

STORYTELLING WITH DATA

Joseph Nelson, Data Science Immersive

HOW TO LIE WITH STATISTICS

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

HOW TO LIE WITH STATISTICS

Mean

What would my
starting salary be?



I'll put it this way:
our average starting
salary is \$80,000!



HOW TO LIE WITH STATISTICS

you → \$ 30,000

all your coworkers { \$ 30,000
\$ 30,000
\$ 30,000
\$ 30,000
\$ 30,000
\$ 30,000

CEO's son → \$ 430,000

Average: \$80,000.



HOW TO LIE WITH STATISTICS

Median

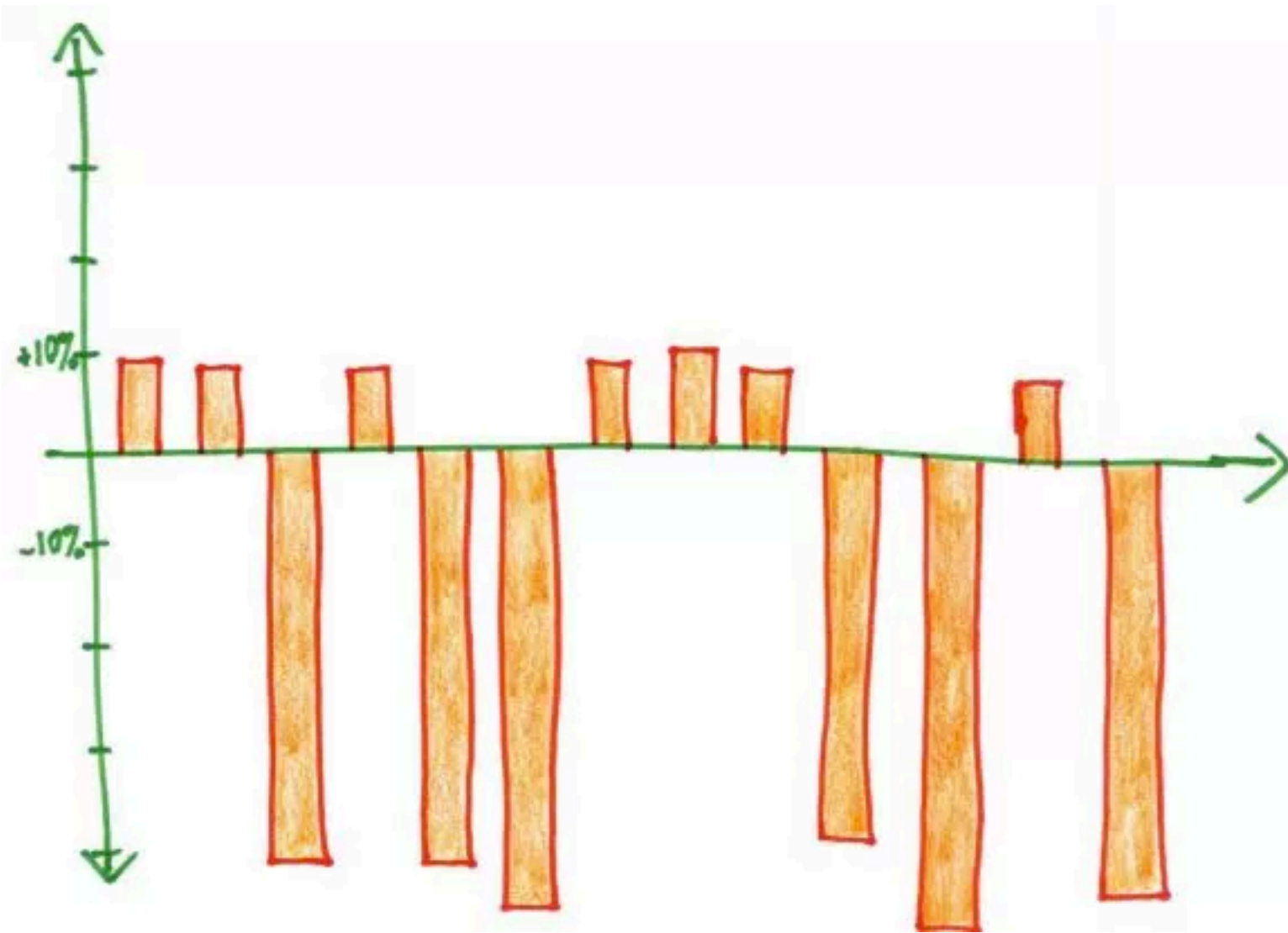
So, why should I
invest with you?



Well, not to brag, but
my fund has a median
gain of 8% per year!



HOW TO LIE WITH STATISTICS



HOW TO LIE WITH STATISTICS

Mode

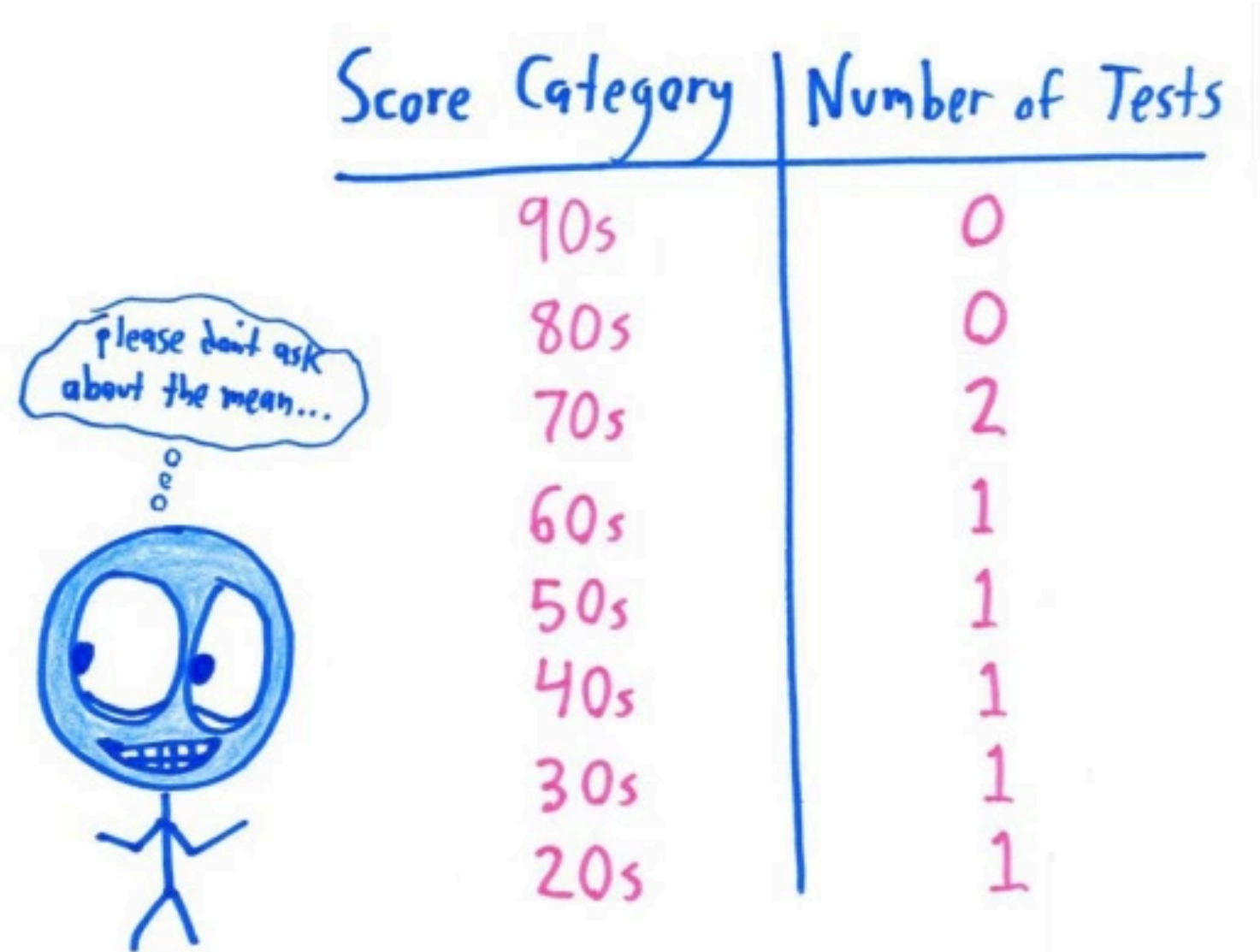
How are you doing
on your tests?



My modal category
is 70-80%!



HOW TO LIE WITH STATISTICS



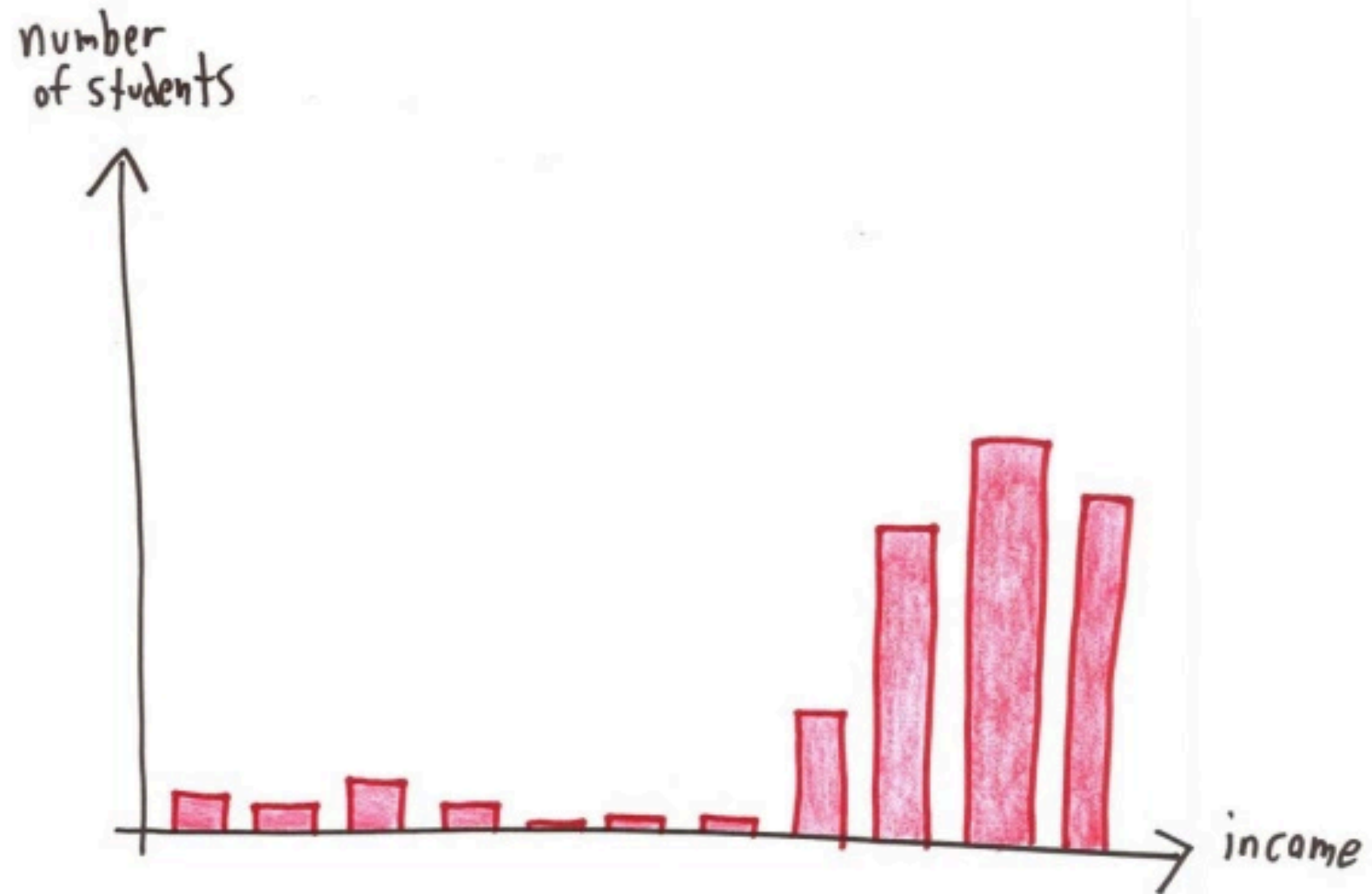
HOW TO LIE WITH STATISTICS

Range

Our students come from a
wide range of
socioeconomic
backgrounds...



HOW TO LIE WITH STATISTICS



HOW TO LIE WITH STATISTICS

Correlation
Coefficient

Try our energy drink —
it's highly correlated with
performance!



HOW TO LIE WITH STATISTICS

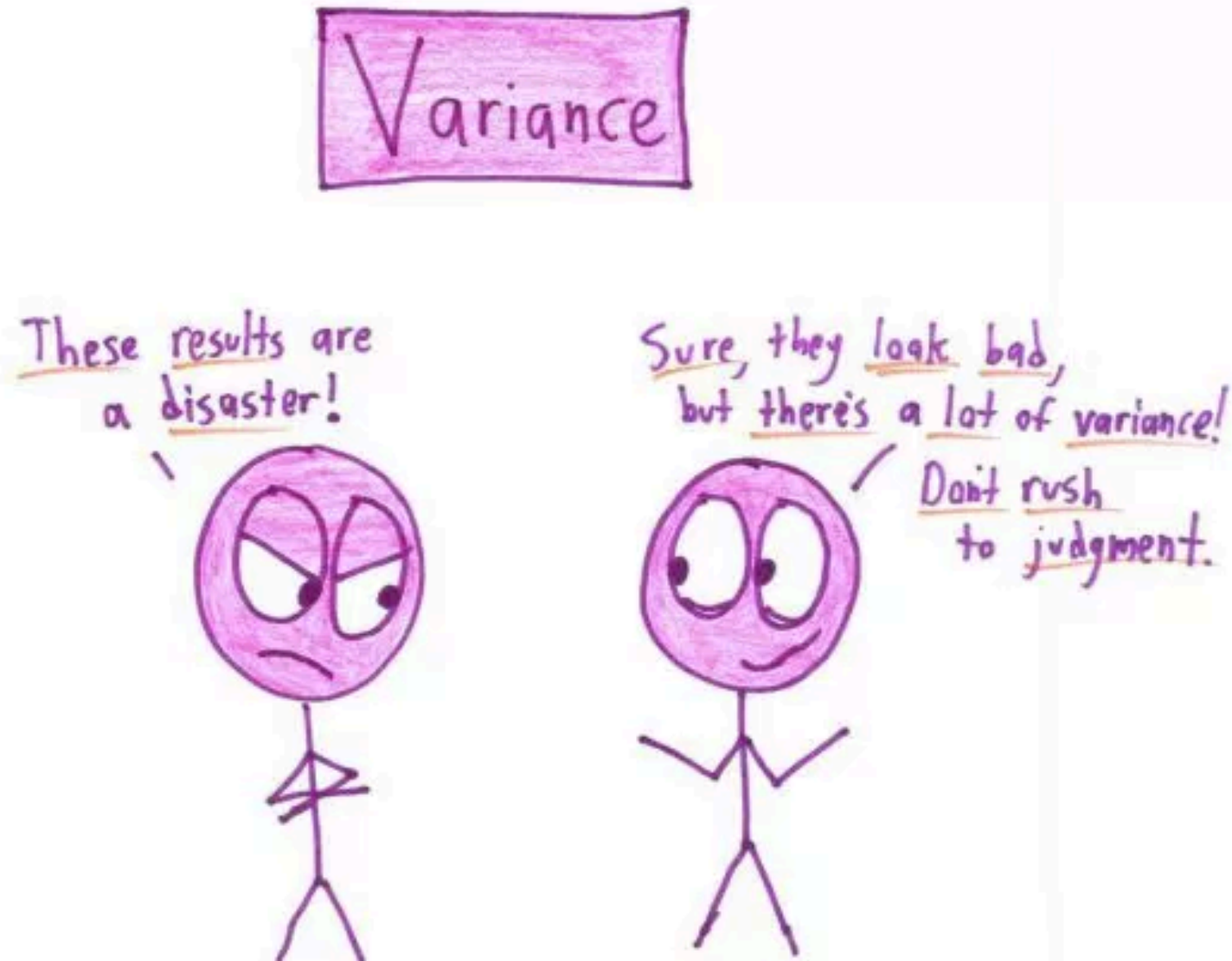
athletic
performance

professional athletes we
paid to guzzle the stuff



amount of
drink consumed

HOW TO LIE WITH STATISTICS

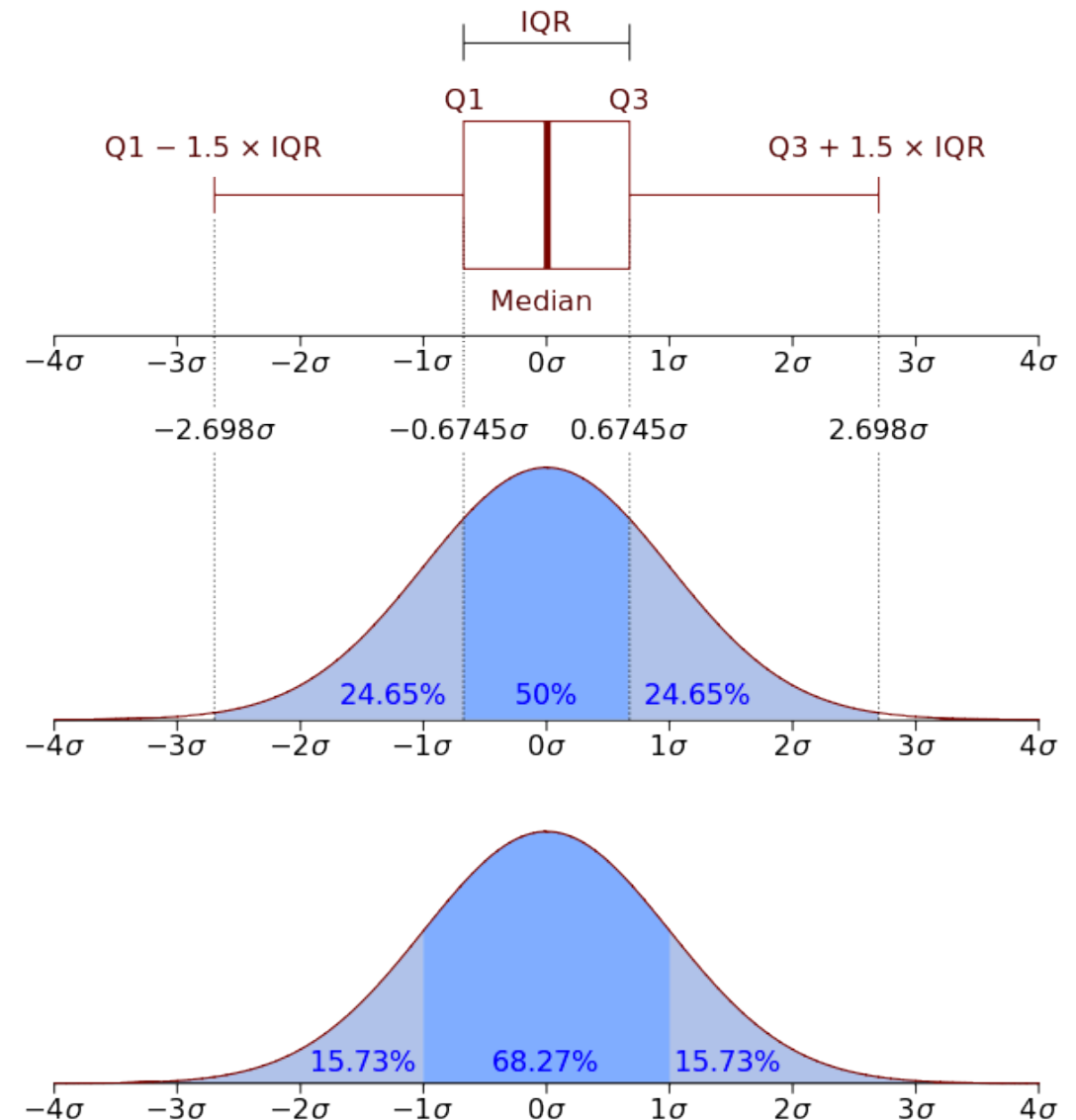


HOW TO LIE WITH STATISTICS



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.
$$\text{IQR} = Q3 - Q1$$
- ▶ Outliers: $Q1 - 1.5(\text{IQR})$, $Q3 + 1.5(\text{IQR})$



CRITERIA FOR GOOD VISUALIZATION

