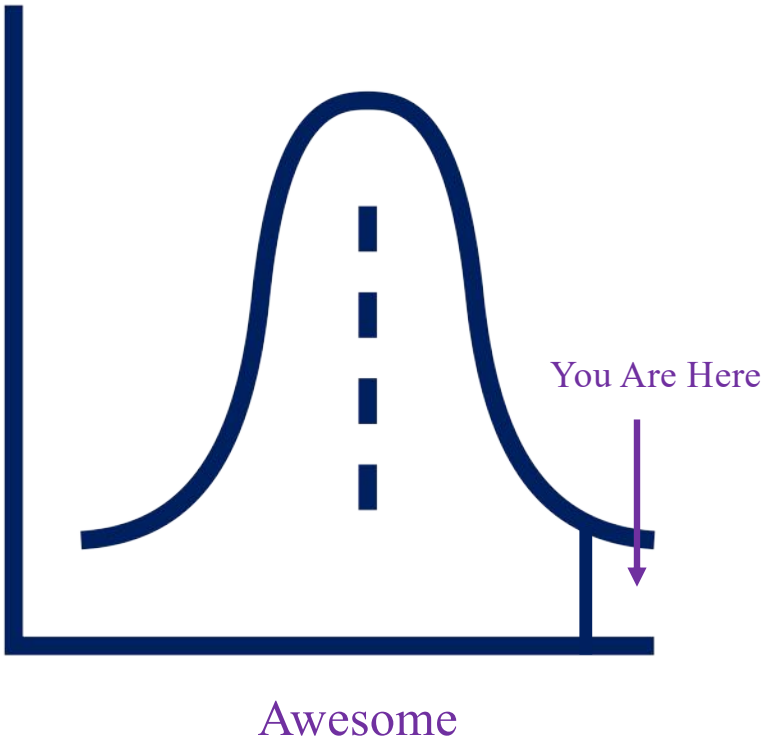


Quantitative Reasoning with Descriptive Statistics

Karyna Pryiomka, MPhil

A few words about me

- Karyna Pryiomka
- Doctoral Candidate in Social/Personality Psychology at the Graduate Center, CUNY
- Lecturer (Psychology Department) at Lehman College, CUNY

Two Categories of Statistics

- **Descriptive Statistics**—tools and procedures used to summarize, organize, and simplify data
- **Inferential Statistics**—tools and procedures used to study samples to make generalization about populations

Types of Descriptive Statistics

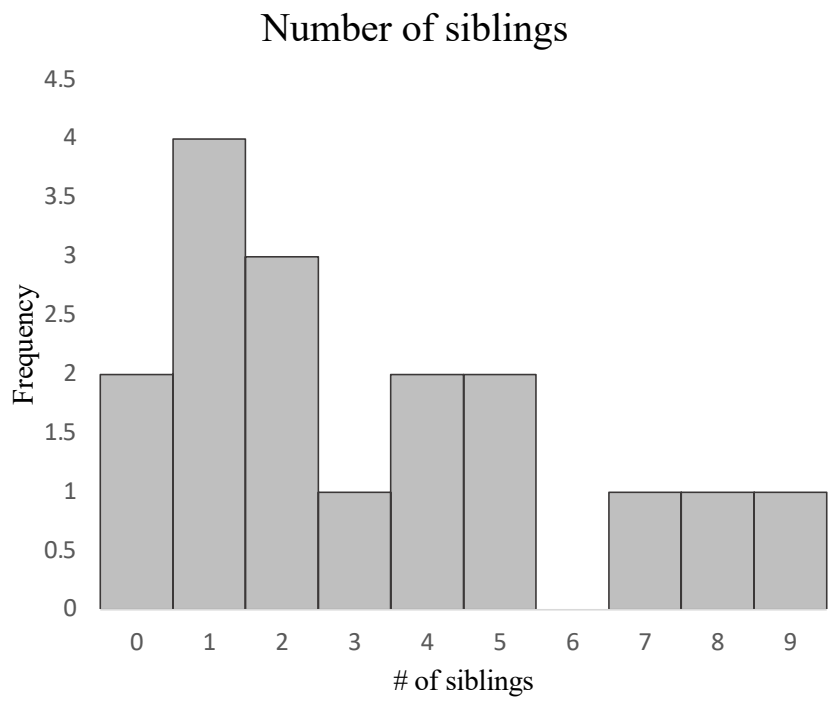
- Frequency Distributions
  - Tables and Graphs
- Measures of Central Tendency
  - Mean
  - Median
  - Mode
- Measures of Variability
  - Range
  - Variance
  - Standard Deviation

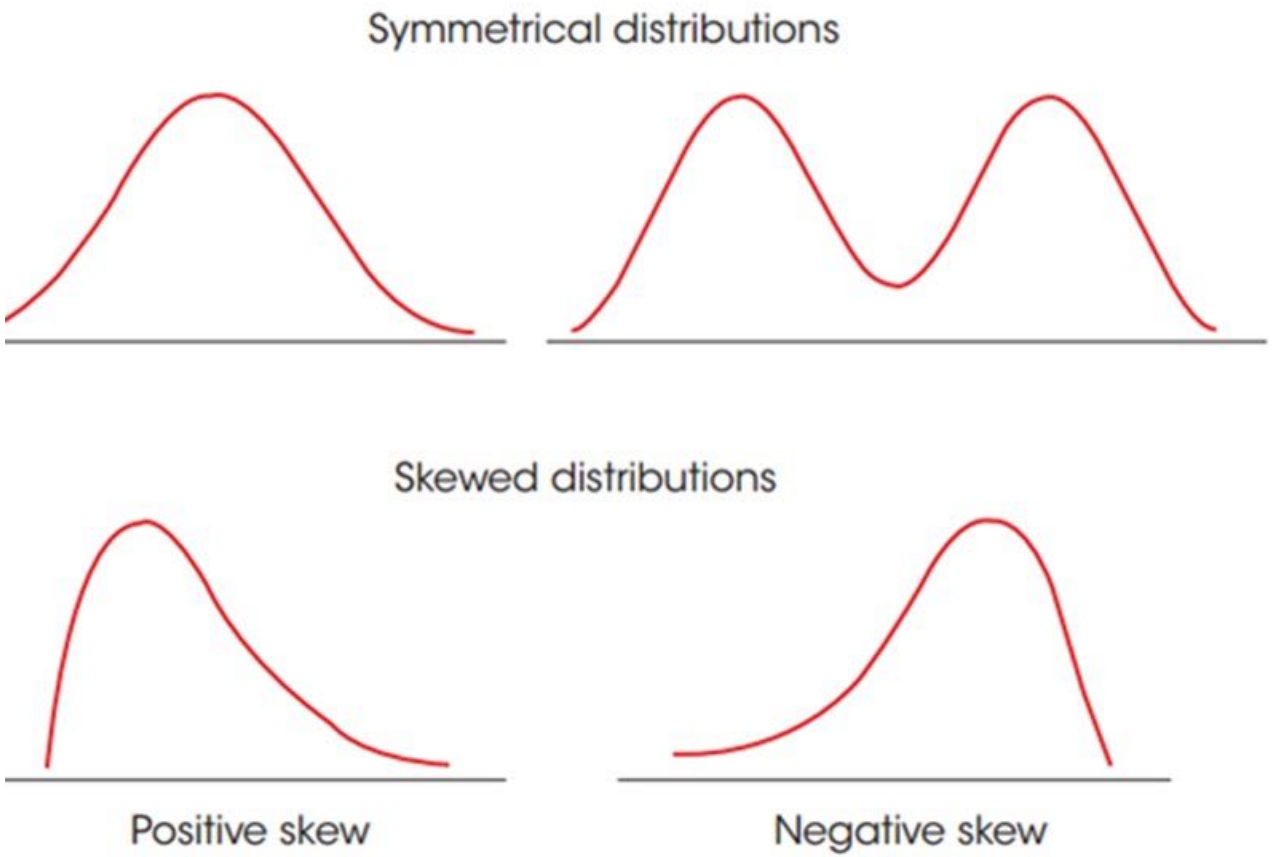
What is a frequency distribution?

- A **frequency distribution** is an organized tabulation showing the number of individuals located in each category on the scale of measurement.
- In other words, it is a summary of how many people fall into each category of the variable that you are measuring.
- Frequency distributions can take a form of a table or a graph.

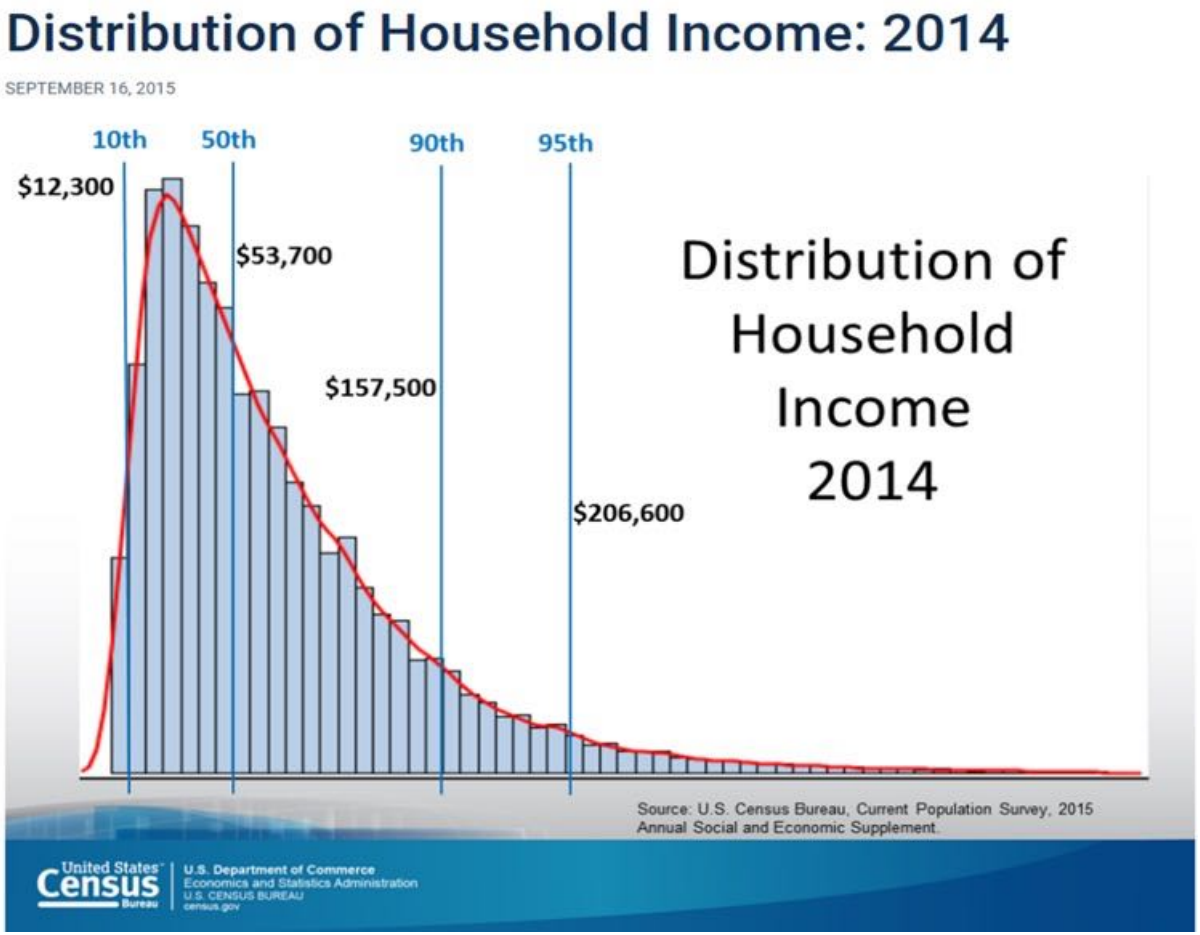
How many siblings do you have?

# of siblings (X)	f	p	%
9	1	0.0588	5.88
8	1	0.0588	5.88
7	1	0.0588	5.88
6	0	0.0000	0.00
5	2	0.1176	11.76
4	2	0.1176	11.76
3	1	0.0588	5.88
2	3	0.1765	17.65
1	4	0.2353	23.53
0	2	0.1176	11.76



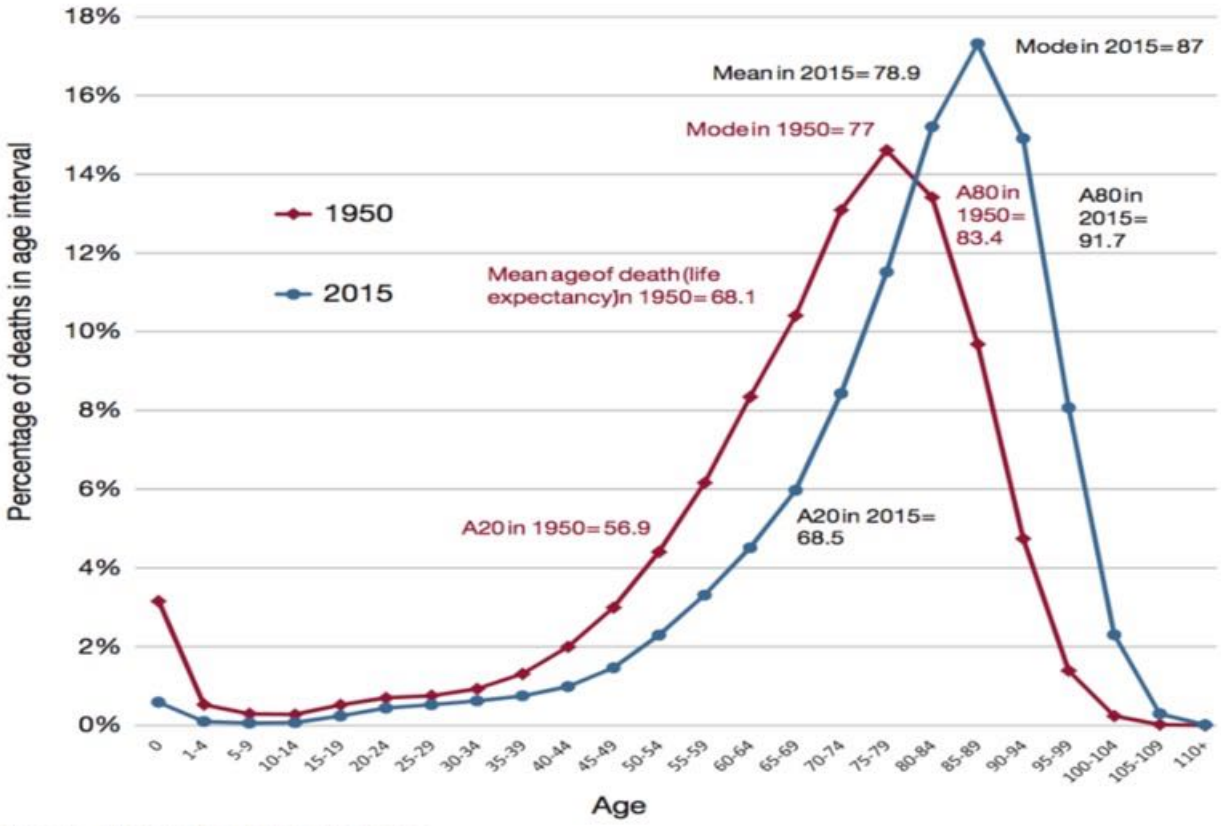


Distribution Shape



https://www.census.gov/library/visualizations/2015/demo/distribution-of-household-income--2014.html

Figure 1. Distribution of Age at Death in the United States, 1950 and 2015



Source: Human Mortality Database.

Measures of Central Tendency

Example Dataset: 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5

- Mean is the sum of all the scores divided by the number of scores in the data
  - Mean = (1+ 1 + 1 + 2 + 2 + 2 + 3 + 3 + 3 + 3 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 5)/19 = 3.05

- Median is the midpoint of the scores in a distribution when they are **listed in order** from smallest to largest
  - Median = 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5

- Mode is the score or category that has the greatest frequency of any score in the frequency distribution.
  - Mode = 4

Which measure of central tendency should I use?

Step 1: Consider scales of measurement

- Nominal – named-based categories only (ex: eye color, academic major)
  - Mode only
- Ordinal – ranked categories (ex: shirt size S, M, L)
  - Mode and Median
- Interval/Ratio – numeric categories (ex: temperature, weight, height, income in $).
  - Mean and Median is preferred

Not your average American

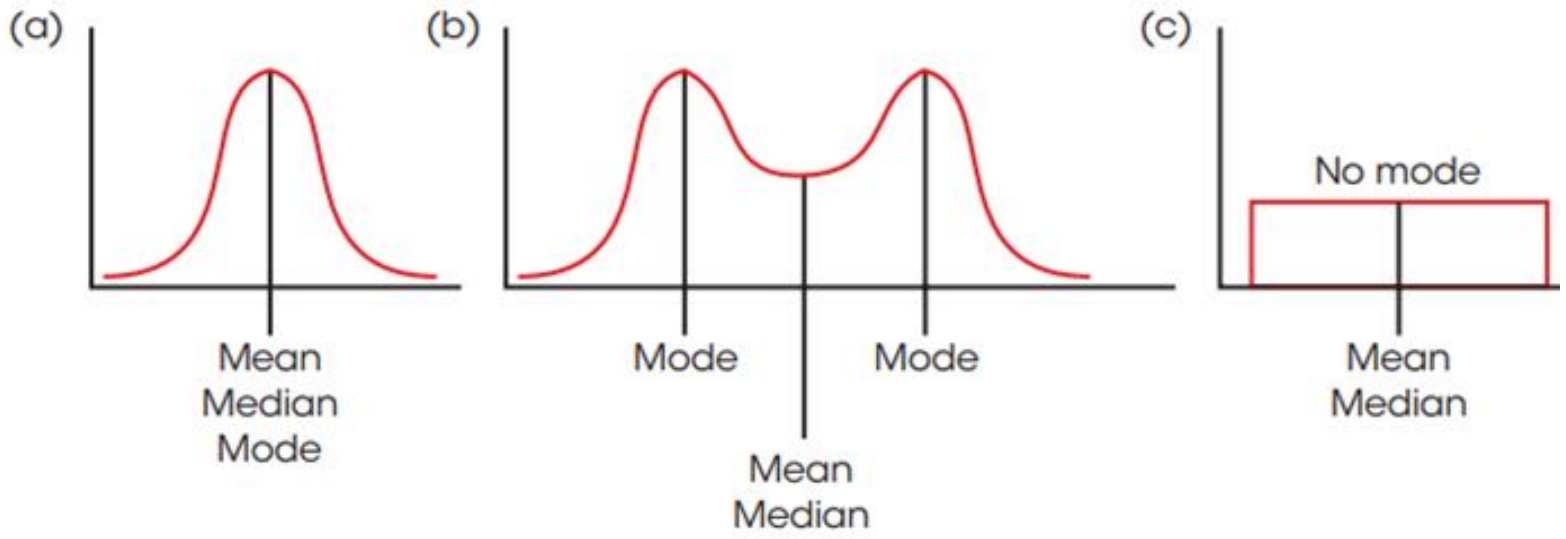
“MALONE: My issue is with this thing that people talk about, the average American. And, like, yeah, there's something sort of silly about the average American 'cause it's like a melting pot. Who is average, really? But my quibble is more pedantic than that because I think people don't mean average American. I think what they actually mean is, who is the person - if I walk outside into America, who is the person I'm most likely to run into?”

-Mode is typically the most useful statistic when we are trying to identify a typical person, customer, client.

The Modal American by NPR Planet Money (https://www.npr.org/2019/08/28/755191639/episode-936-the-modal-American)

Step 2: Consider distribution shape

Symmetrical Distributions



The Romanoffs - Season 1

2018 | TV Parental Guideline Rating: TV-MA

1,250

Prime Video

$0.00 with a Prime membership

The Romanoffs - Season

2018 | TV Parental Guideline R

1,250

Prime Video

$0.00 with a Prime membership

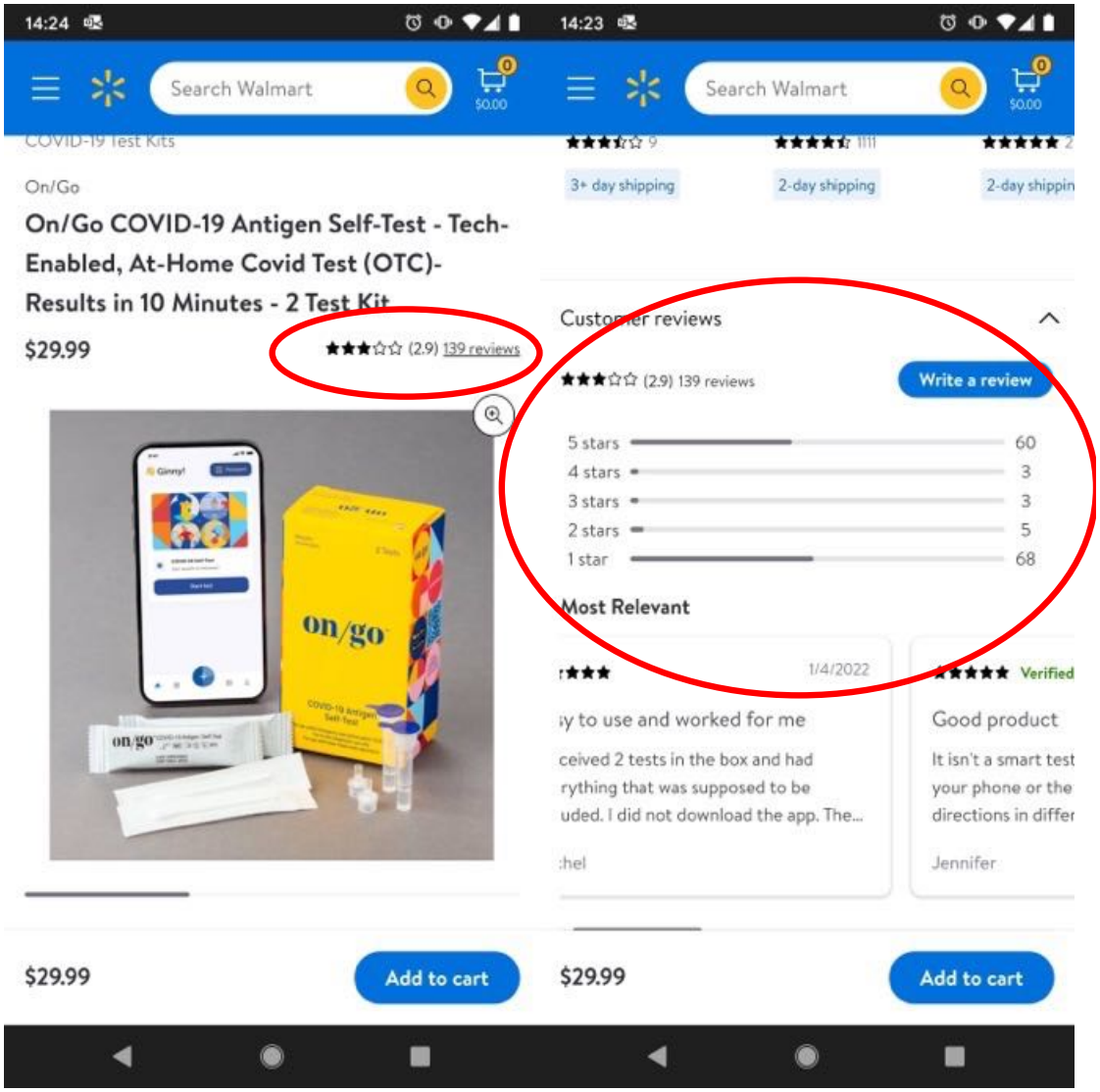
Customer reviews

3.3 out of 5

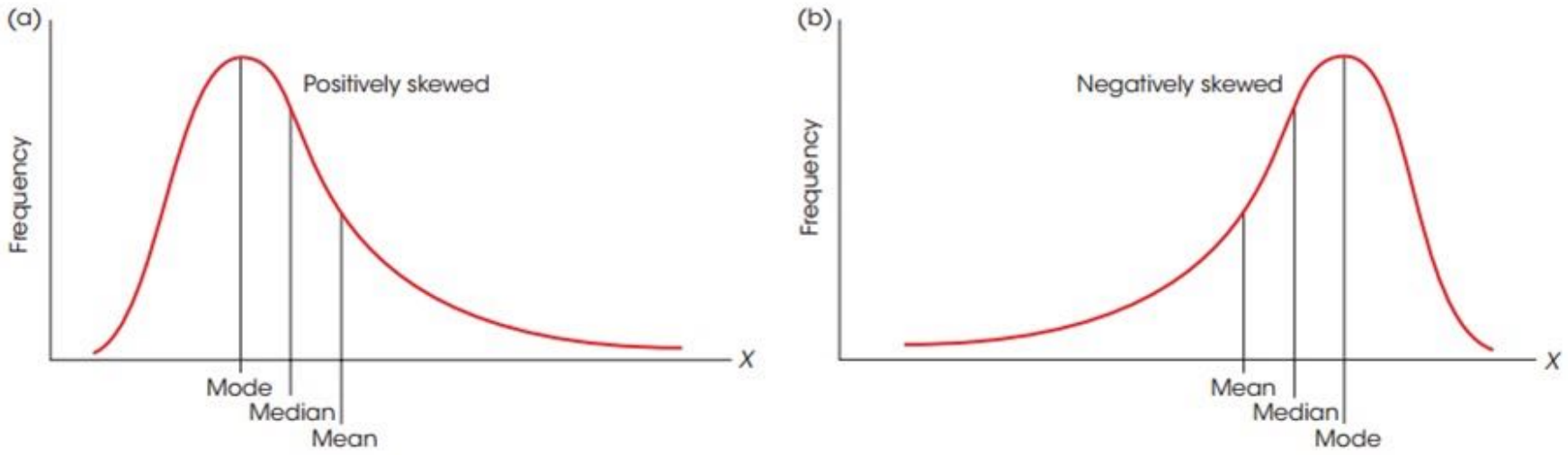
1,250 customer ratings

5 star	44%
4 star	9%
3 star	5%
2 star	13%
1 star	29%

Always look at the distribution shape!



Skewed Distributions and Measures of Central Tendency

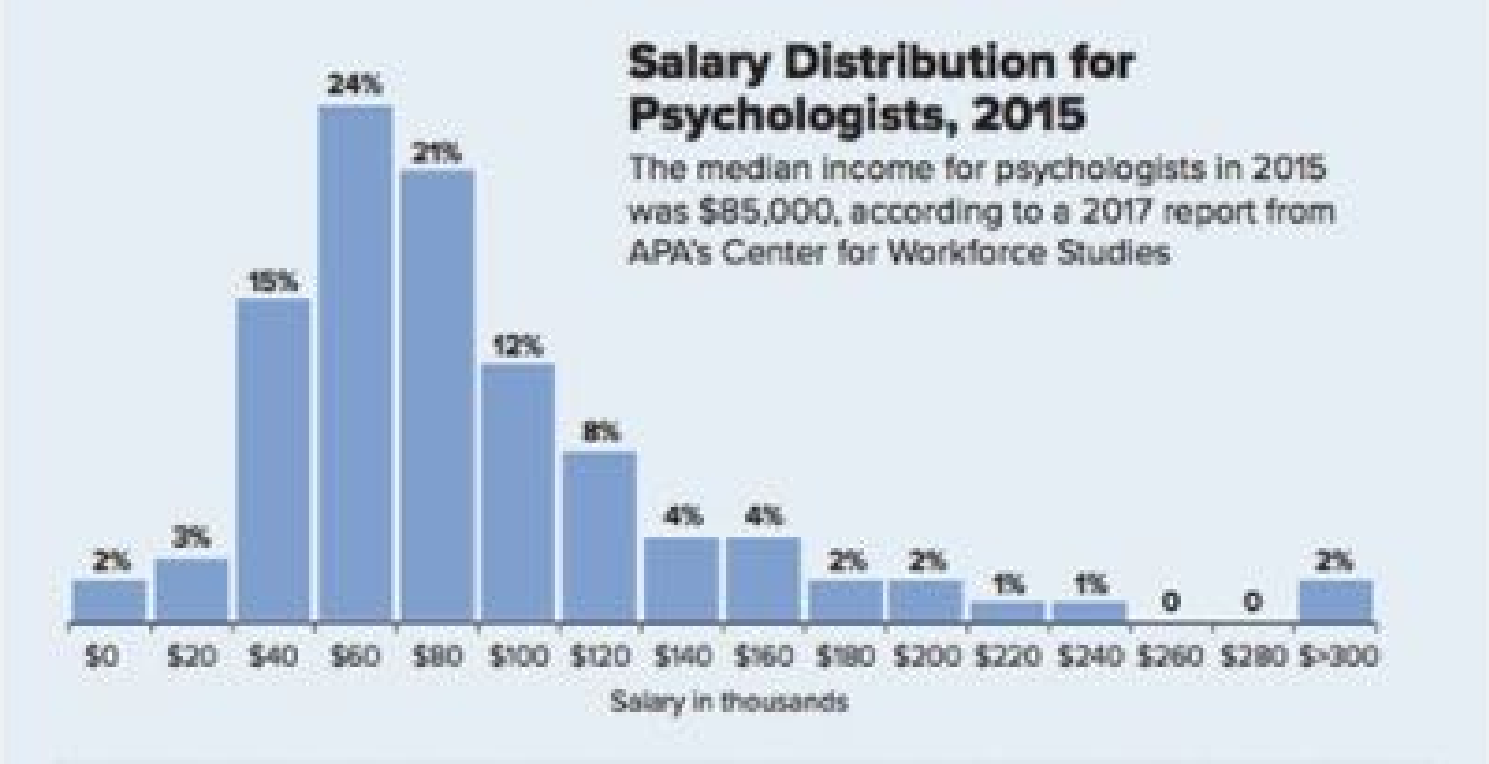


Consider Salary Distributions (Psychologists)

Occupation code	Occupation title (click on the occupation title to view its profile)	Level	Employment	Employment RSE	Employment per 1,000 jobs	Median hourly wage	Mean hourly wage	Annual mean wage	Mean wage RSE
19-3030	Psychologists	broad	127,070	1.6%	0.913	$39.51	$43.61	$90,710	0.9%
19-3031	Clinical, Counseling, and School Psychologists	detail	111,320	1.5%	0.800	$38.37	$42.93	$89,290	1.1%
19-3032	Industrial-Organizational Psychologists	detail	780	13.3%	0.006	$46.28	$54.18	$112,690	5.8%

https://www.bls.gov/oes/current/oes_nat.htm#19-0000

Consider Salary Distributions (Psychologists)



https://www.apa.org/monitor/2017/09/psychologists-earn

Consider Salary Distributions (Legal Professions)

Occupation code	Occupation title (click on the occupation title to view its profile)	Level	Employment	Employment RSE	Employment per 1,000 jobs	Median hourly wage	Mean hourly wage	Annual mean wage	Mean wage RSE
23-0000	**Legal Occupations**	major	1,154,740	0.5%	8.302	$40.82	$54.00	$112,320	0.7%
23-1000	Lawyers, Judges, and Related Workers	minor	721,750	0.6%	5.189	$59.79	$69.70	$144,970	0.7%
23-1010	Lawyers and Judicial Law Clerks	broad	672,820	0.7%	4.837	$60.12	$70.70	$147,050	0.7%
23-1011	Lawyers	detail	658,120	0.7%	4.731	$61.03	$71.59	$148,910	0.7%
23-1012	Judicial Law Clerks	detail	14,690	1.3%	0.106	$27.65	$30.70	$63,860	2.7%
23-1020	Judges, Magistrates, and Other Judicial Workers	broad	48,930	0.6%	0.352	$55.36	$55.96	$116,390	1.0%

https://www.bls.gov/oes/current/oes_nat.htm#19-0000

Consider Salary Distributions (Legal Professions)

Note: Graph is based on 18,536 salaries reported for full-time jobs lasting a year or more — a few salaries above $210,000 are excluded from the graph for clarity, but not from the percentage calculations. The left-hand peaks of the graph reflect salaries of $45,000 to $75,000, which collectively accounted for nearly half (49.3%) of reported salaries. The right-hand peak shows that salaries of $190,000 accounted for 21.3% of reported salaries. However, more complete salary coverage for jobs at large firms heightens this right-hand peak and diminishes the left-hand peaks, and as a result the unadjusted mean overstates the average starting salary by an estimated 5.6%. Nonetheless, as both the unadjusted arithmetic mean and the adjusted mean show, relatively few salaries are close to either mean. For purposes of this graph, all reported salaries were rounded to the nearest $5,000. However, the mean is based on salaries as reported and the adjusted mean reflects a weighting of those salaries.

https://www.nalp.org/salarydistrib#2020

Step 3: Consider underlying data

- Extreme values/outliers
  - 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 100

- Undetermined values
  - How many hours a week do you study?
  - 4-5, 6-7, about 15.

- Open-ended distributions

X	**f**
4 and over	4
3	5
2	0
1	7

Selecting a Measure of Central Tendency

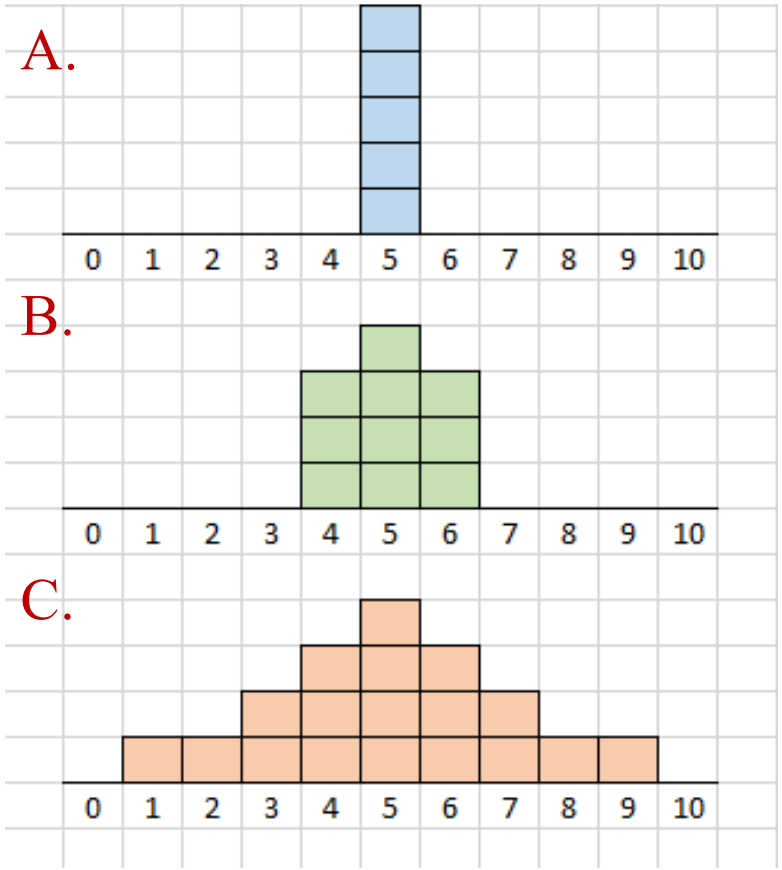
Measure of Central Tendency	Appropriate to choose if …	Could be misleading if…
Mean	•You can calculate $\sum X$ •You know the value of every score	•Extreme scores •Skewed distribution •Undetermined values •Open-ended distribution •Ordinal scale •Nominal scale
Median	•Extreme scores •Skewed distribution •Undetermined values •Open-ended distribution •Ordinal scale	•Nominal scale
Mode	•Nominal scales •Discrete variables •Describing shape	•Interval or ratio data, except to accompany mean or median

Essentials of Statistics for the Behavioral Sciences, 9th edition, Authors: Gravetter & Wallnau.

A few words about data variability

Josh wants to know if playing music during quizzes affects students' performance. Josh teaches three sections of the same class, so he decides to play classical music in one class, light instrumental rock in another, and no music in his third class. Below are his results. Does it look like music had an effect on quiz scores in Josh's classes?

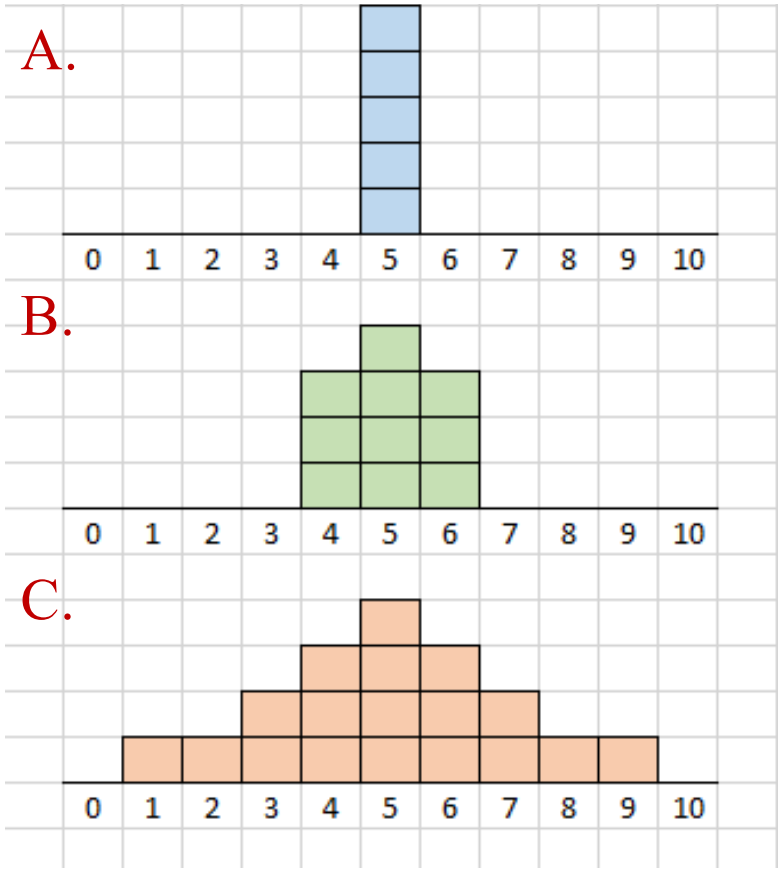
$M_{music_class} = 5$ $M_{music_rock} = 5$ $M_{no_music} = 5$



Be Careful

Identical mean scores do not always indicate identical distributions.

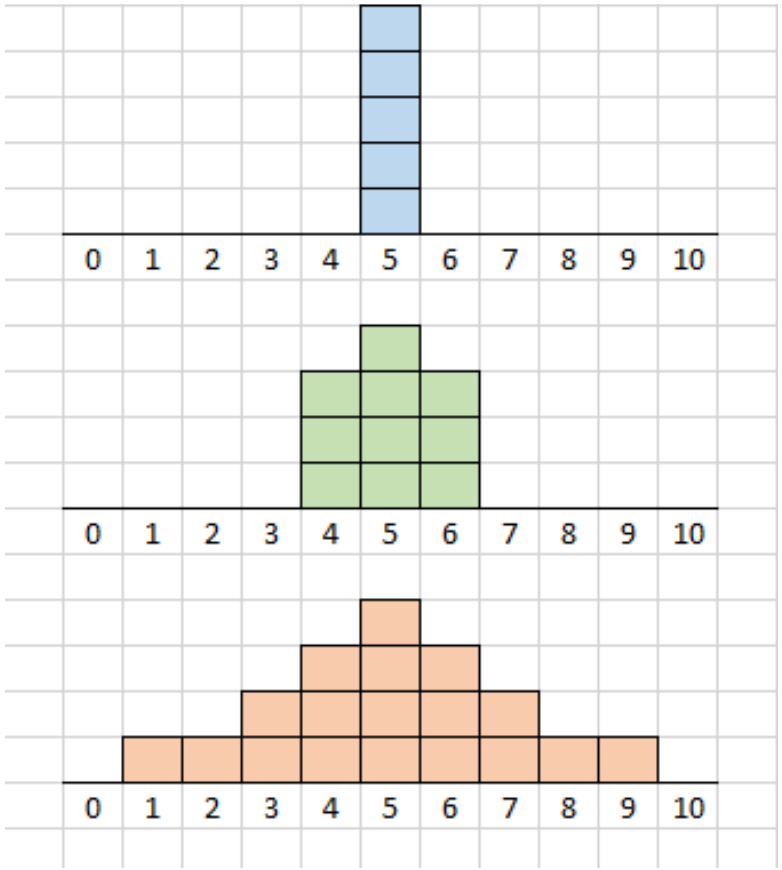
Which class would you rather be in?



Consider your goals

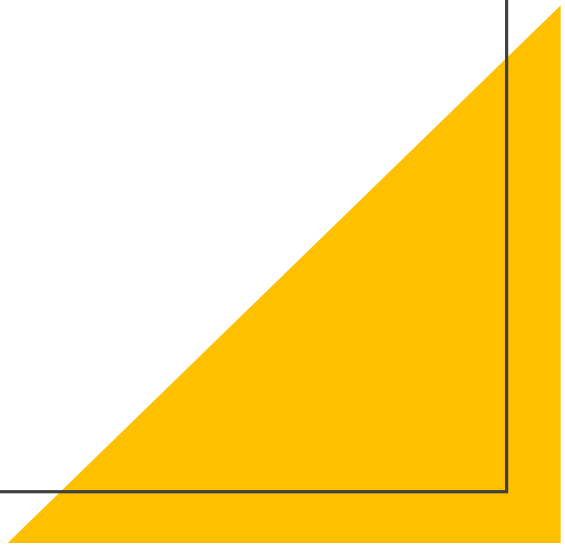
Assuming that 5 or higher is a passing score,

which class would you rather be in?



If the scores are very spread out (high variability), each individual score provides a poor representation of the entire distribution.

Summary

- Descriptive statistics play a vital role in our ability to accurately draw conclusions from data.
- When analyzing data, make sure to explore key descriptive statistics for each variable or combinations of variables.
- Do NOT skip the graphs
- When consuming data-based information, look for missing details. Use your knowledge of descriptive statistics to ask critical questions. Demand to see the graph☺

Thank You!