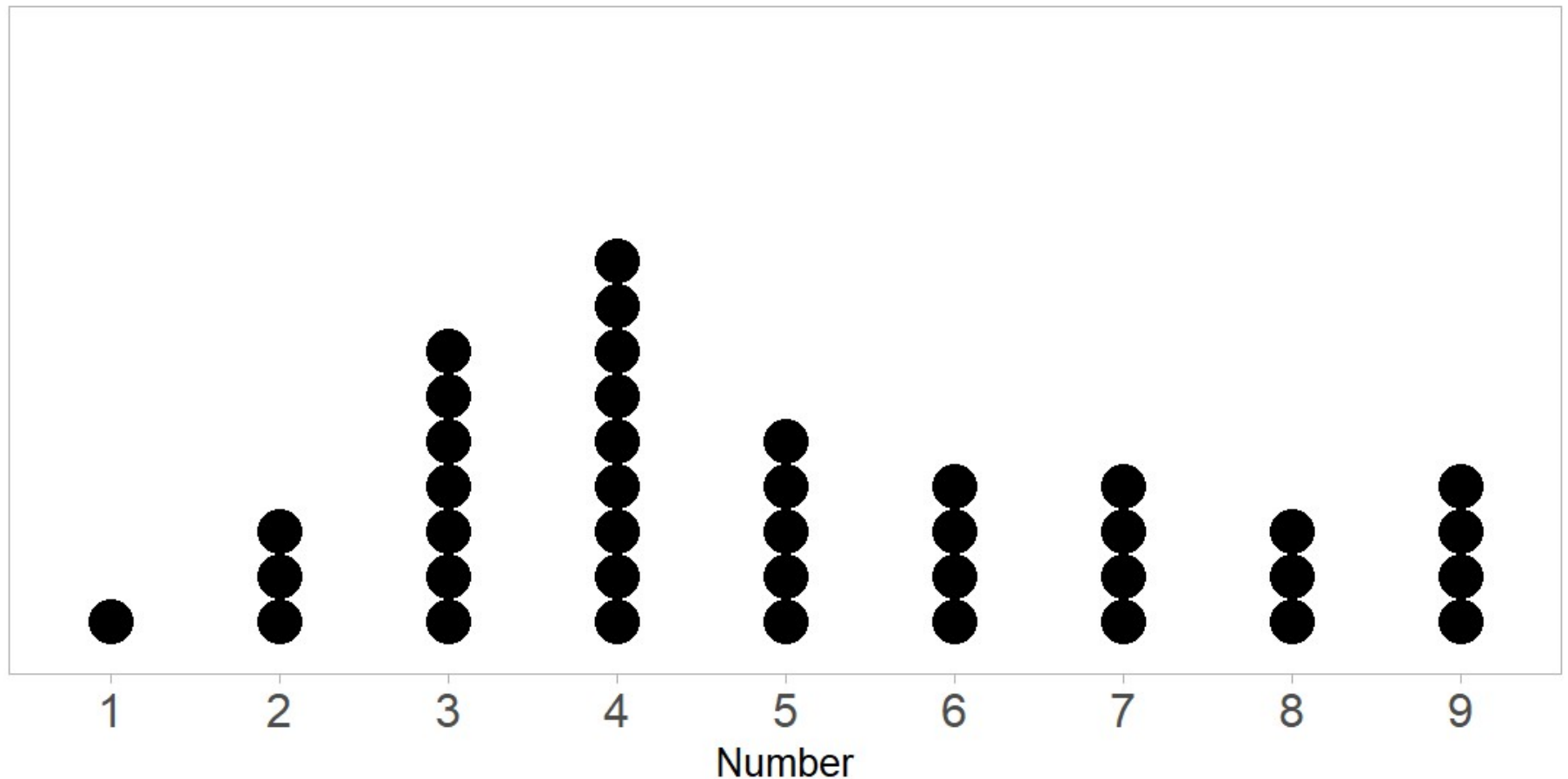
Mode is also beneficial for summarizing **categorical data** (like survey responses)

---

## Measures of Central Tendency—Mode

**Mode** is the most often occurring value that appears in a dataset

Finding the Mode of A Distribution



## Measures of Central Tendency—Mode

**Mode** is the most often occurring value that appears in a dataset

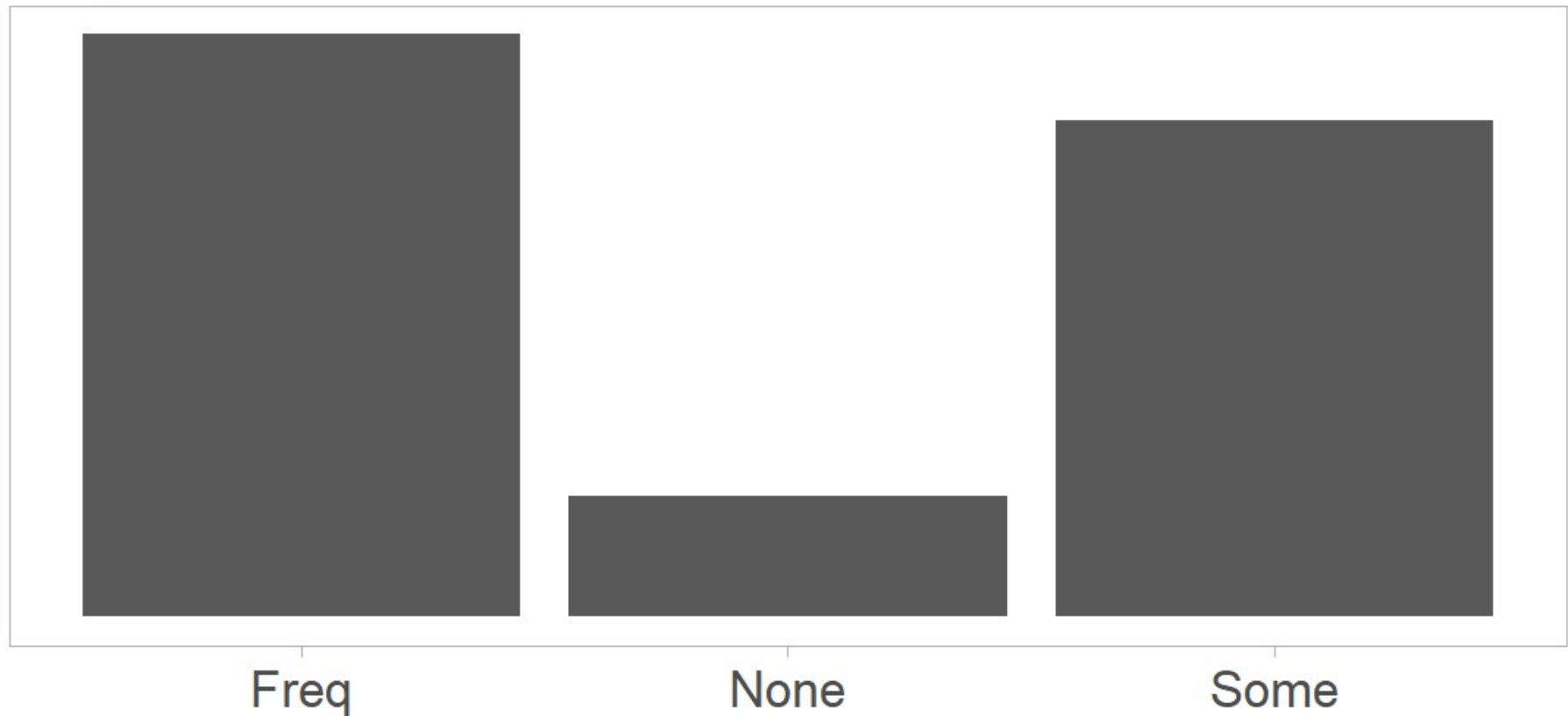
Finding the Mode of A Distribution





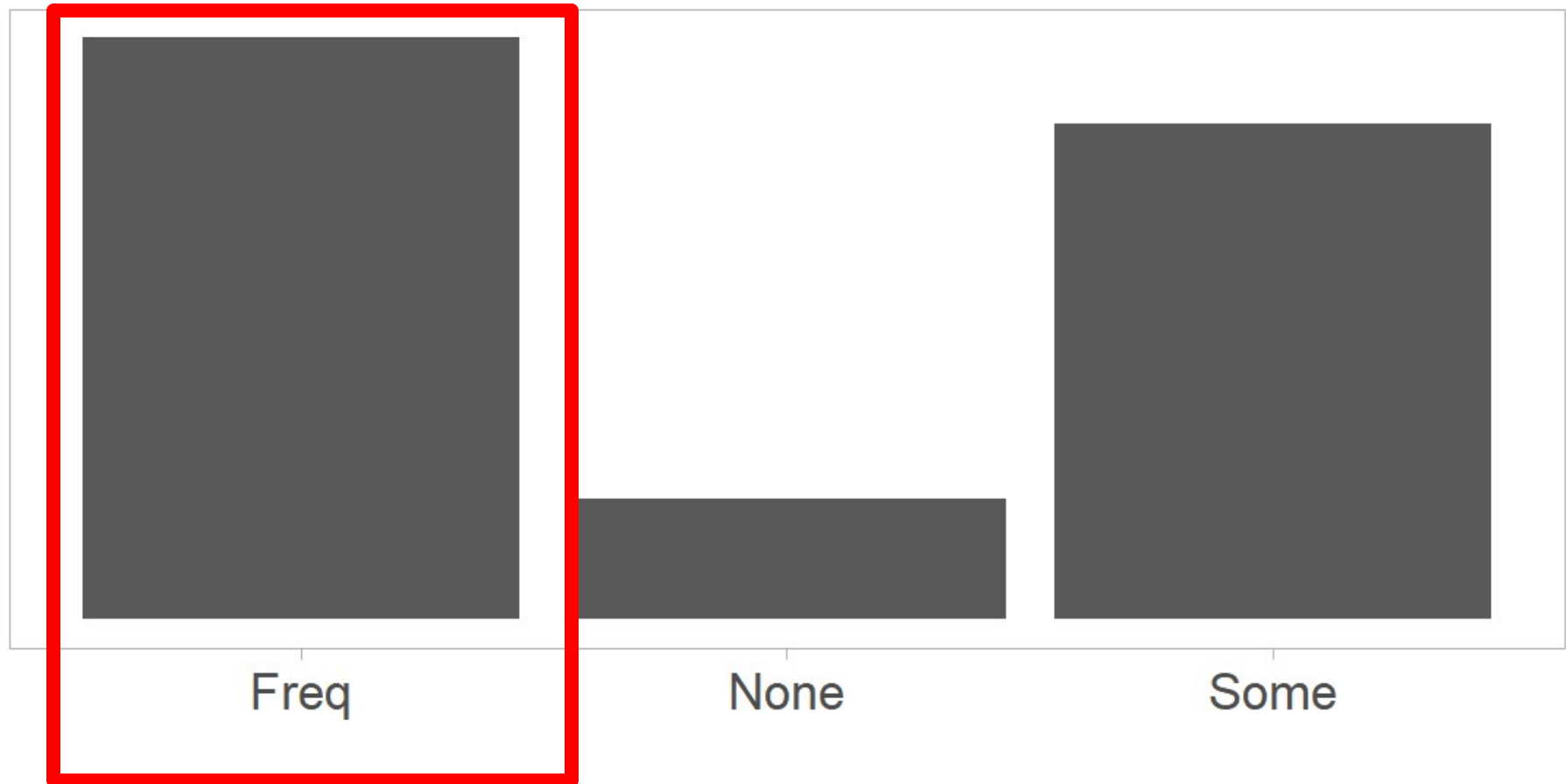
## Survey Data Lends Itself to Analysis with the Mode

Response to How Often Do You Exercise

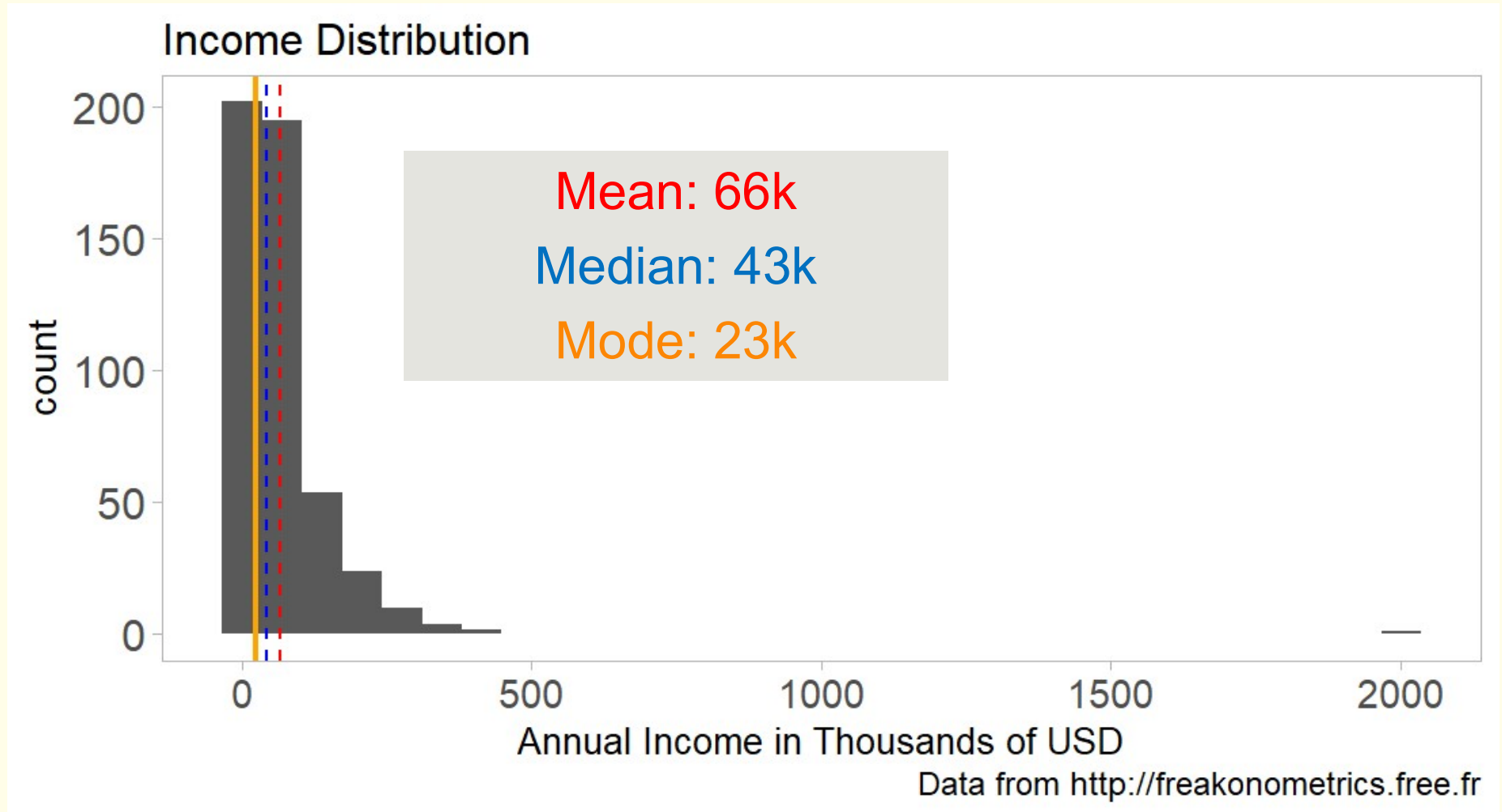


## Survey Data Lends Itself to Analysis with the Mode

Response to How Often Do You Exercise



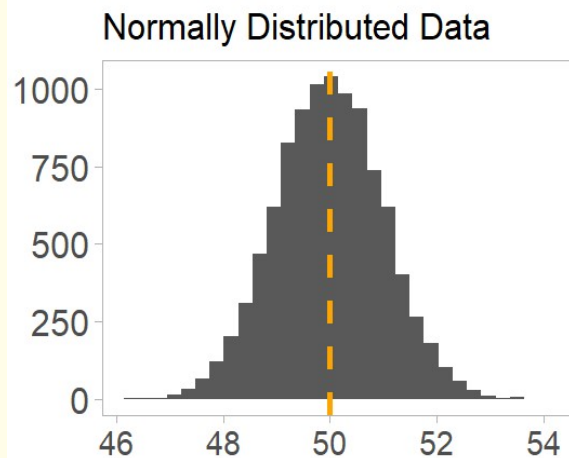
# Measures of Central Tendency—Mode



# When to Use Each Measure?

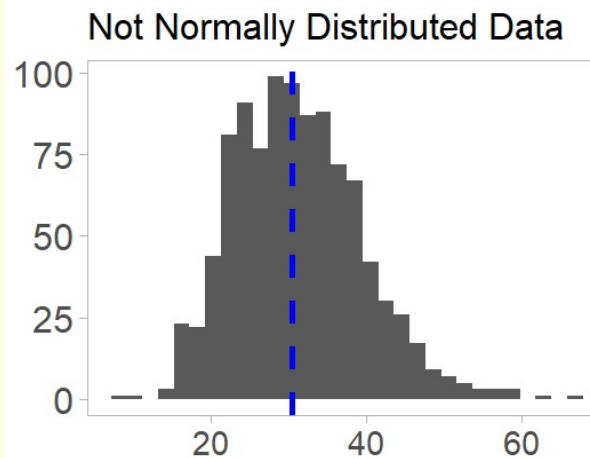
## Mean

Data are centered  
(normally  
distributed)



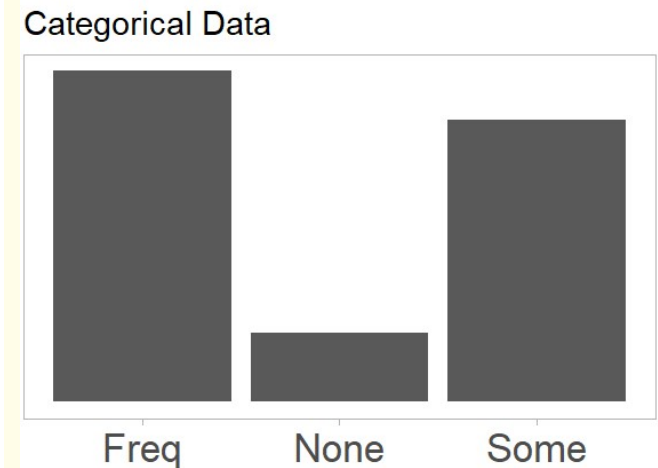
## Median

Data are skewed  
or  
has outliers



## Mode

When the data  
are categorical



## General Guidelines for Determining Bin Sizes

- The bins should be equally spaced
  - There should be at least 10 bins in a histogram
    - More if you don't see any variation
  - The bins should include the minimum and the maximum
    - Outliers? YES!
  - Try to use “nice” intervals that make sense
    - E.g. by 1s, 2s, 5s, 10s
  - *Consider context of the analysis*
-

## A Measure of Central Tendency is Only One Number

The problem with a one number summary of data is that it doesn't tell us anything about the **distribution** of the data.

---

## A Measure of Central Tendency is Only One Number

The problem with a one number summary of data is that it doesn't tell us anything about the **distribution** of the data.

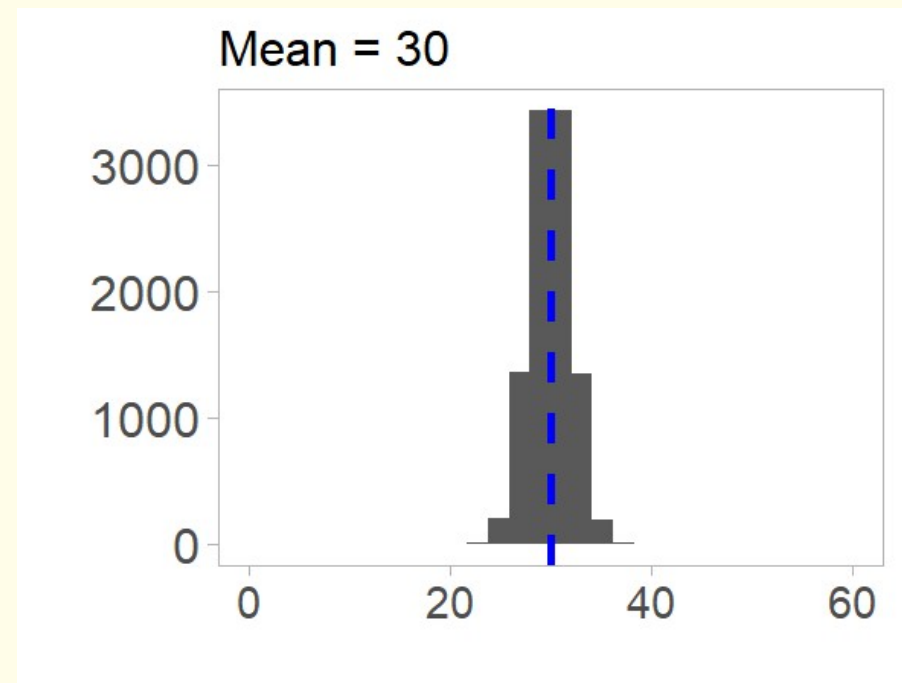
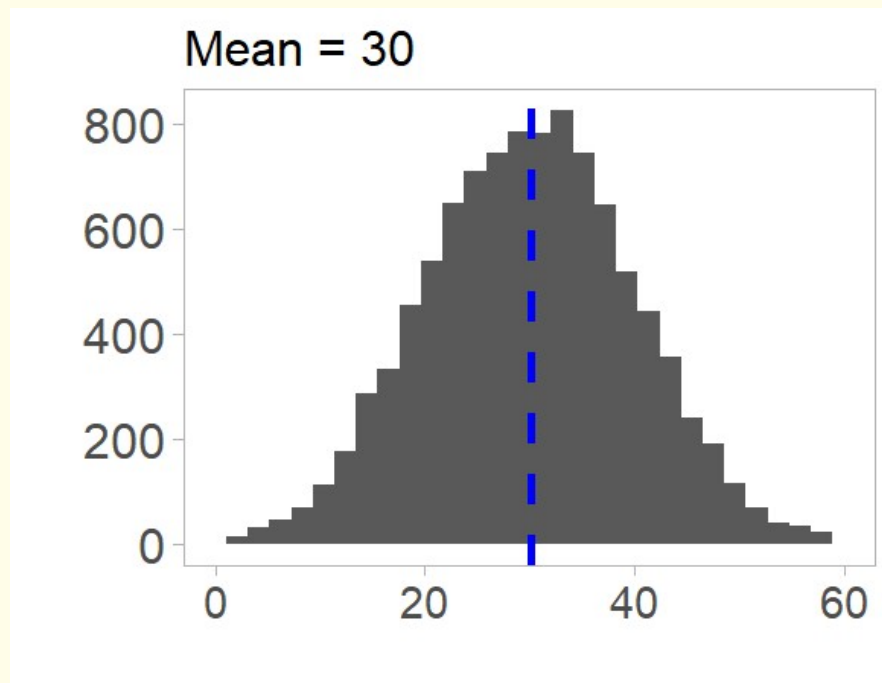
Example: I have two groups with a mean score of 30.

They're equivalent, right?

---

## A Measure of Central Tendency is Only One Number

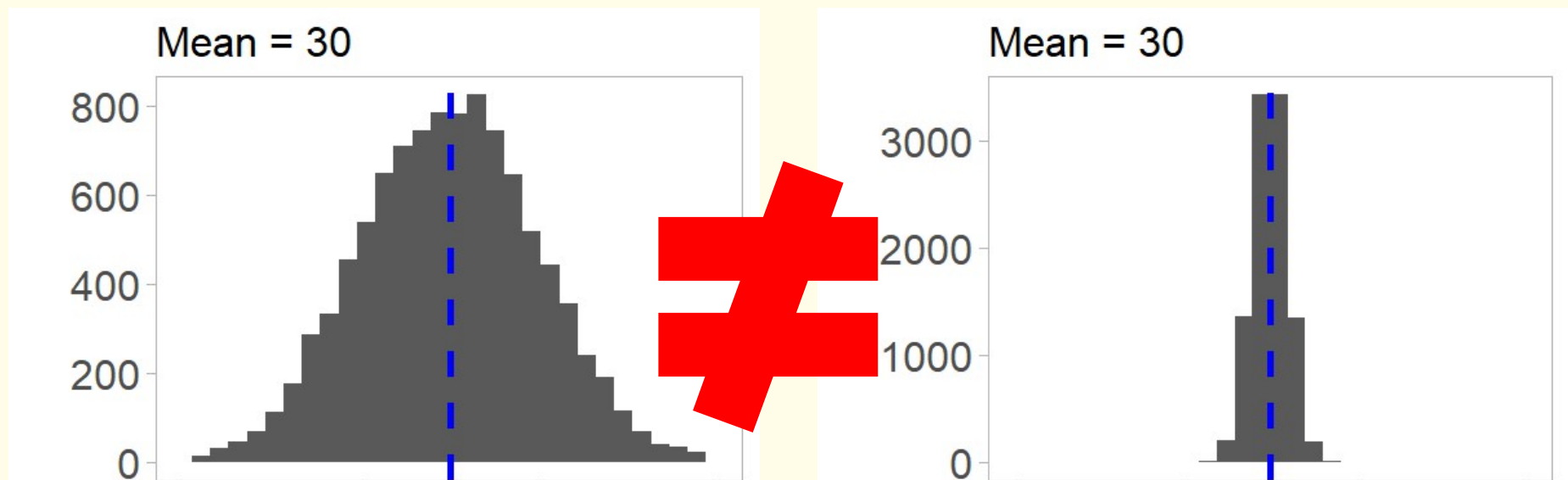
The problem with a one number summary of data is that it doesn't tell us anything about the **distribution** of the data.





## A Measure of Central Tendency is Only One Number

The problem with a one number summary of data is that it doesn't tell us anything about the **distribution** of the data.



These don't look like they are from the same population!  
Same mean, but different "spread" around that mean

## The Standard Deviation is a Measure of Spread

Our measures of central tendency do not provide us with a measure of **spread** or **variation** about the central measure (mean, median, mode)

---

## The Standard Deviation is a Measure of Spread

One measure of variation about the mean is called **standard deviation**

---

## The Standard Deviation is a Measure of Spread

One measure of variation about the mean is called **standard deviation**

Standard deviation is the square root of the mean square error

---

## The Standard Deviation is a Measure of Spread

One measure of variation about the mean is called **standard deviation**

Standard deviation is the square root of the mean square error

...(again, let excel do the work for you)

---

## When Data are Normal the Standard Deviation Tells A LOT!

When data are normally distributed (following a bell shaped curve) the standard deviation signals additional information regarding the spread of the data.

---

## When Data are Normal the Standard Deviation Tells A LOT!

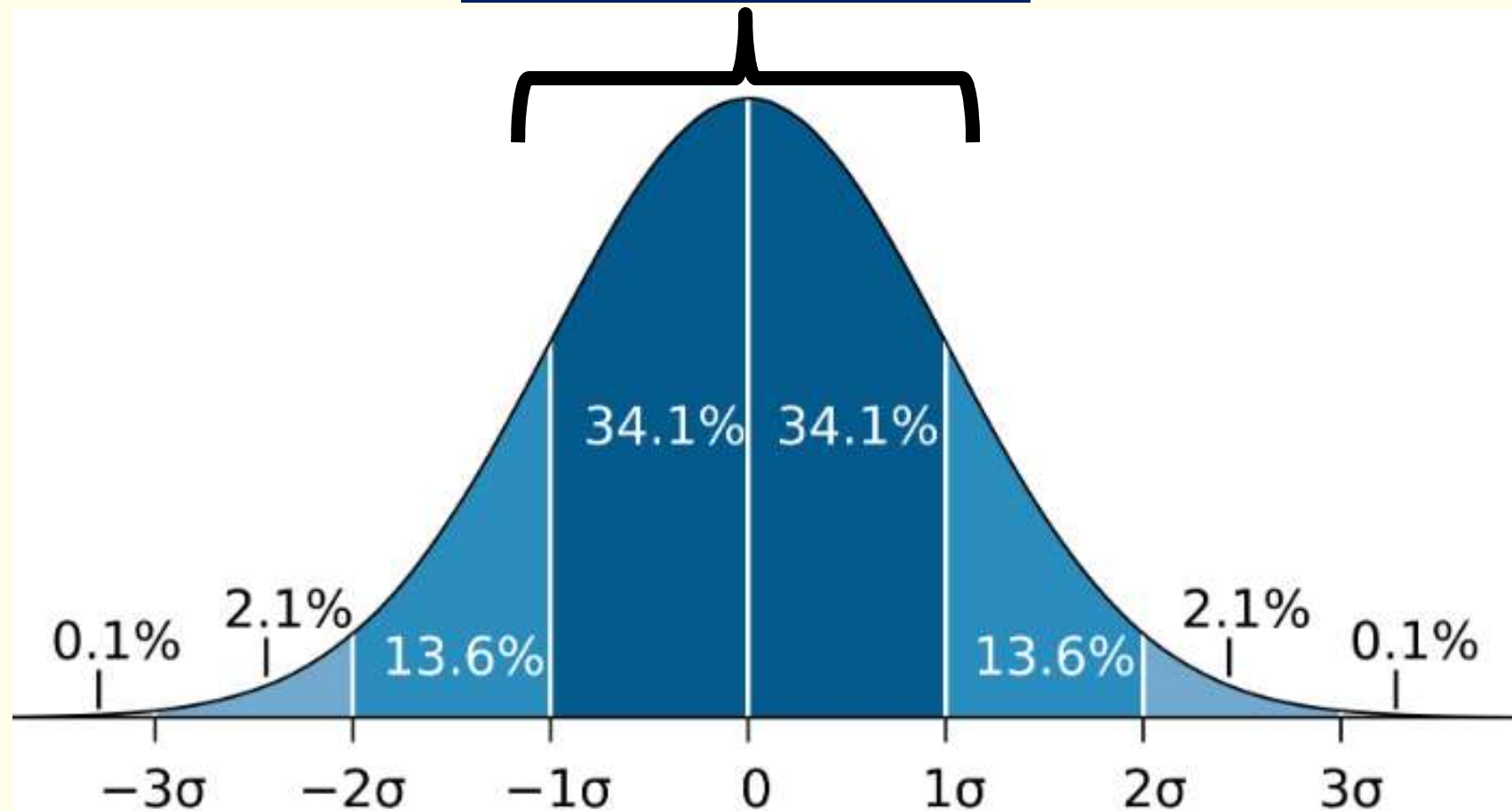
When data are normally distributed (following a bell shaped curve) the standard deviation signals additional information regarding the spread of the data.

We won't go into detail today about more rigorous tests of normality.

---

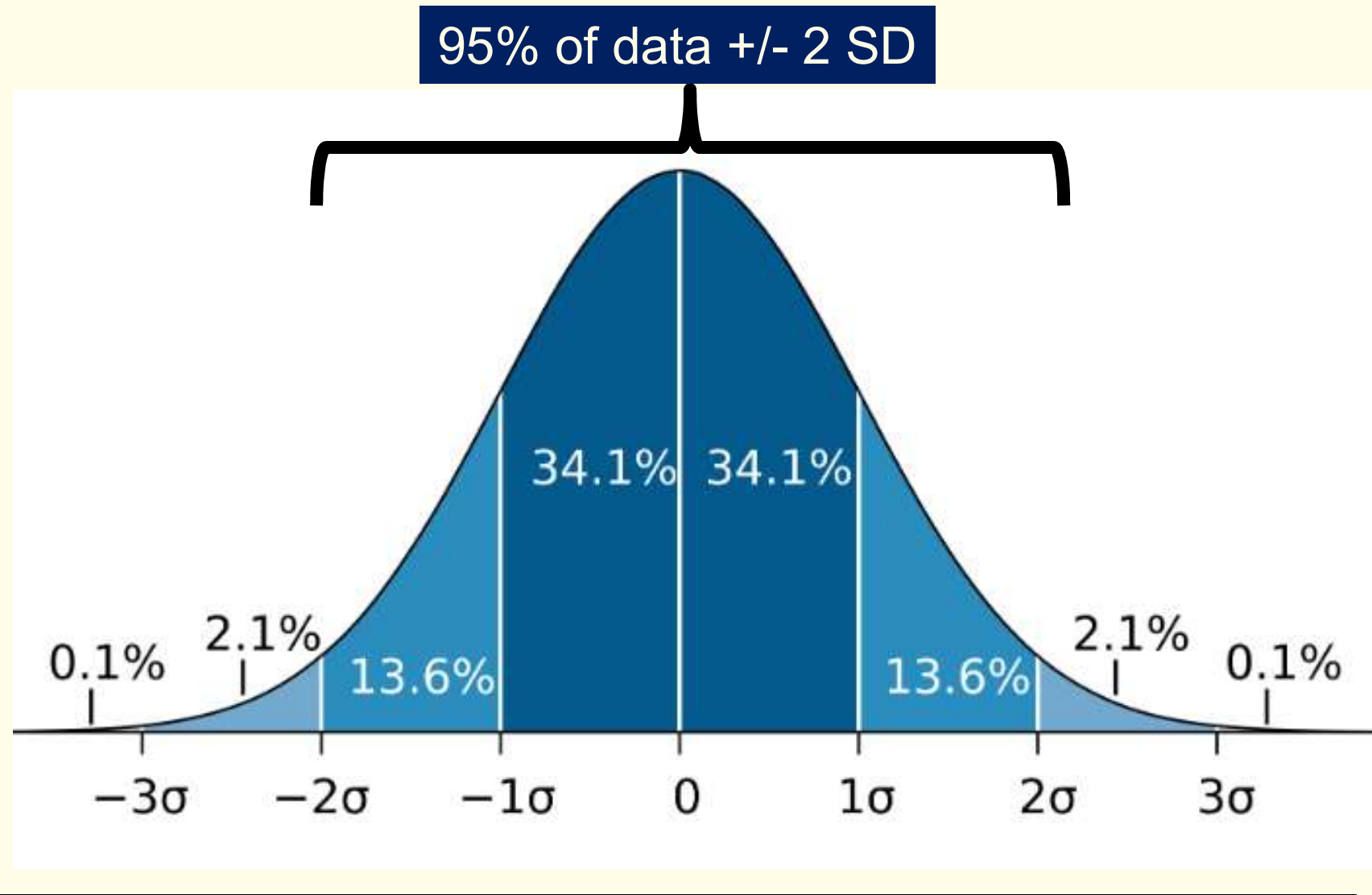
## 68-95-99.7% Rule For Normal Distributions

68% of data  $\pm 1$  SD



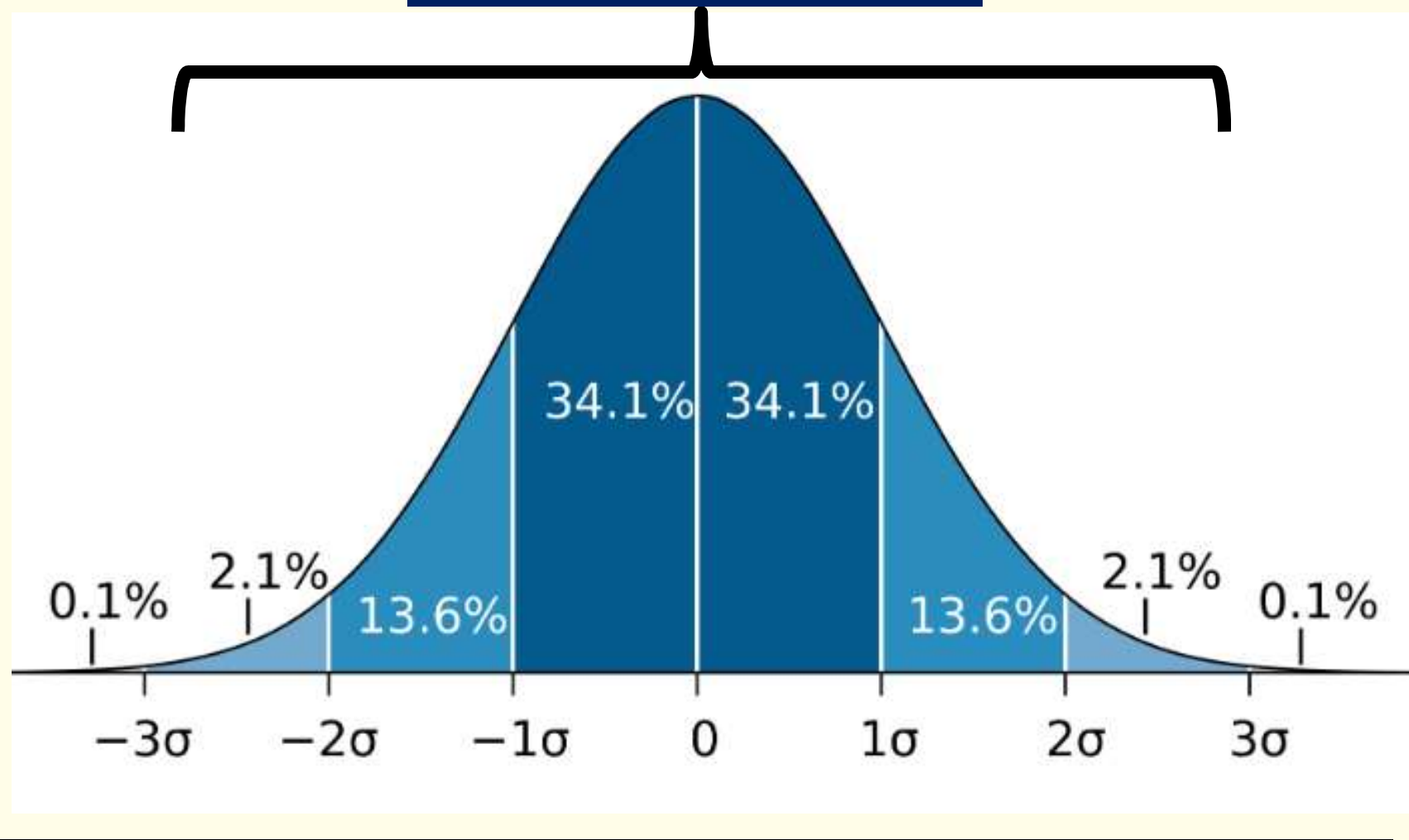


## 68-95-99.7% Rule For Normal Distributions



## 68-95-99.7% Rule For Normal Distributions

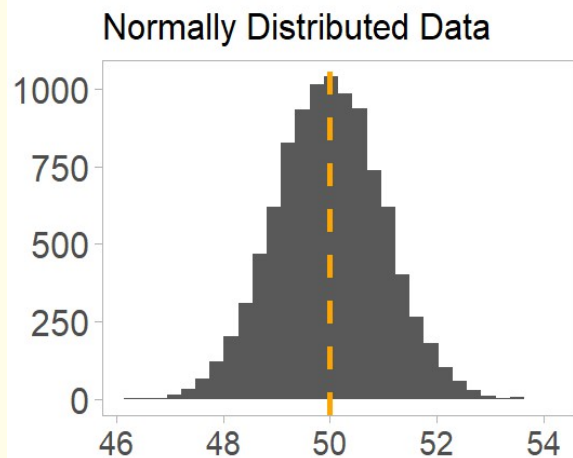
99.7% of data  $\pm 2$  SD



## To Summarize Today's Measures

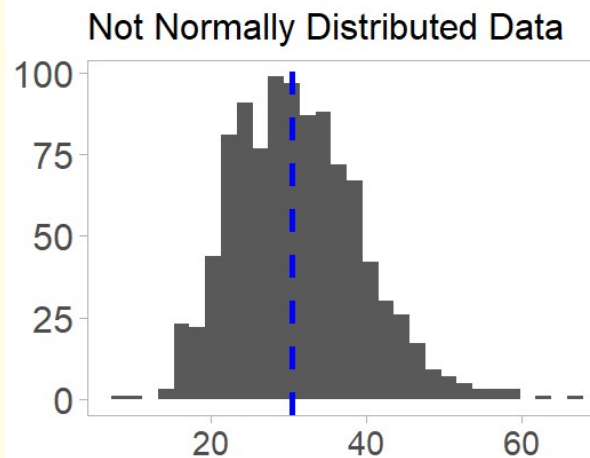
### Mean

Data are centered  
(normally  
distributed)



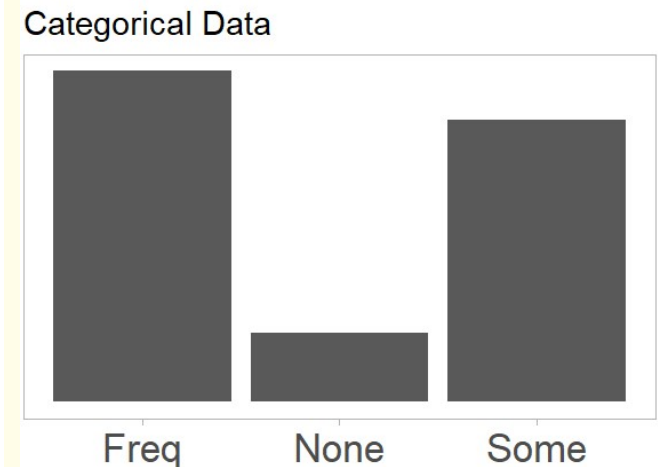
### Median

Data are skewed  
or  
has outliers



### Mode

When the data  
are categorical



### Standard Deviation

Measure of spread about the mean.

If the data are normally distributed we can use  
the 68-95-99.7% rule

