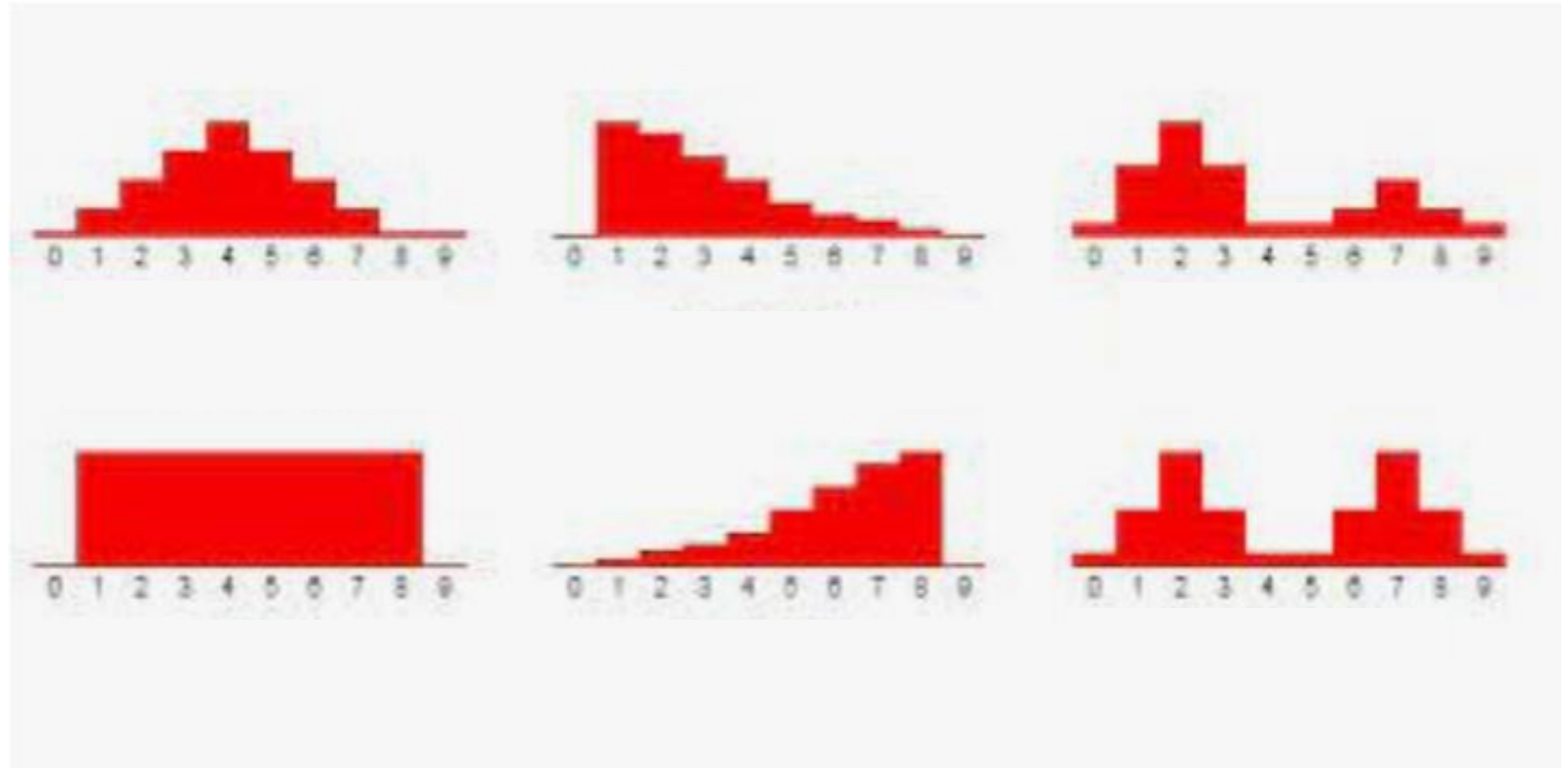


# Summarizing Distributions

Mean, Median  
Standard Deviation, Interquartile Range

# Review: Describing Distributions

- Uniform distribution
- Bell curve
- Symmetric
- Skewed (left and right)
- Bimodal



# Today we will review tools for *summarizing* distributions

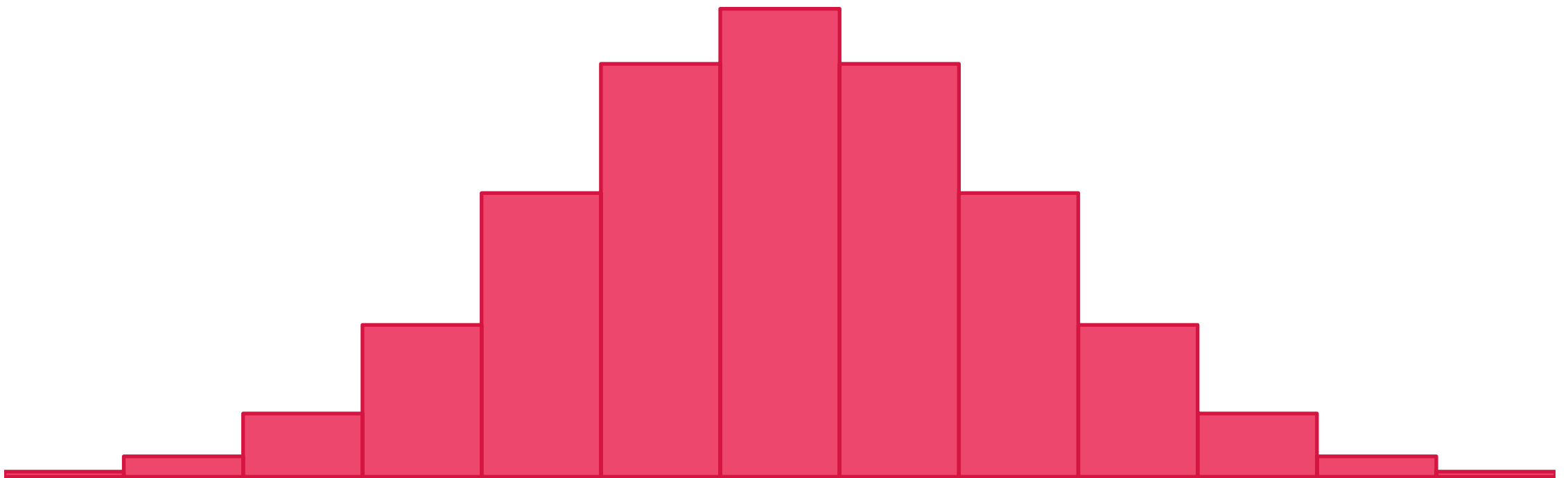
---

- Central tendencies of Mean and Median and when to use which one
- Range and Standard deviation
- Five Number Summary and Interquartile Range

## Essential Question

---

How can I **summarize** a whole distribution using just **a few numbers**?



## When *should* I summarize a distribution?

---

When the summary can **accurately portray** the distribution as a whole!

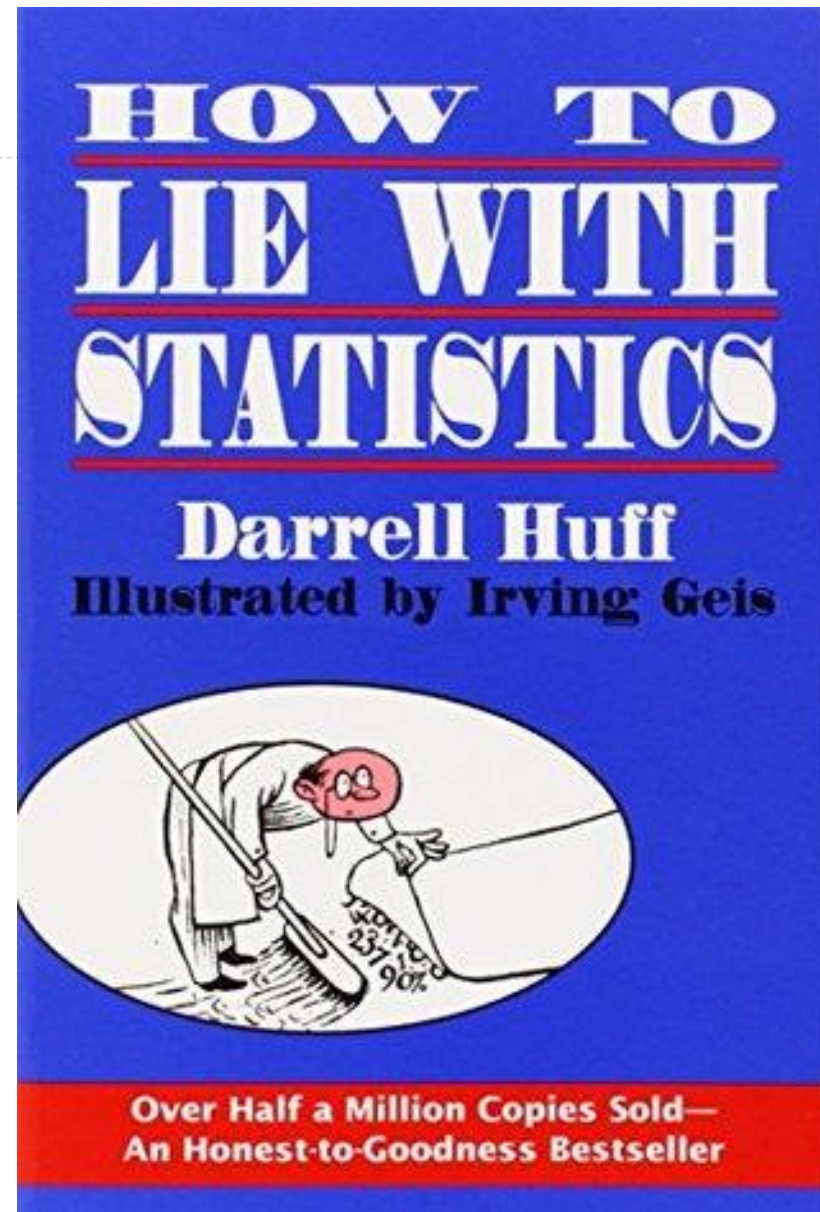
*“There are three kinds of lies:*

*lies,*

*damned lies,*

*and statistics”*

Mark Twain



# Doing math with a different mindset

---

## ALGEBRA

The stuff we do in algebra ...

... tells us something **insightful**

... is hopefully **easy to calculate**

## STATISTICS

The stuff we do to summarize  
distributions ...

... is **easy to calculate**

... hopefully tells us something  
**insightful**

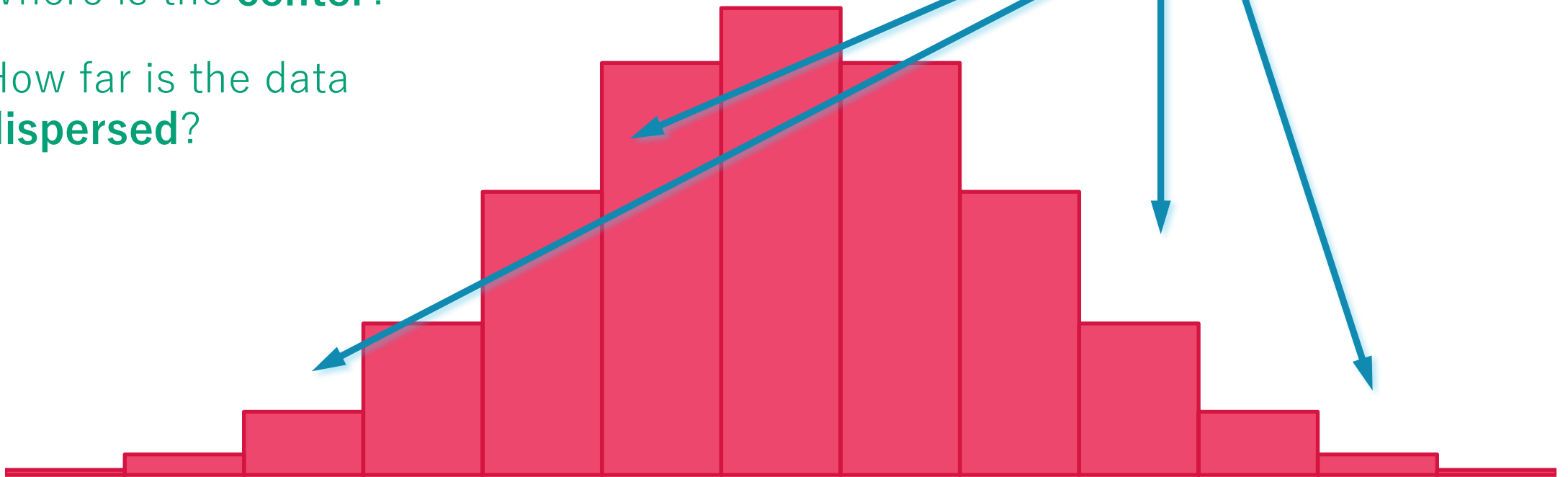
# Summarizing Bell Curves

**Central tendency:**  
Values tend to cluster

**Dispersion:** Values are spread out  
around the central value

Where is the **center**?

How far is the data  
**dispersed**?





# Tools for Summarizing Distributions

---

## CENTRAL TENDENCY

- Mean
- Median
- Mode

## DISPERSION

- Range
- Standard Deviation
- Interquartile Range (IQR)

## Calculate the Mean and find the Median (Center)

---

- **Mean:**  $\frac{\textit{Add the data}}{\textit{The number of pieces of data}}$
- We sometimes use  $\mu$  (mu) as shorthand for the mean
- **Median:** Put the numbers in numerical order and find the number in the middle. If there is an even number of data, find the mean of the middle two numbers.

## Spread/Variation

---

- **Range:** The difference between the largest and smallest number. This is a single number, not an interval.
- **Standard deviation:** The average distance each piece of data is from the mean
- We sometimes use  $\sigma$  (sigma) as shorthand for the standard deviation

## Calculate the Range

---

- Maximum - Minimum

## Use your calculator to find the standard deviation

---

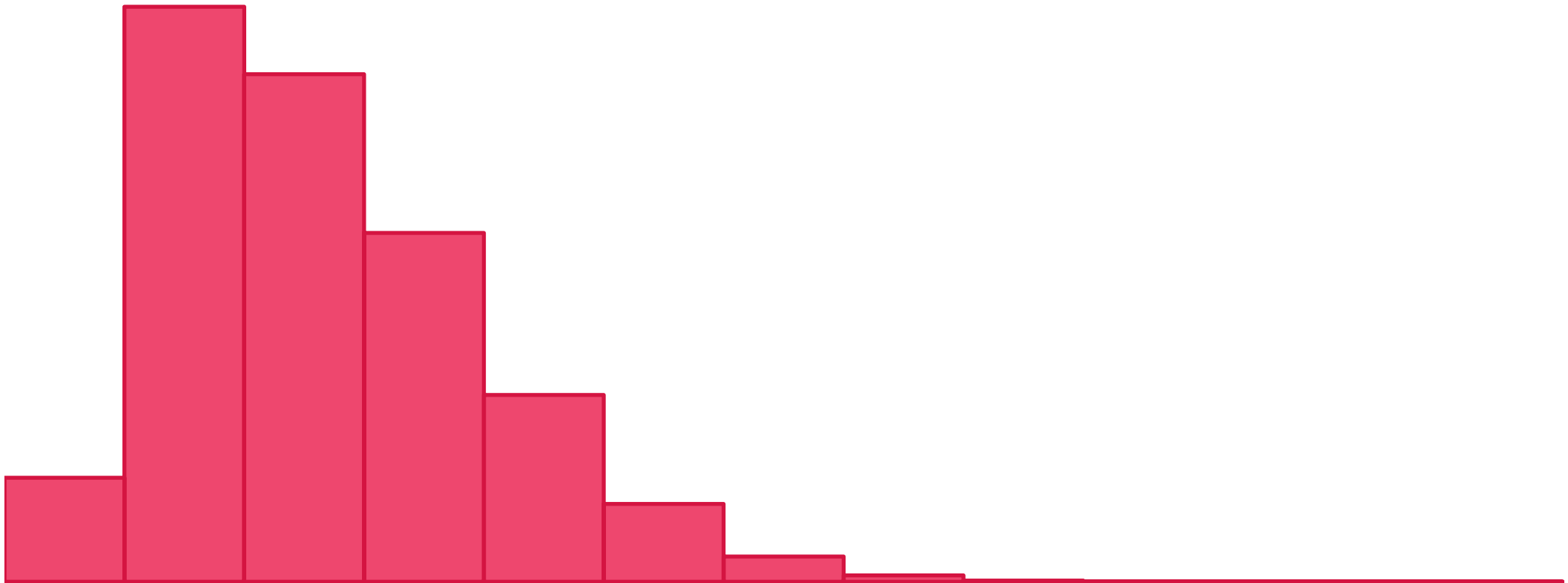
- *Stat* → *1:Edit* Enter the data into L1. If you have data in the list already, use the arrow keys to highlight L1 → *clear* → *enter*
- *stat* → *CALC* → *1:1 – Var Stats* → *L1* (or whatever list you used if you used a different one, the calculator defaults to L1) → *Calculate* (leave *FreqList* blank)
- $\bar{x}$  read as x-bar, is the mean
- $s_x$  is the standard deviation
- Continue to scroll down and you will see Q1 and Q3 so you can find the Interquartile Range (IQR)

## What if the distribution is skewed?

---

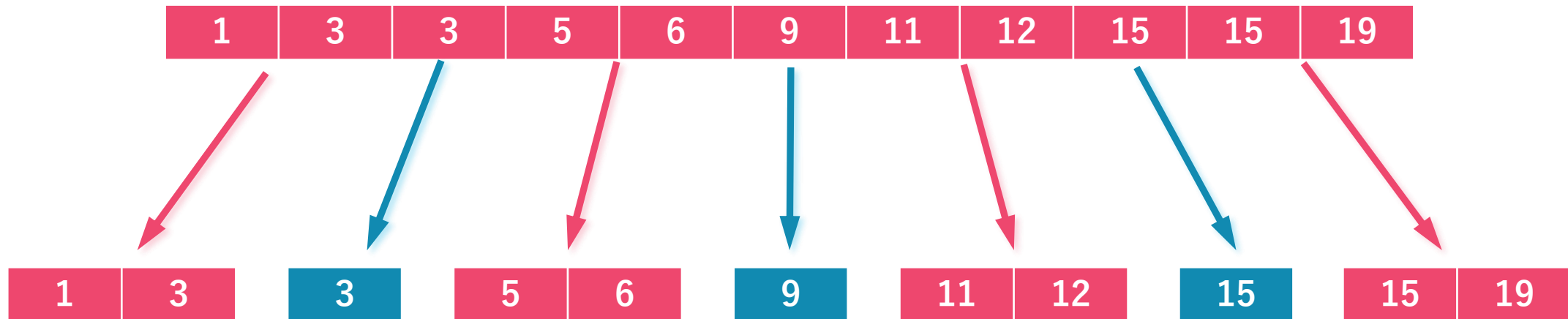
What should the **center** of this distribution be? Are there multiple valid opinions?

Does your definition of the center also change your definition of **dispersion**?



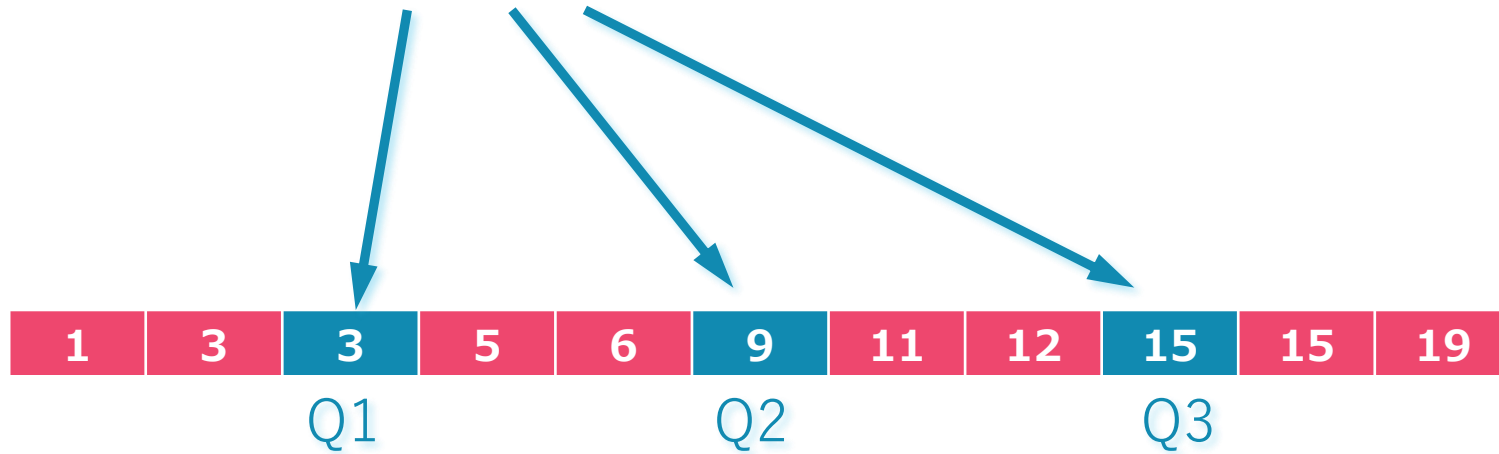
## A Different Strategy

Take the data set, put the values in **ascending order**, and **divide into groups**.  
Then, to summarize the distribution, look at the numbers **between** the groups.



# Quantiles

Quantiles are **cut points** that divide the data into equal-sized intervals.



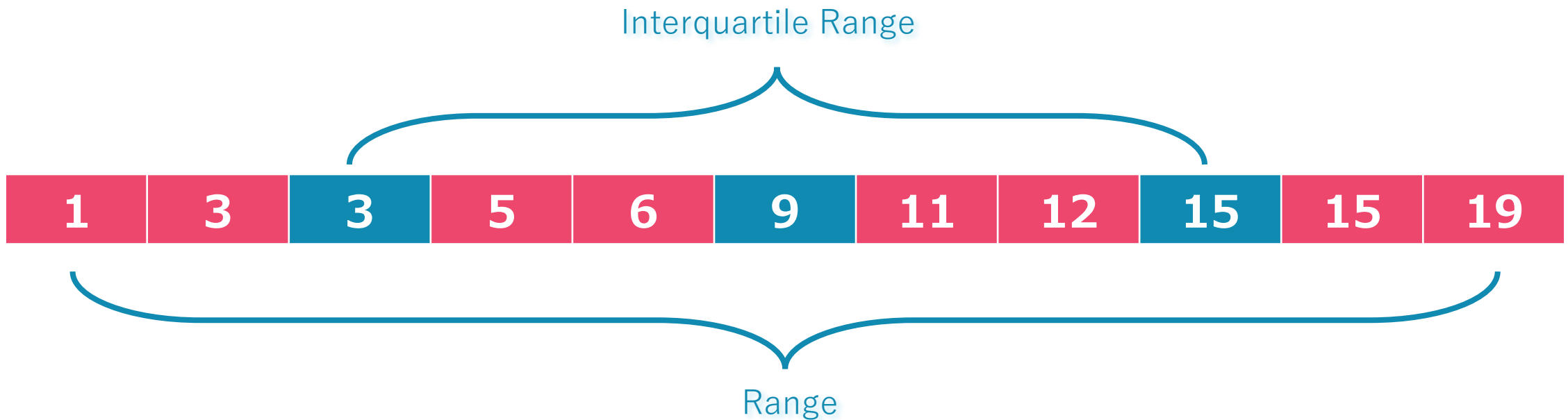
The data above has been split into **quartiles**: four groups, three cut points.

Name	Number of Intervals
Median	2
Quartile	4
Decile	10
Percentile	100



## Interquartile Range (IQR)

Range of the **middle half** of the distribution, from the **first quartile** to the **third quartile**:



# Calculating Ranges

---

## INTERQUARTILE RANGE

Range of the **middle half** of the distribution, from the first to third quartile

$$Q3 - Q1$$

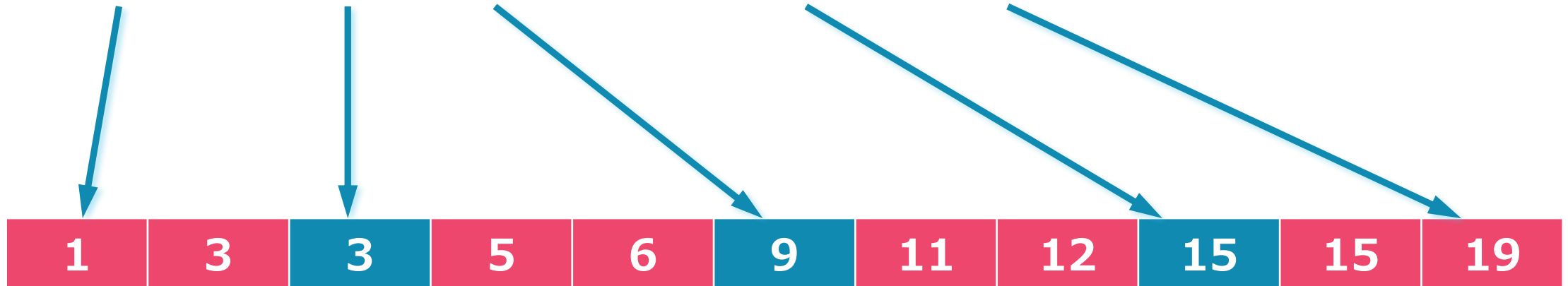
## RANGE

Range of the **entire** distribution

$$\text{Max} - \text{Min}$$

## 5 number summary

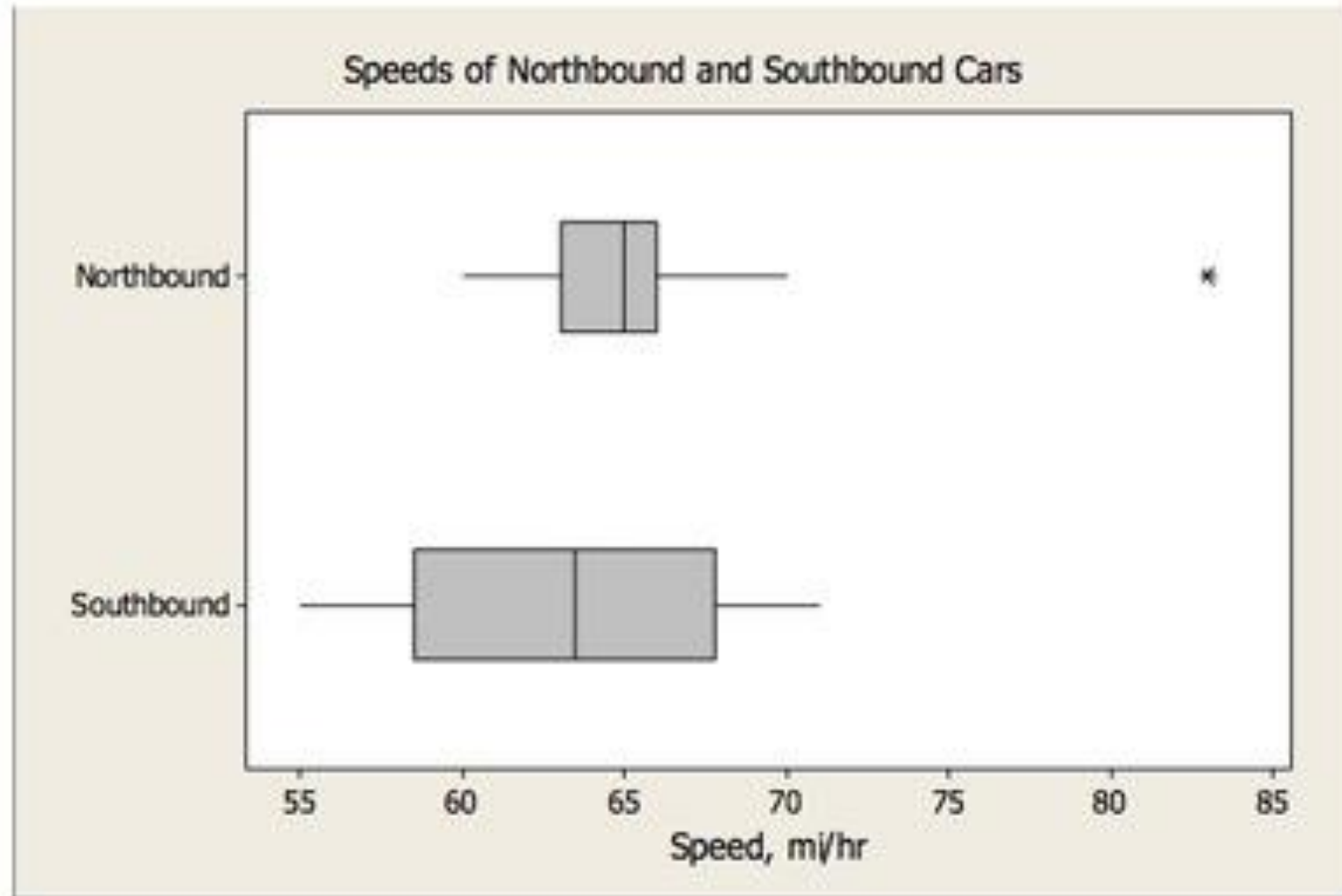
- Minimum, Q1, Median (Q2), Q3, Maximum



Measure of center: **Median (Q2)**

Measures of dispersion:  
**Range** (max – min)  
**IQR** (Q3-Q1)

# Visualizing 5-number summaries: Box plots



## Rule of thumb

---

- When should we use mean or median as our center?
- Mean is often used when the data is symmetrical
- Median is often used when the data is skewed, why?
- When we use mean as the center, we use standard deviation to describe the spread since the calculation of standard deviation is based on the mean.
- When we use median as the center, we use IQR to describe the spread since IQR is calculated using the 5 number summary.

# Homework

---

- Worksheet, see Teams