GENERAL ASSEMBLY

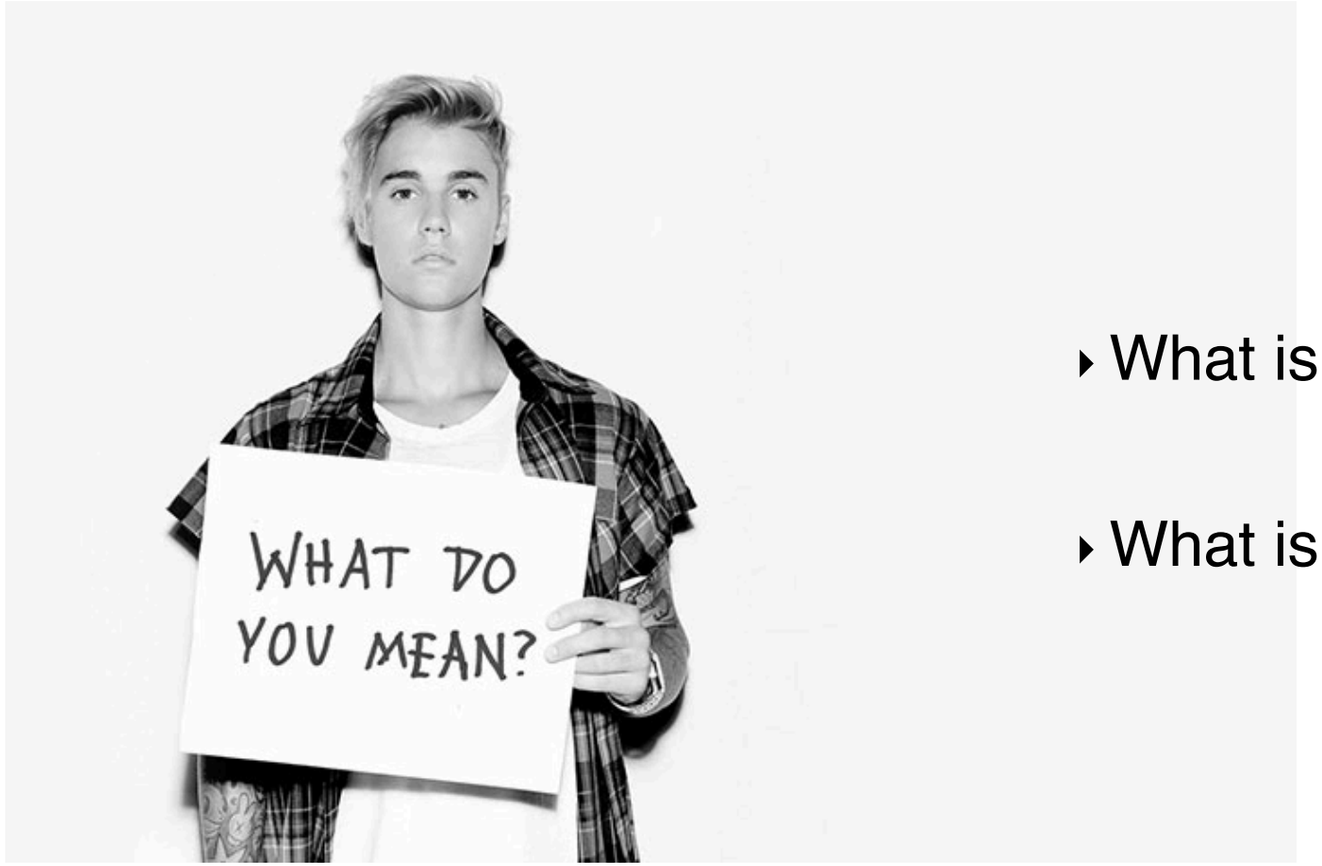
STORYTELLING WITH DATA

Joseph Nelson, Data Science Immersive

BASIC DESCRIPTIVE STATISTICS

- Mean
- Median
- Mode
- Max
- Min
- Quartile
- Inter-quartile Range
- Variance
- Standard Deviation
- Correlation

MEAN



- What is the mean?
- What is another name for the mean?

MEAN

- What is the mean?
- The mean of a set of values is the sum of the values divided by the number of values. It is also called the average.
- It is also known as the average.
- Example: Find the mean of 19, 13, 15, 25, and 18

MEDIAN

- What is the median?
- How do you find the median?

MEDIAN

- What is the median?
- How do you find the median?
- Bonus: Why might the median be advantageous instead of the mean? When does this condition NOT hold?

MEDIAN

- The median refers to the midpoint in a series of numbers.

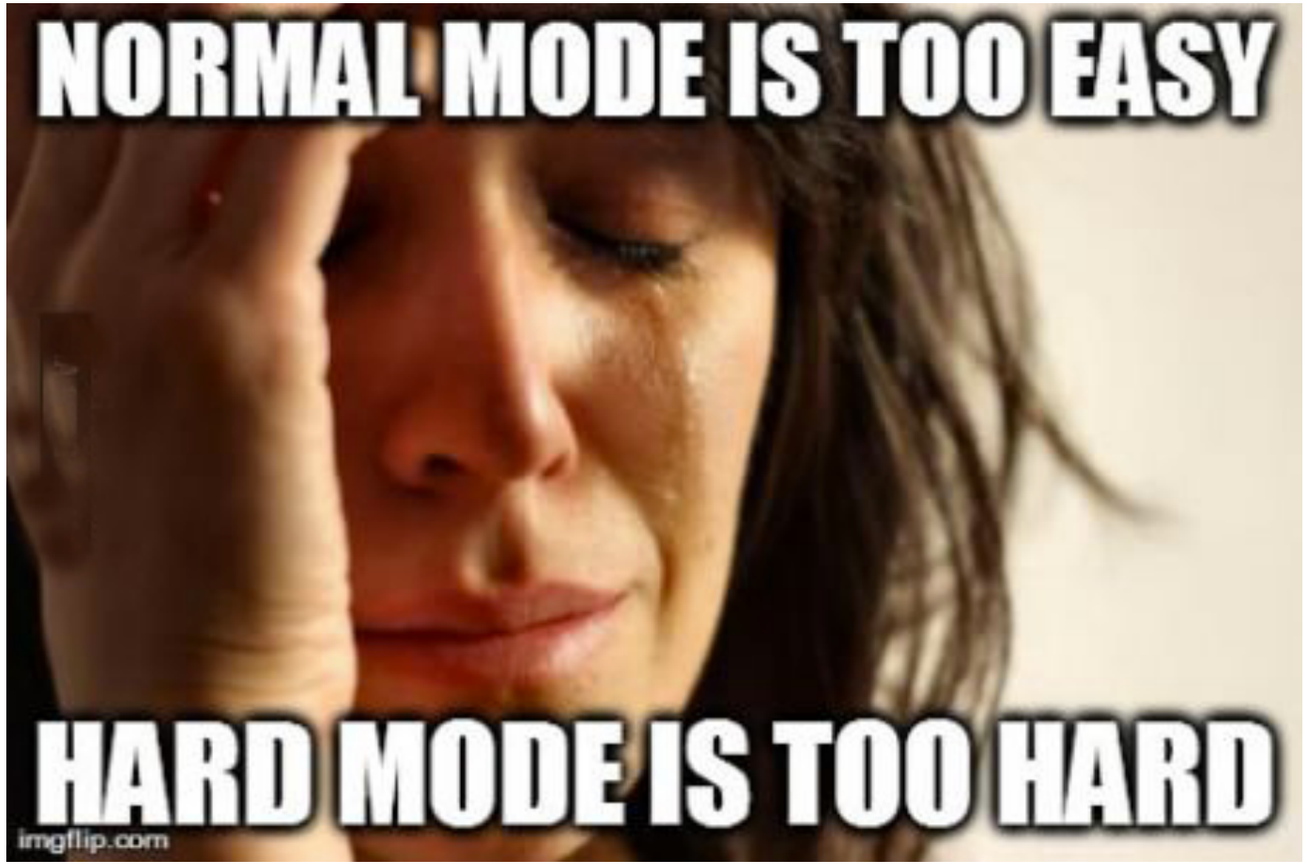
- To find the median, arrange the numbers in order from smallest to largest. If there is an odd number of values, the middle value is the median. If there is an even number of values, the average of the two middle values is the median.

MEDIAN

- The median refers to the midpoint in a series of numbers.
- Example #1: Find the median of 19, 29, 36, 15, and 20
- Example #2: Find the median of 67, 28, 92, 37, 81, 75
- Bonus: Median may be more useful than average in a highly skewed population.

MODE

- What is the mode?
- What is the mode in the following: 1, 2, 3, 4, 5



MODE

- What is the mode?
- The mode of a set of values is the value that occurs most often.
- A set of values may have more than one mode or no mode.

CHECK FOR UNDERSTANDING

- For the following groups of numbers, calculate the mean, median and mode by hand:

- A. 18, 24, 17, 21, 24, 16, 29, 18

- B. 75, 87, 49, 68, 75, 84, 98, 92

- C. 55, 47, 38, 66, 56, 64, 44, 39

CHECK FOR UNDERSTANDING

- Answers:

- A. Mean = 20.875 Median = 19.5 Mode = 18, 24 Max = 29 Min = 16

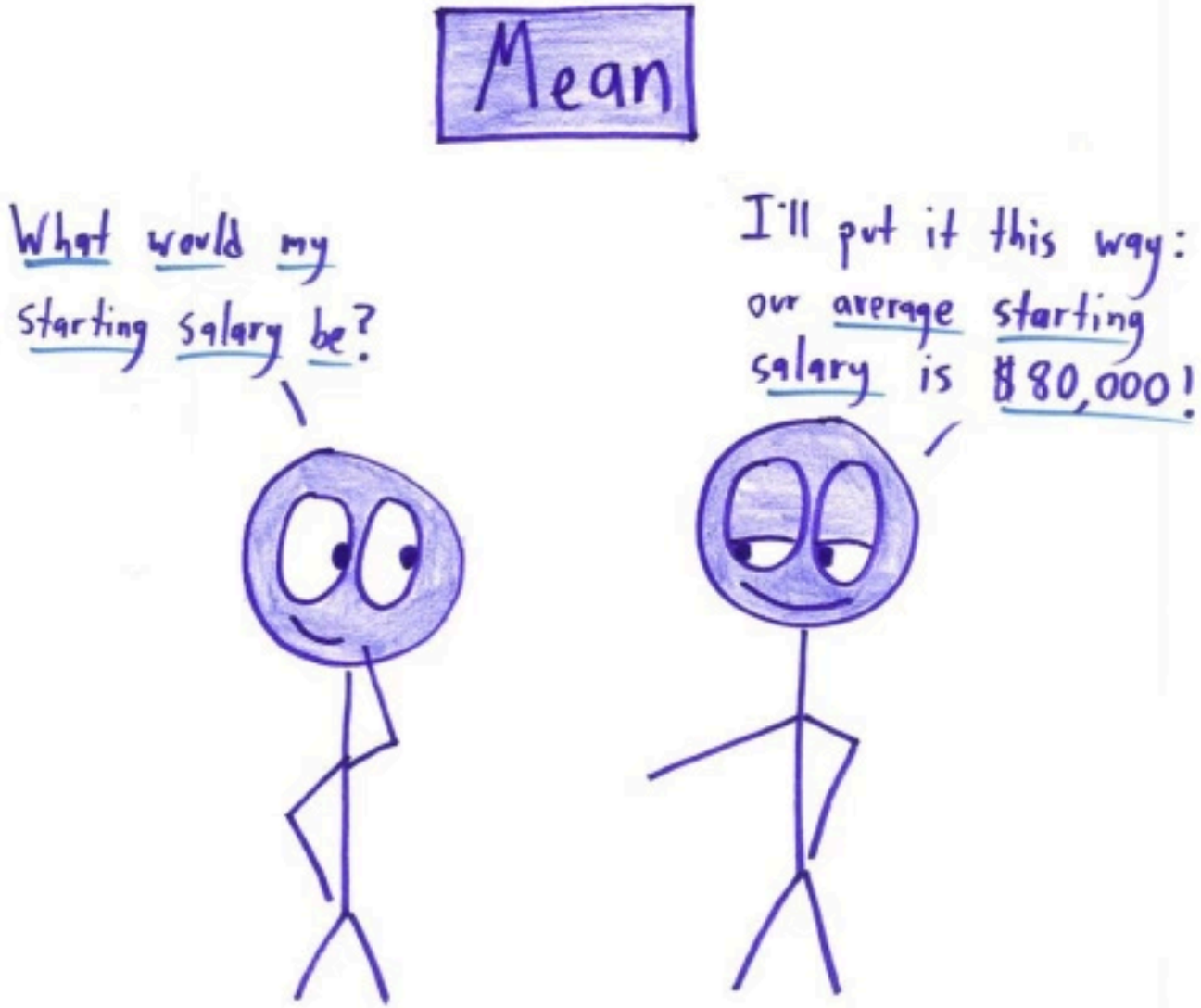
- B. Mean = 78.5 Median = 79.5 Mode = 75 Max = 98 Min = 49

- C. Mean = 51.125 Median = 51 Mode = none Max = 66 Min = 38

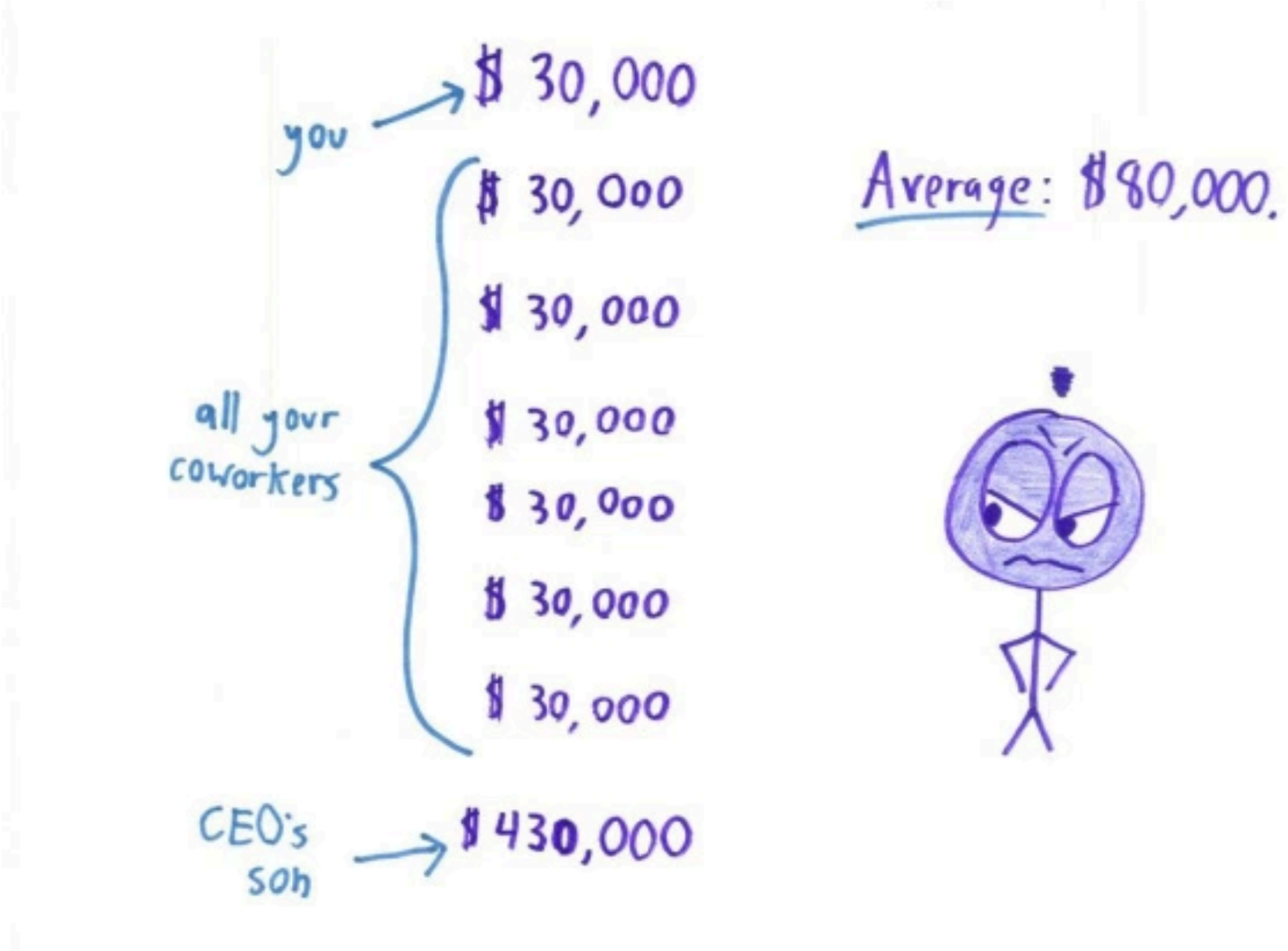
HOW TO LIE WITH STATISTICS

- For each picture:
- 1) What could go wrong
- 2) How to fix it

HOW TO LIE WITH STATISTICS

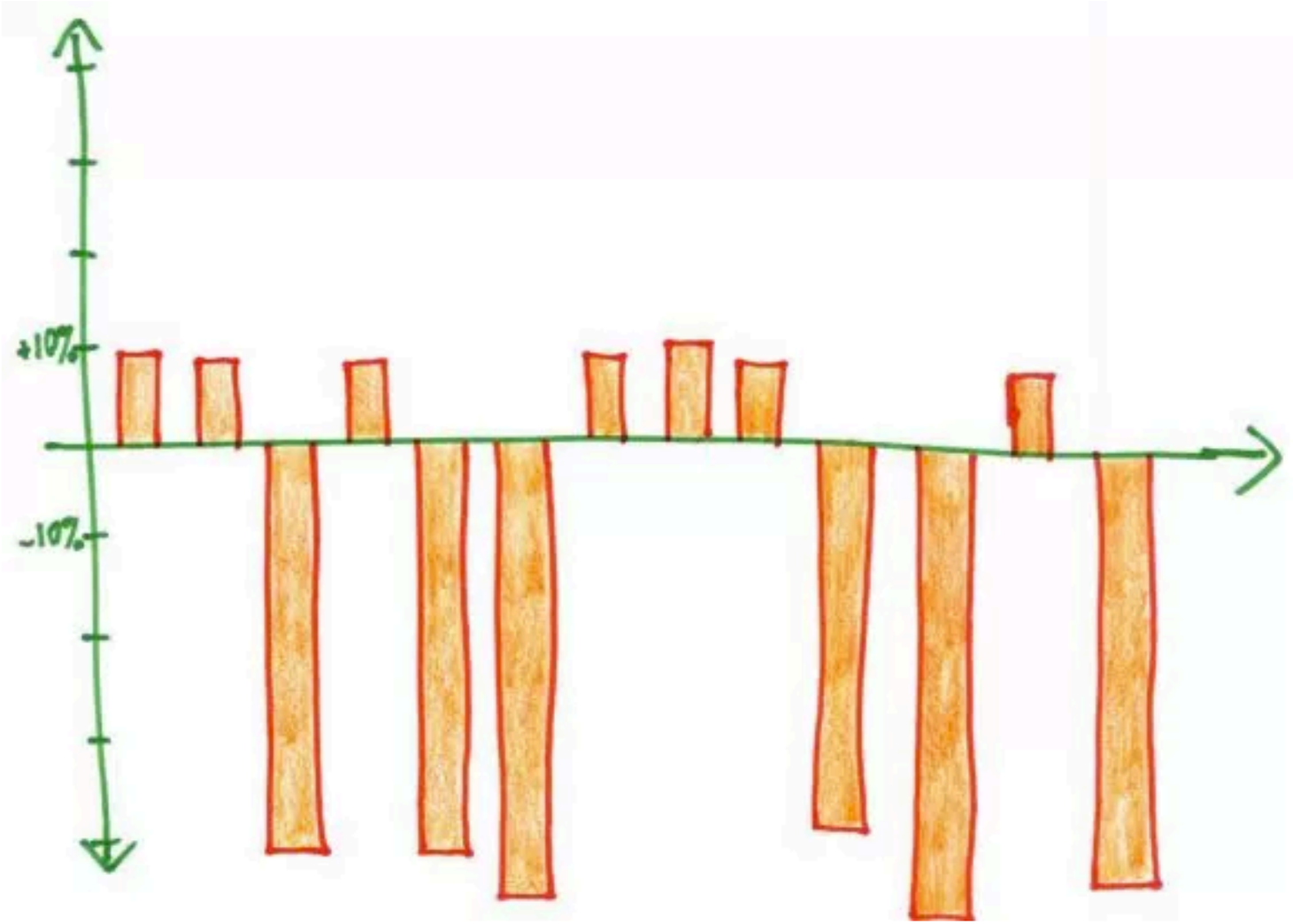


HOW TO LIE WITH STATISTICS

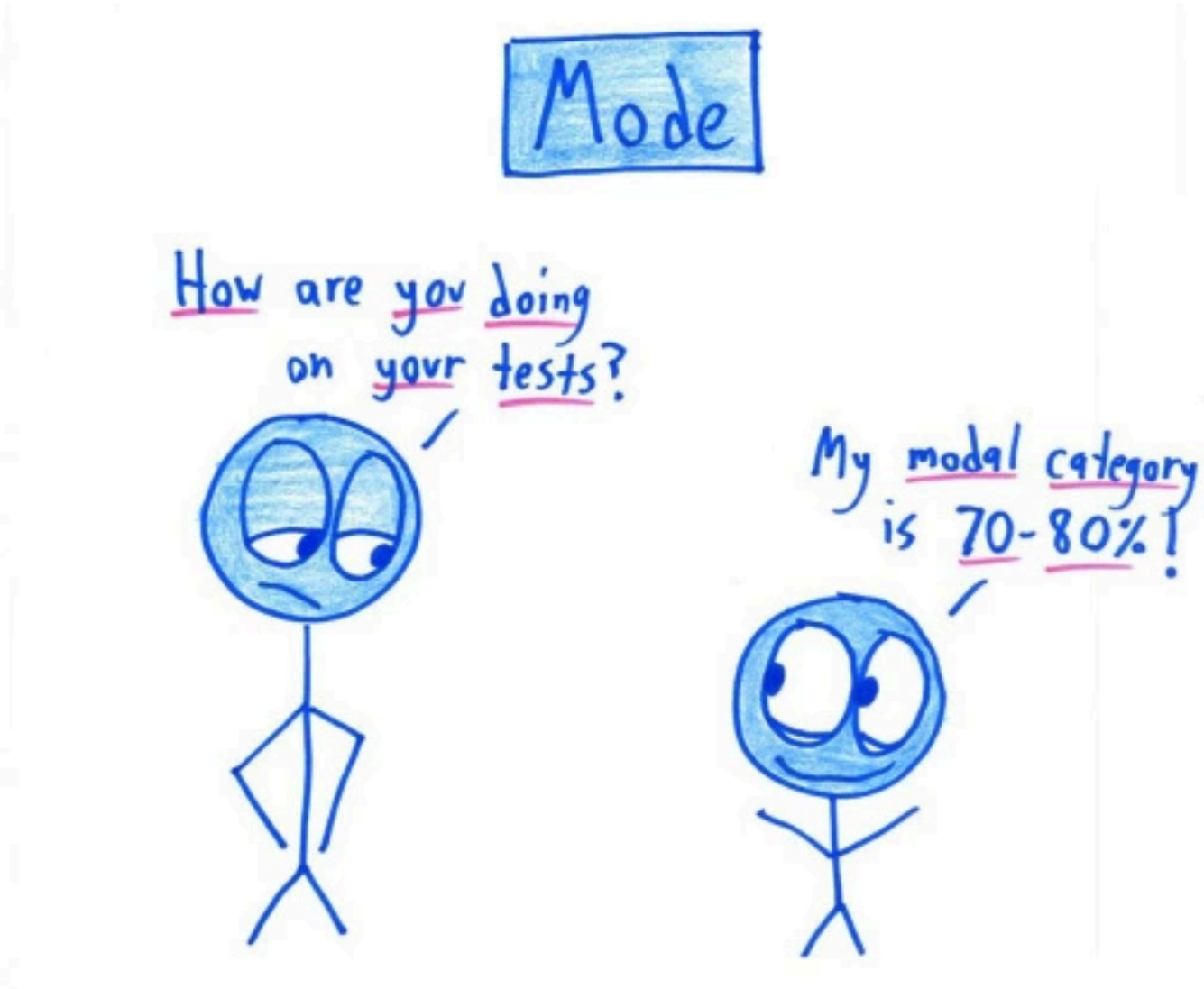


HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS



HOW TO LIE WITH STATISTICS

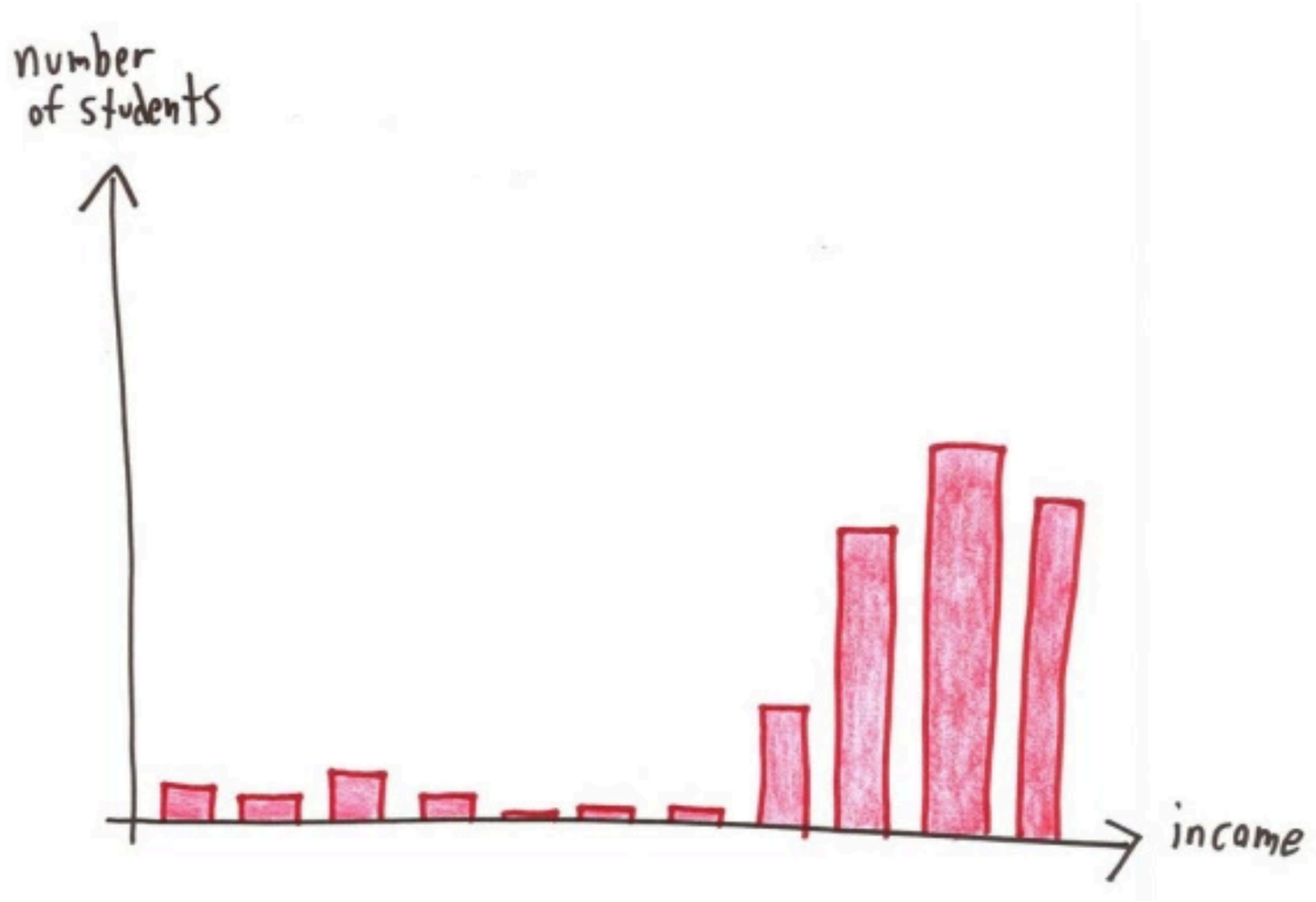


HOW TO LIE WITH STATISTICS

Score Category	Number of Tests
90s	0
80s	0
70s	2
60s	1
50s	1
40s	1
30s	1
20s	1

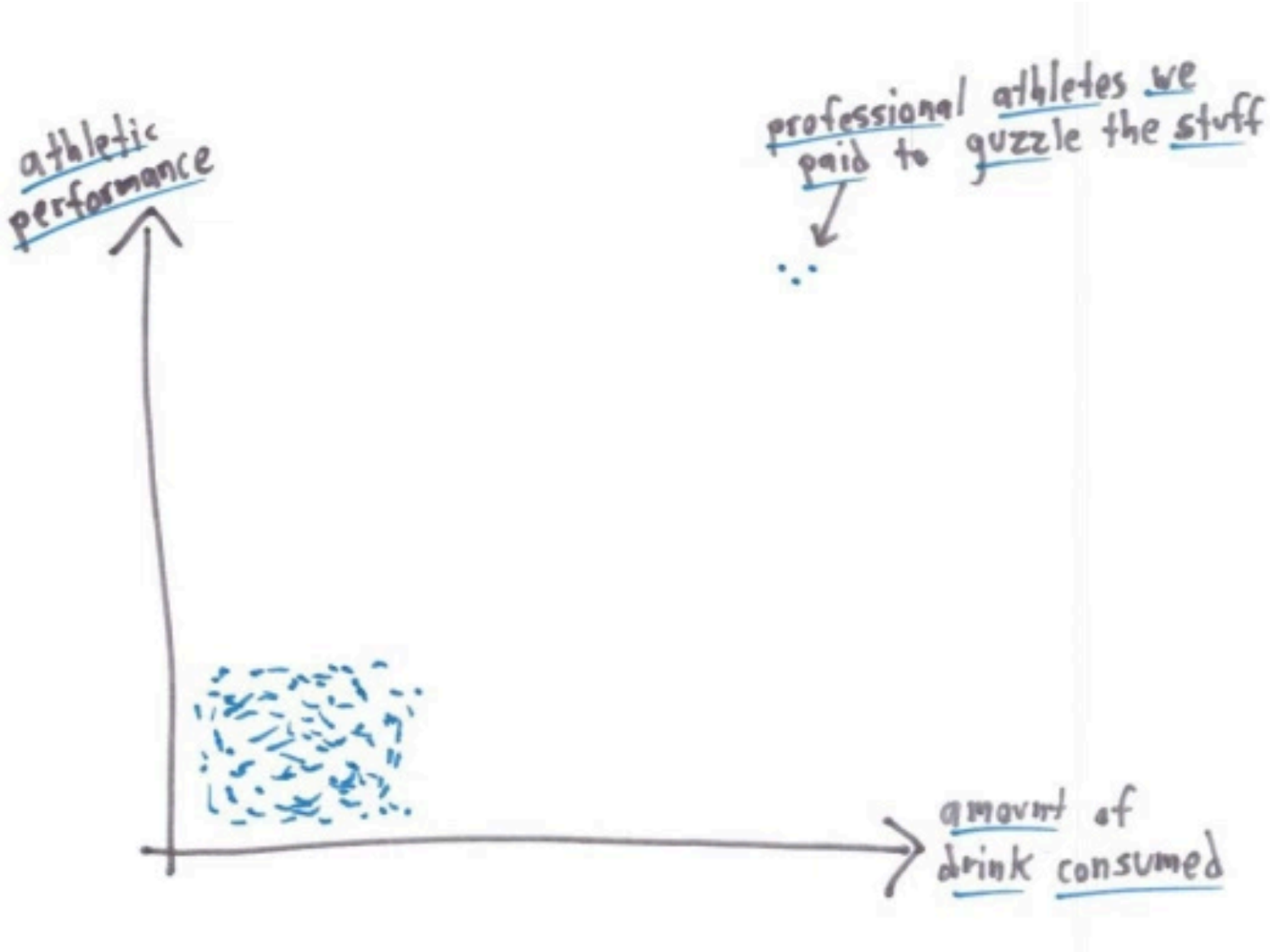
HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

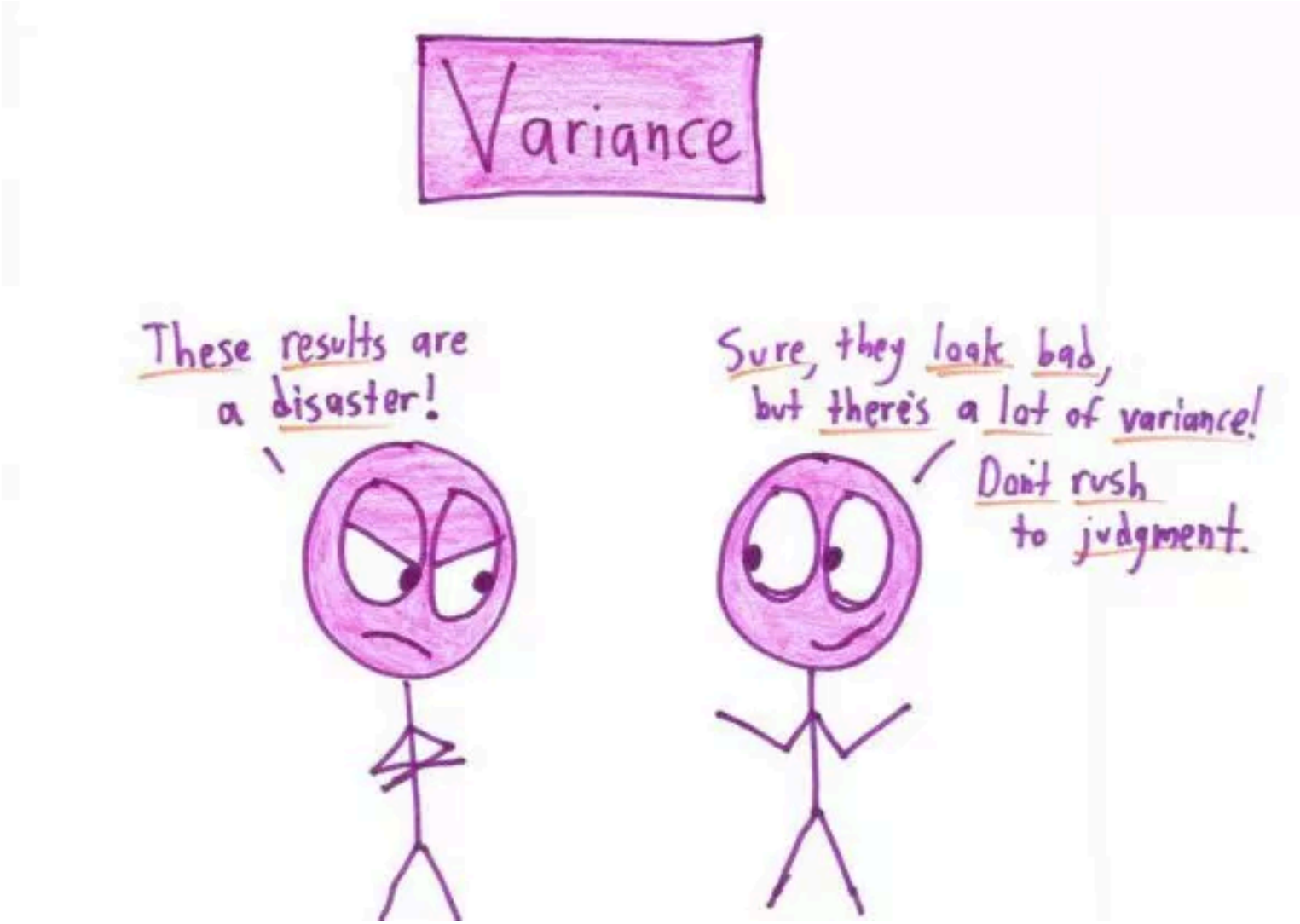


HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS



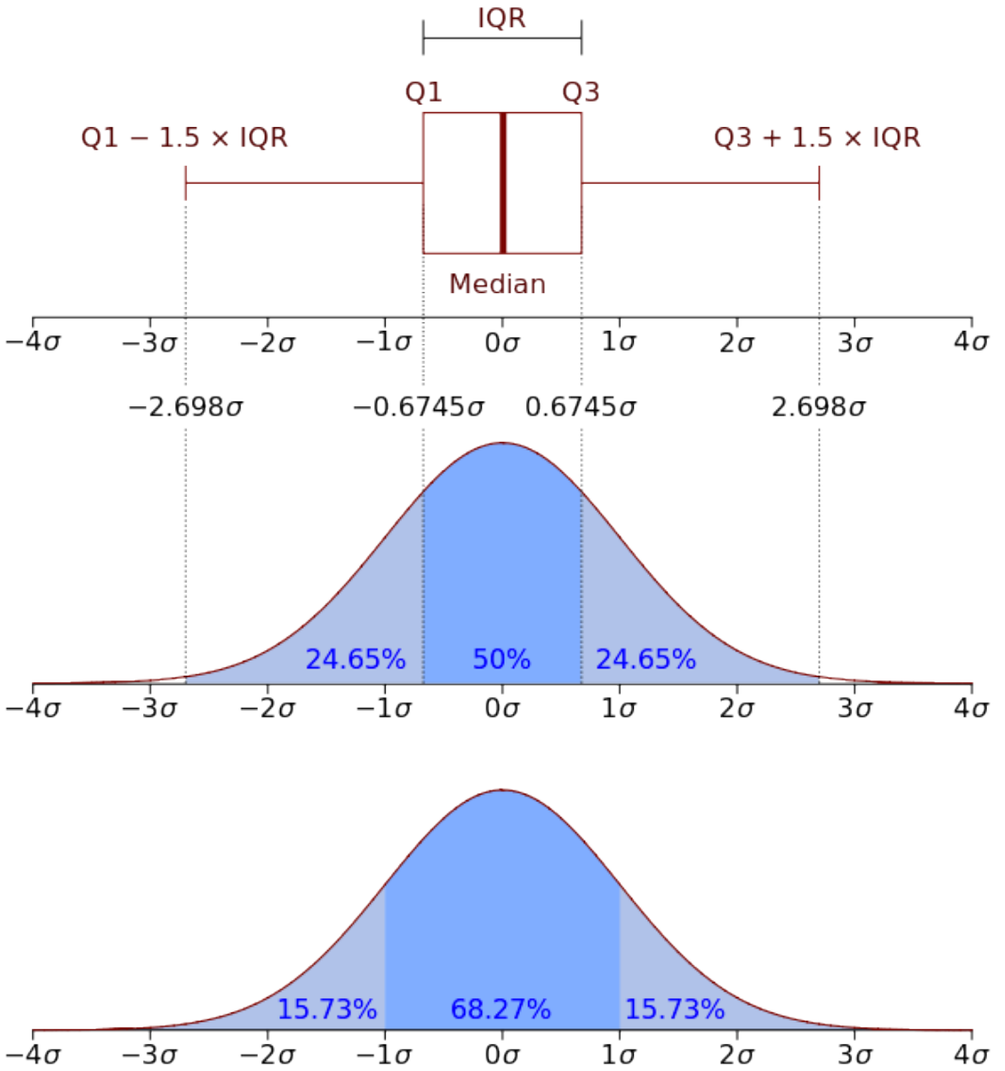
HOW TO LIE WITH STATISTICS



HOW TO LIE WITH STATISTICS

QUARTILES AND THE INTER QUARTILE RANGE

- Quartiles divide a rank-ordered data set into four equal parts.
- The values that divide each part are called the first, second, and third quartiles; and they are denoted by Q1, Q2, and Q3, respectively.
- The interquartile range (IQR) is a measure of variability, based on dividing a data set into quartiles. It is the “middle 50” of your data. Also called the H-spread. IQR = Q3-Q1
- Outliers: Q1 - 1.5(IQR), Q3 + 1.5(IQR)



CRITERIA FOR GOOD VISUALIZATION