GENERAL ASSEMBLY

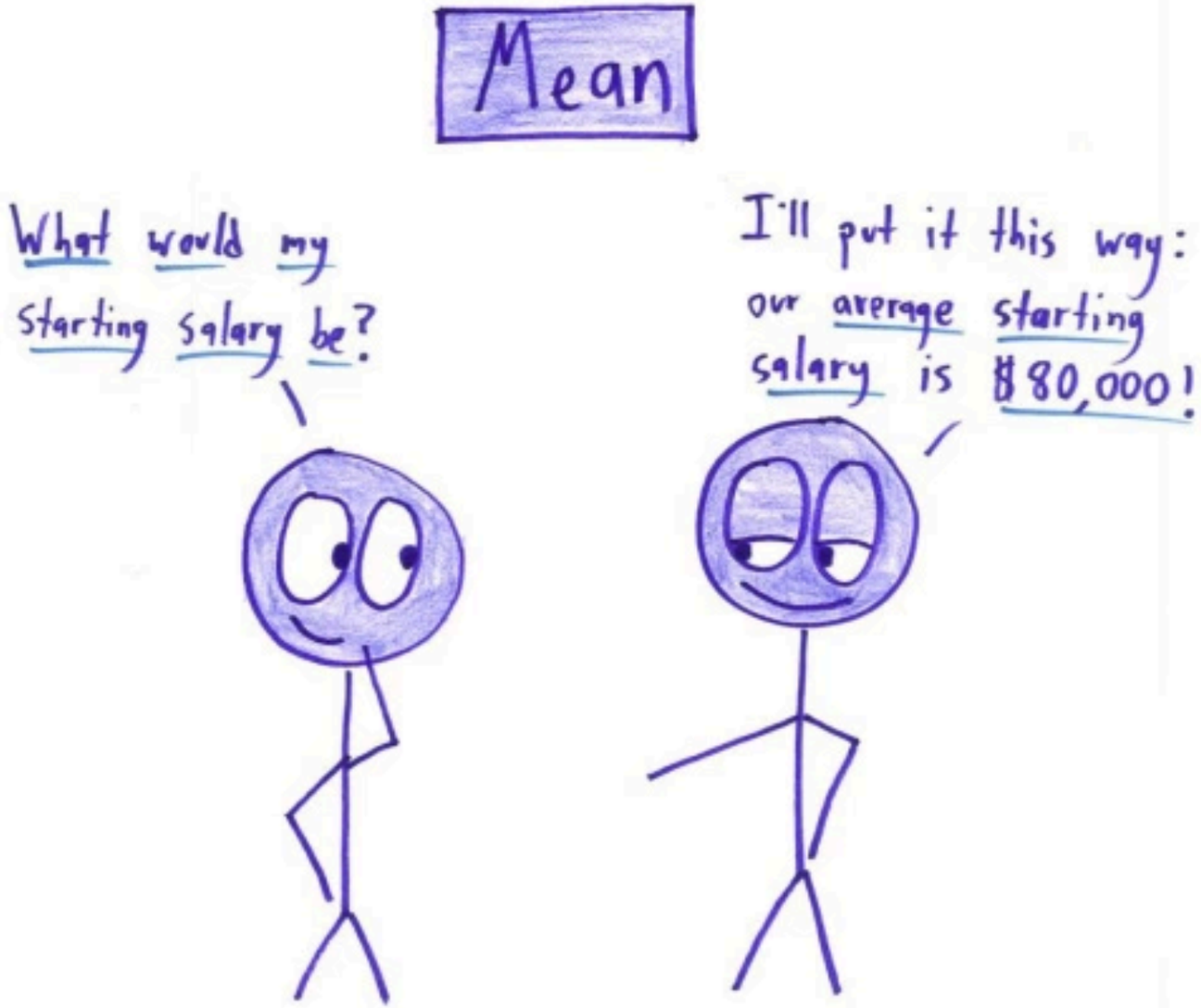
STORYTELLING WITH DATA

Joseph Nelson, Data Science Immersive

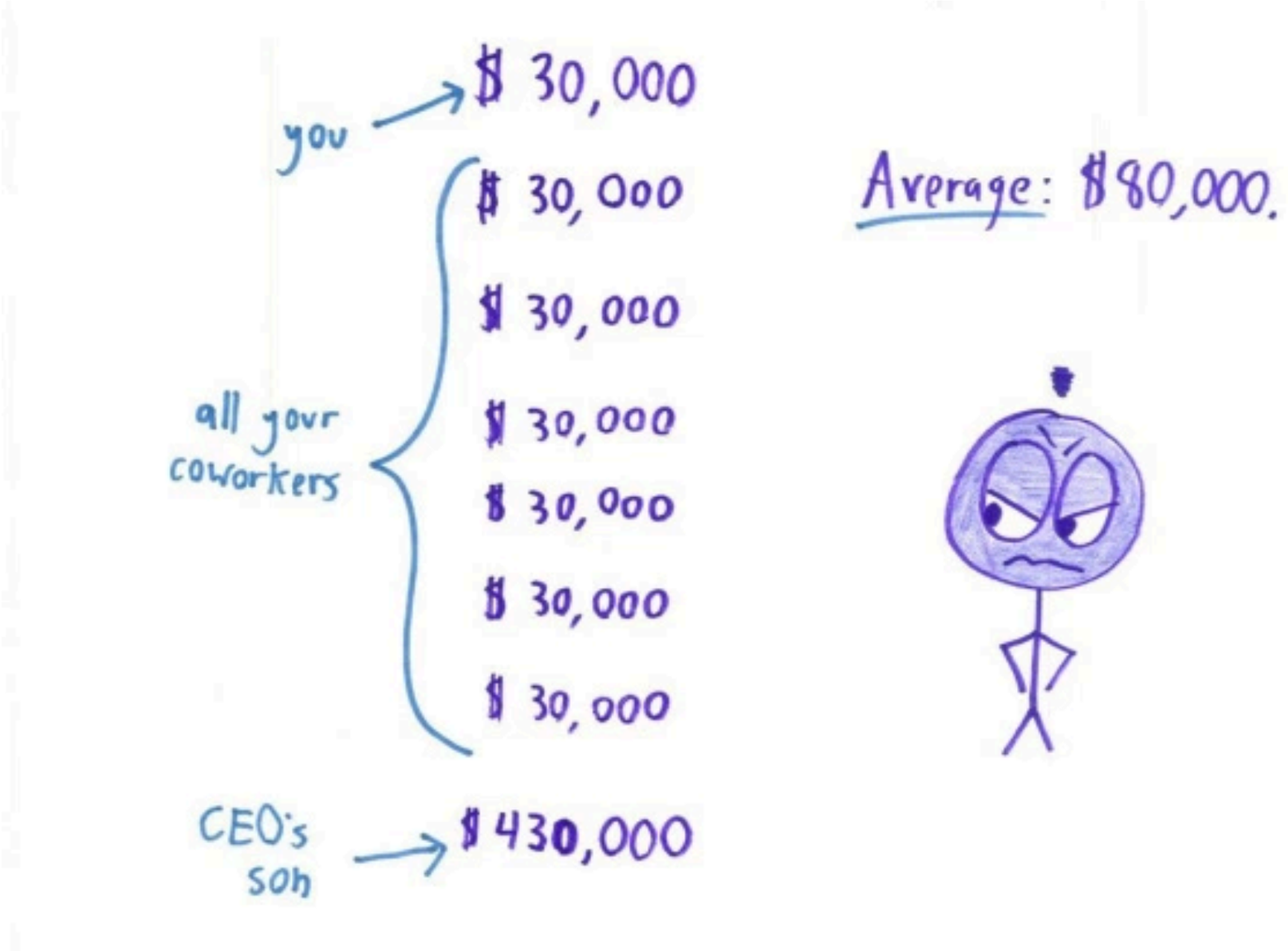
HOW TO LIE WITH STATISTICS

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

HOW TO LIE WITH STATISTICS

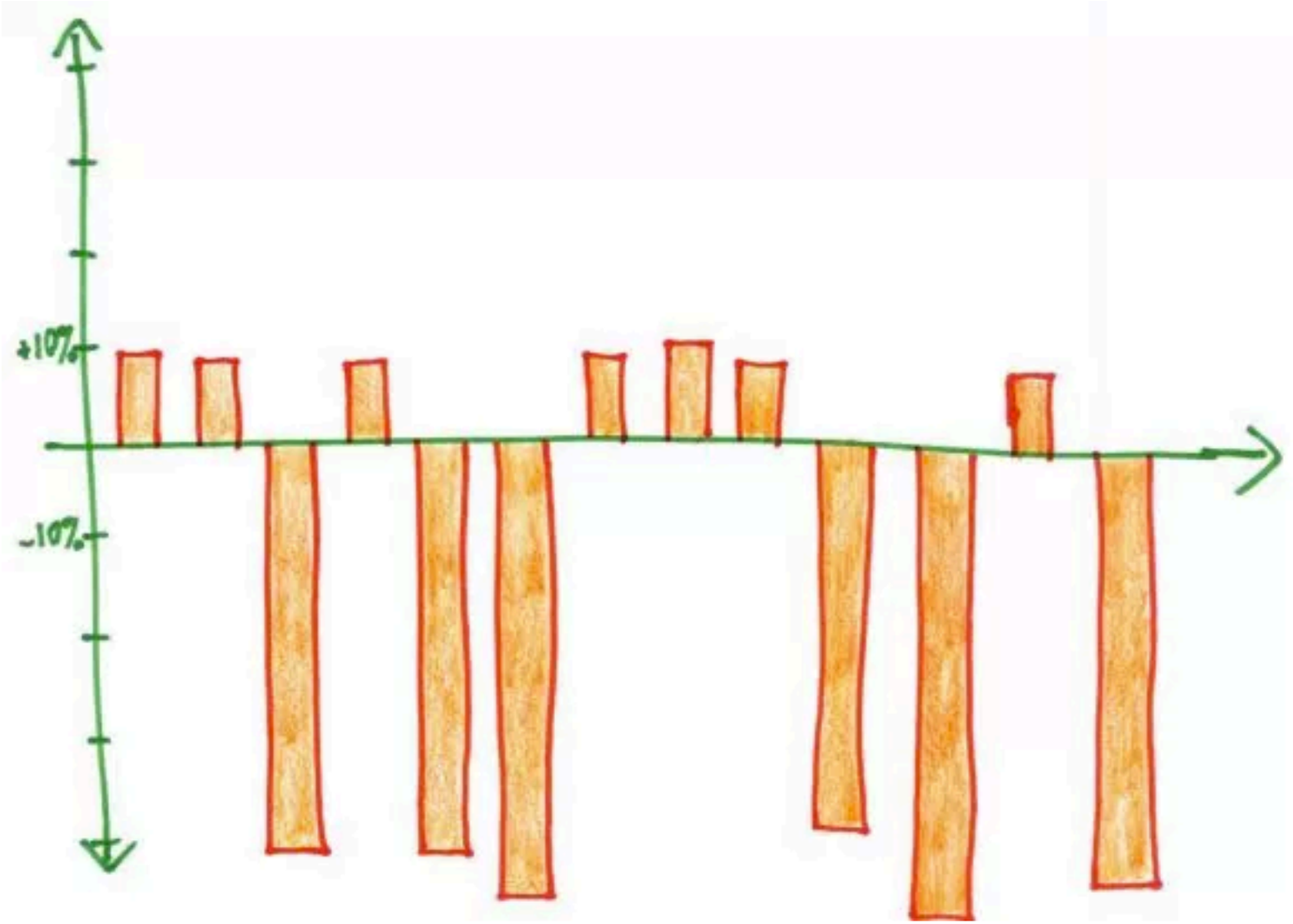


HOW TO LIE WITH STATISTICS

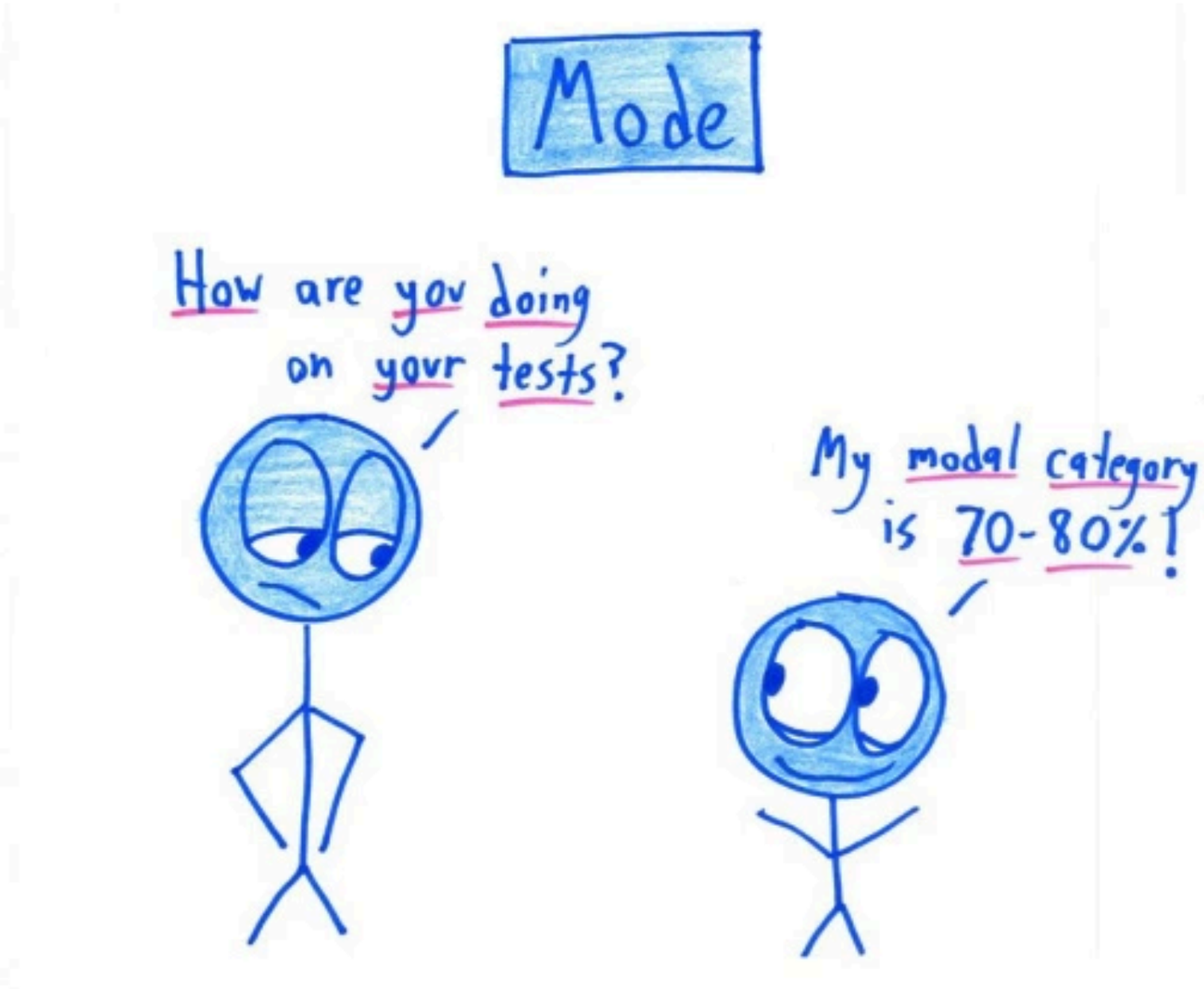


HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS



HOW TO LIE WITH STATISTICS

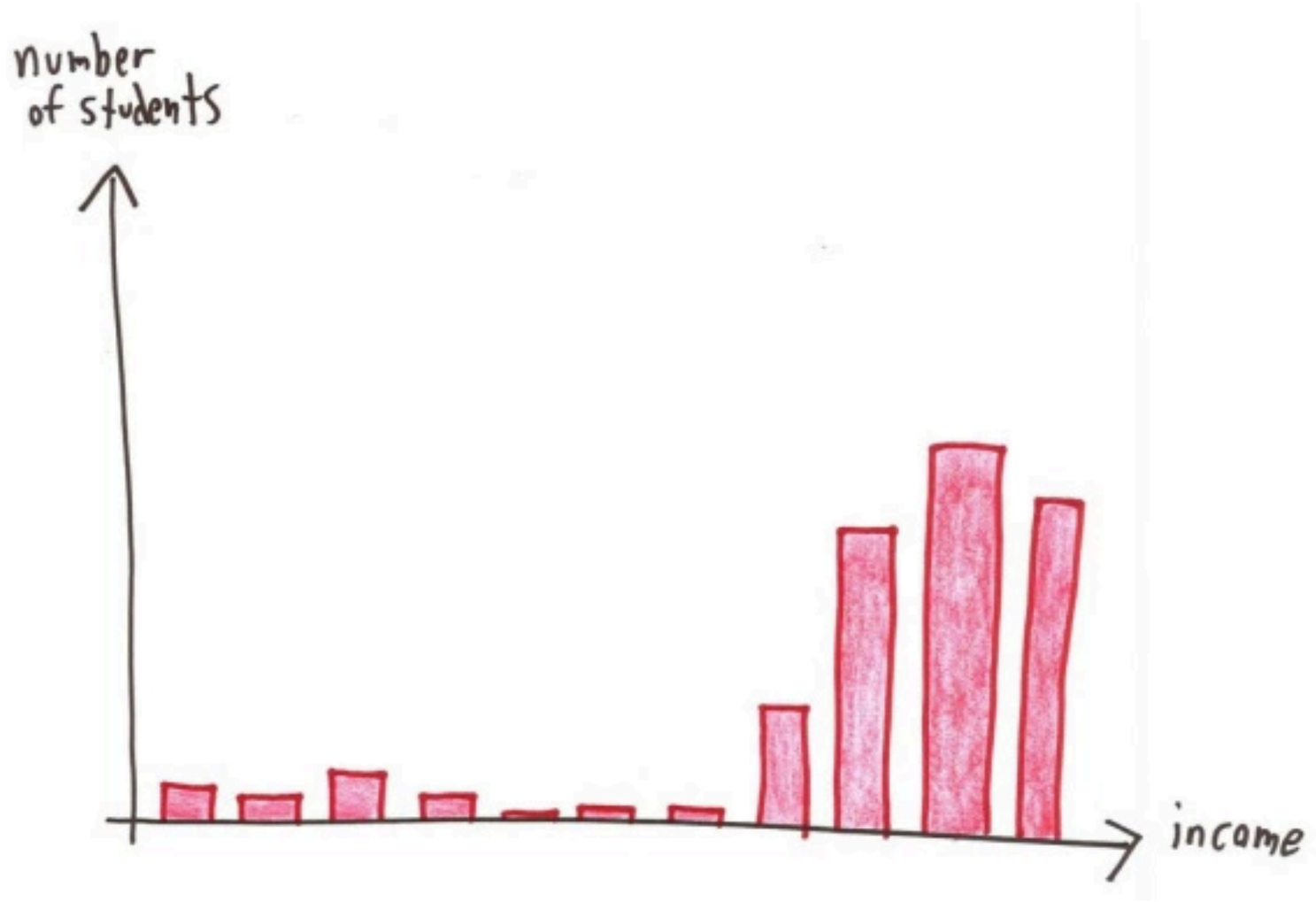


HOW TO LIE WITH STATISTICS

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

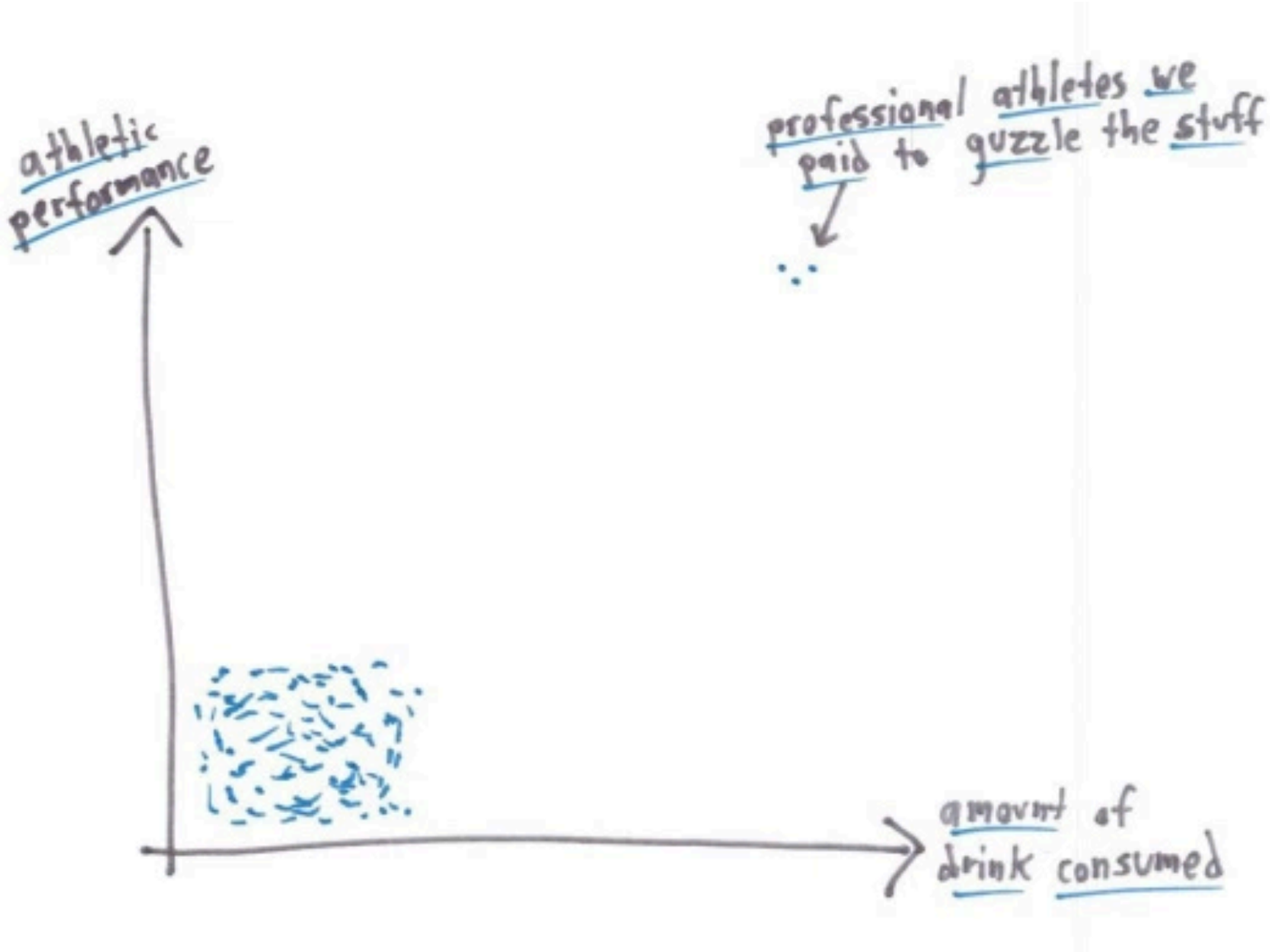
HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

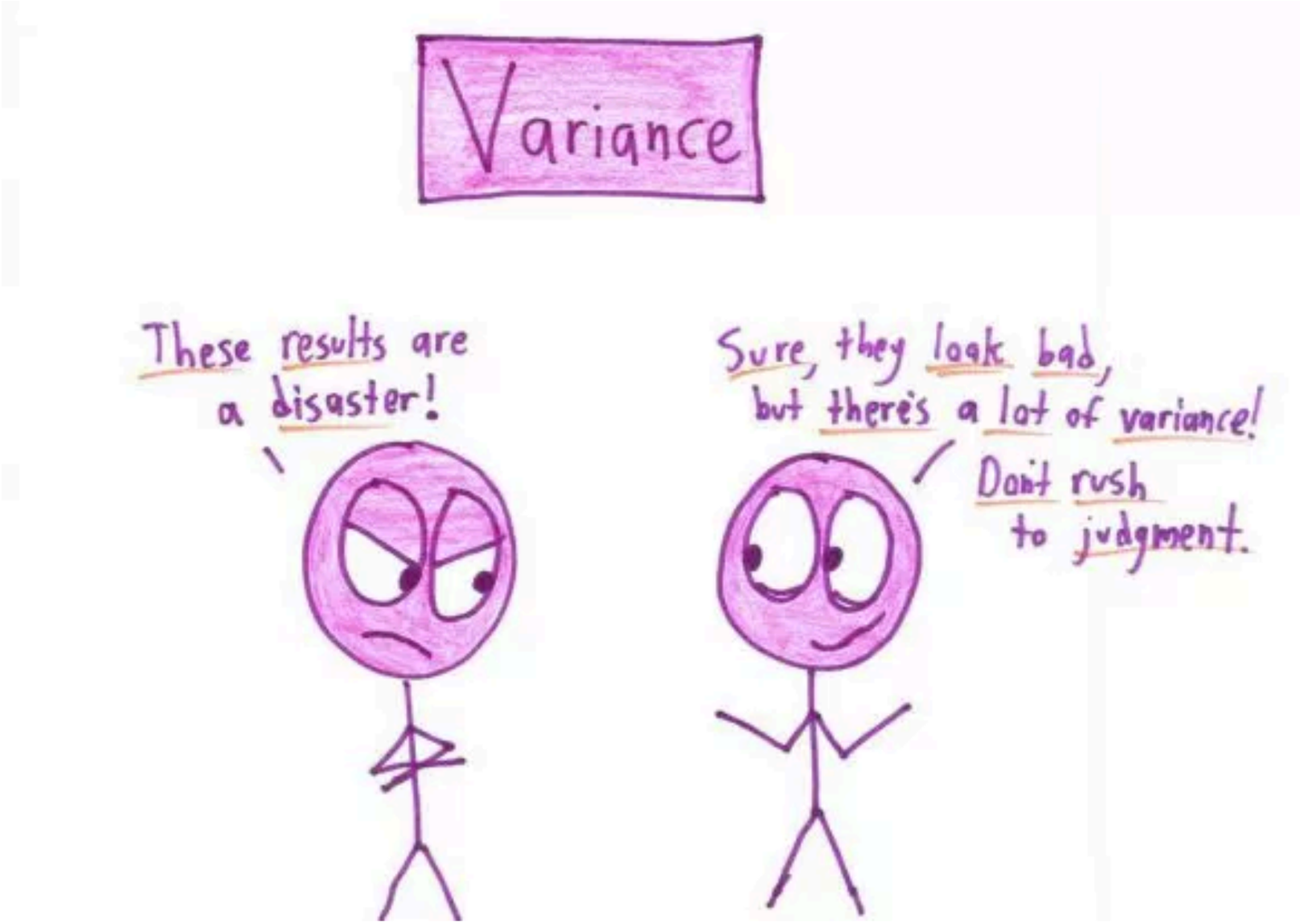


HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS



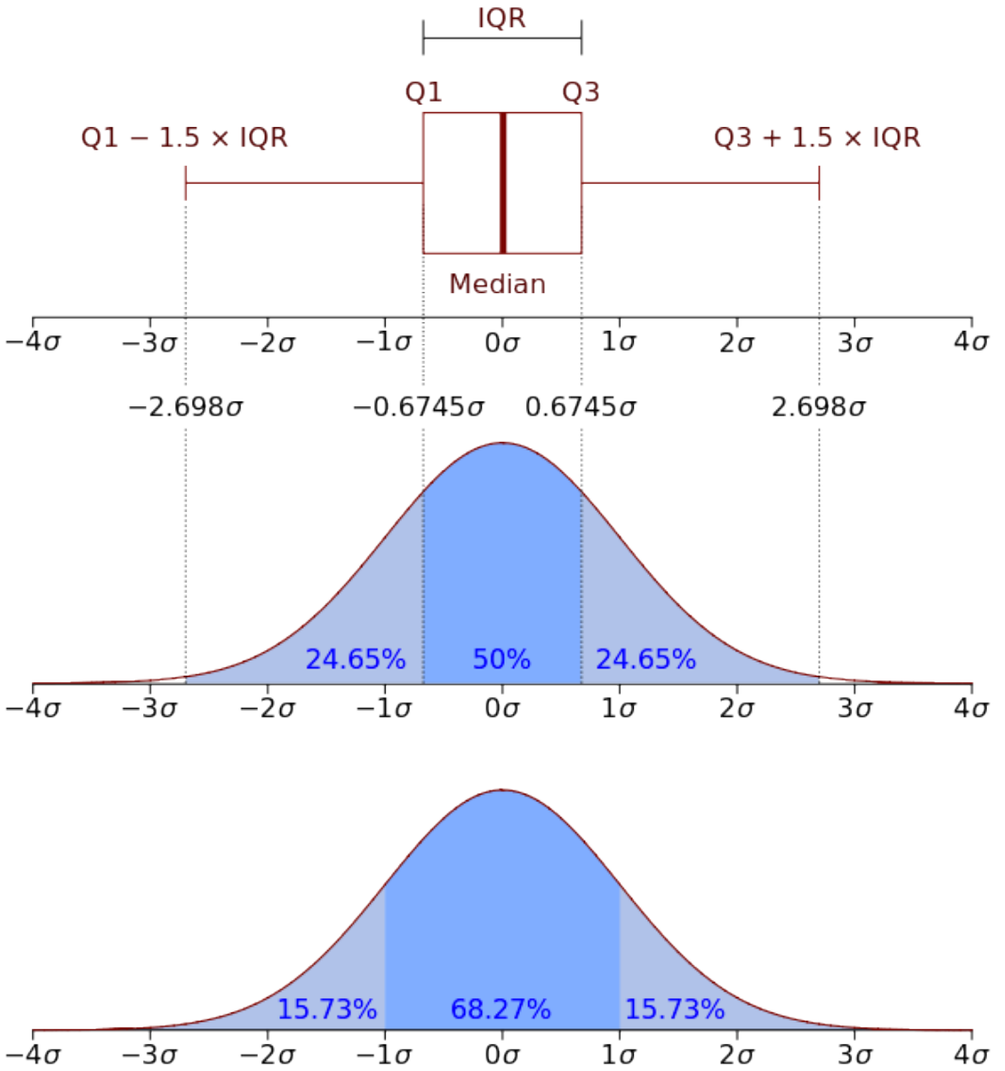
HOW TO LIE WITH STATISTICS



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