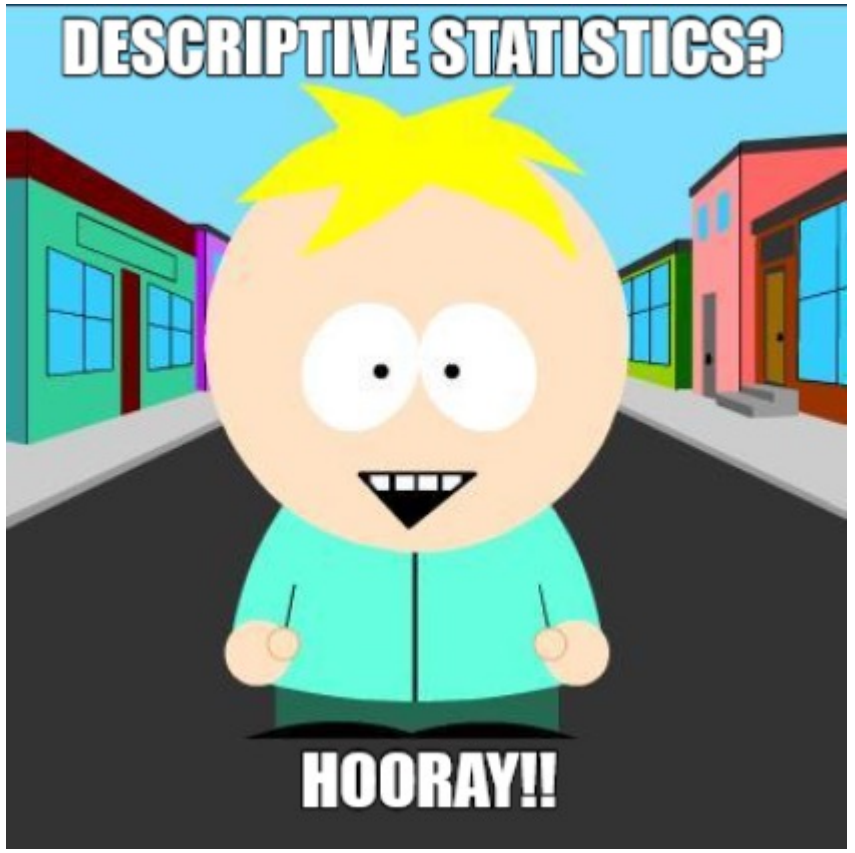
Basics to R

Descriptive Statistics

December 4, 2020

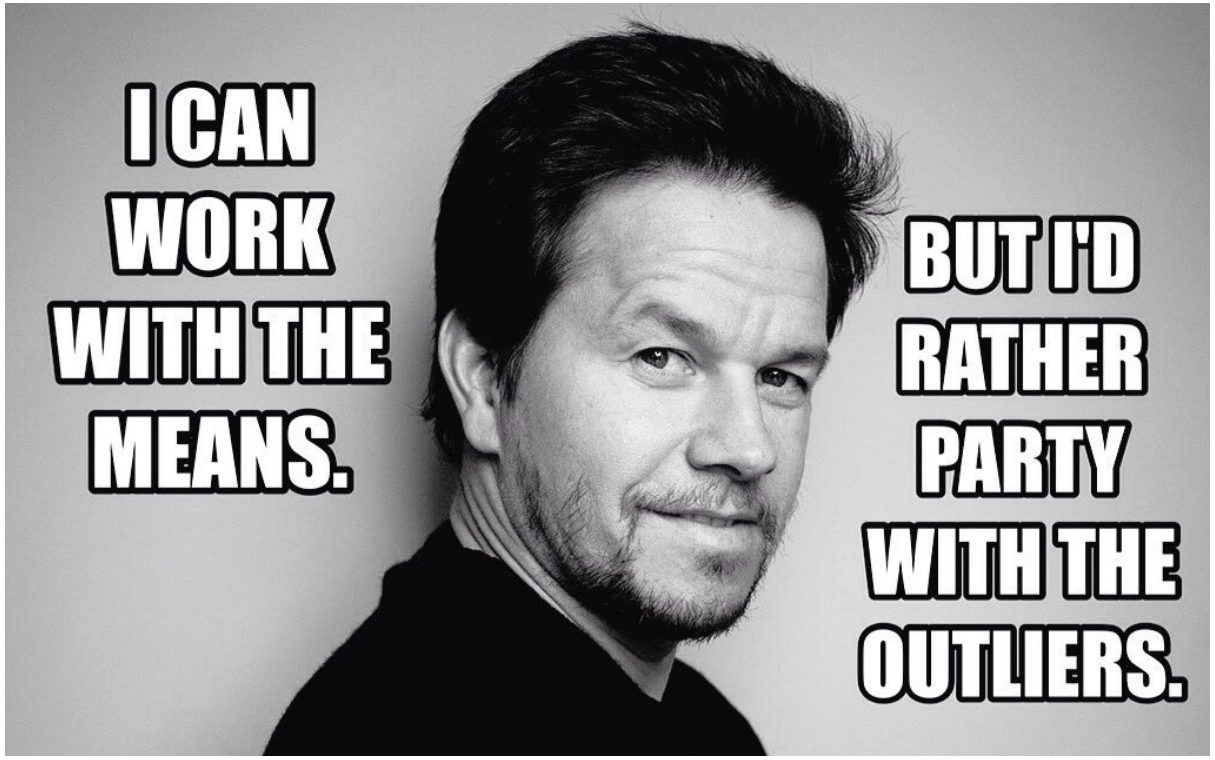
Outline for Today

- What are descriptive statistics?
- Why are they important?
- What do I need to look at?
- Using R for descriptive statistics and basic data visualization

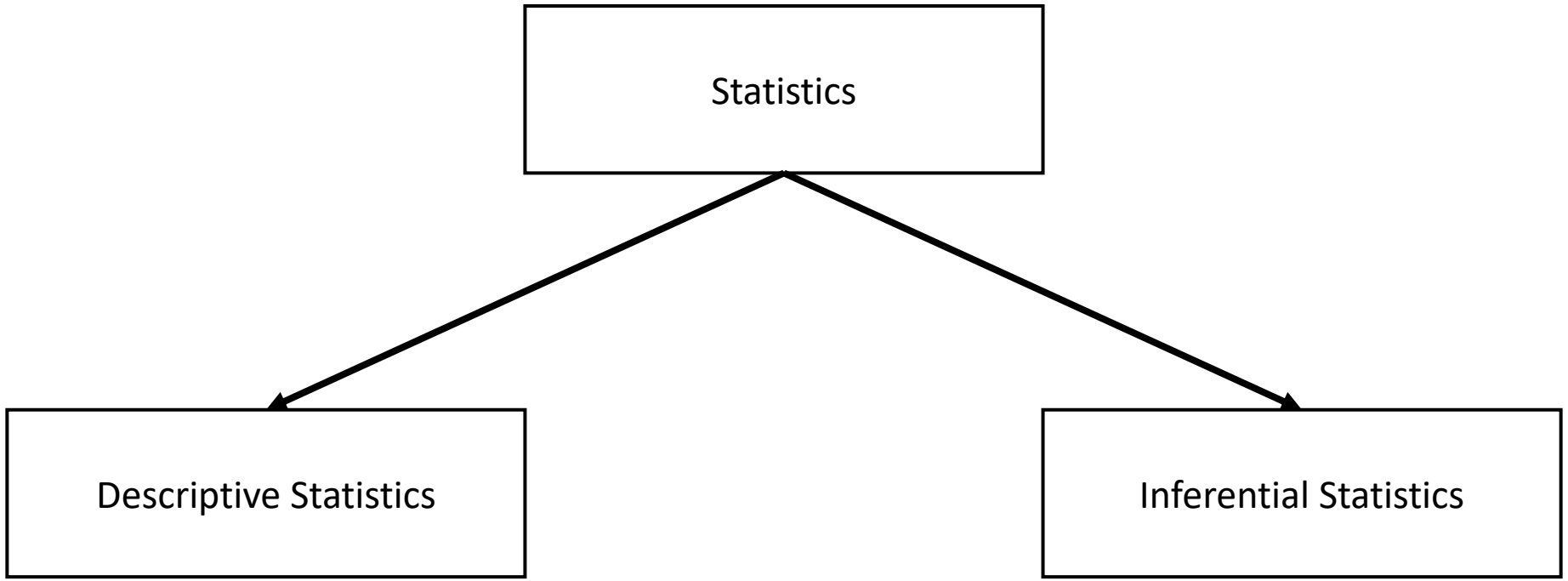


What are descriptive statistics?

- Simple summaries about the sample
- Can be quantitative (e.g., the mean) or visual (e.g., histograms)
- Common descriptive statistics:
  - **Measures of central tendency:** mean, mode, median
  - **Measures of variability:** standard deviation, variance, range, IQR
  - **Modality**
  - **Skew**
  - **Kurtosis**



What are descriptive statistics?



But why should I care?

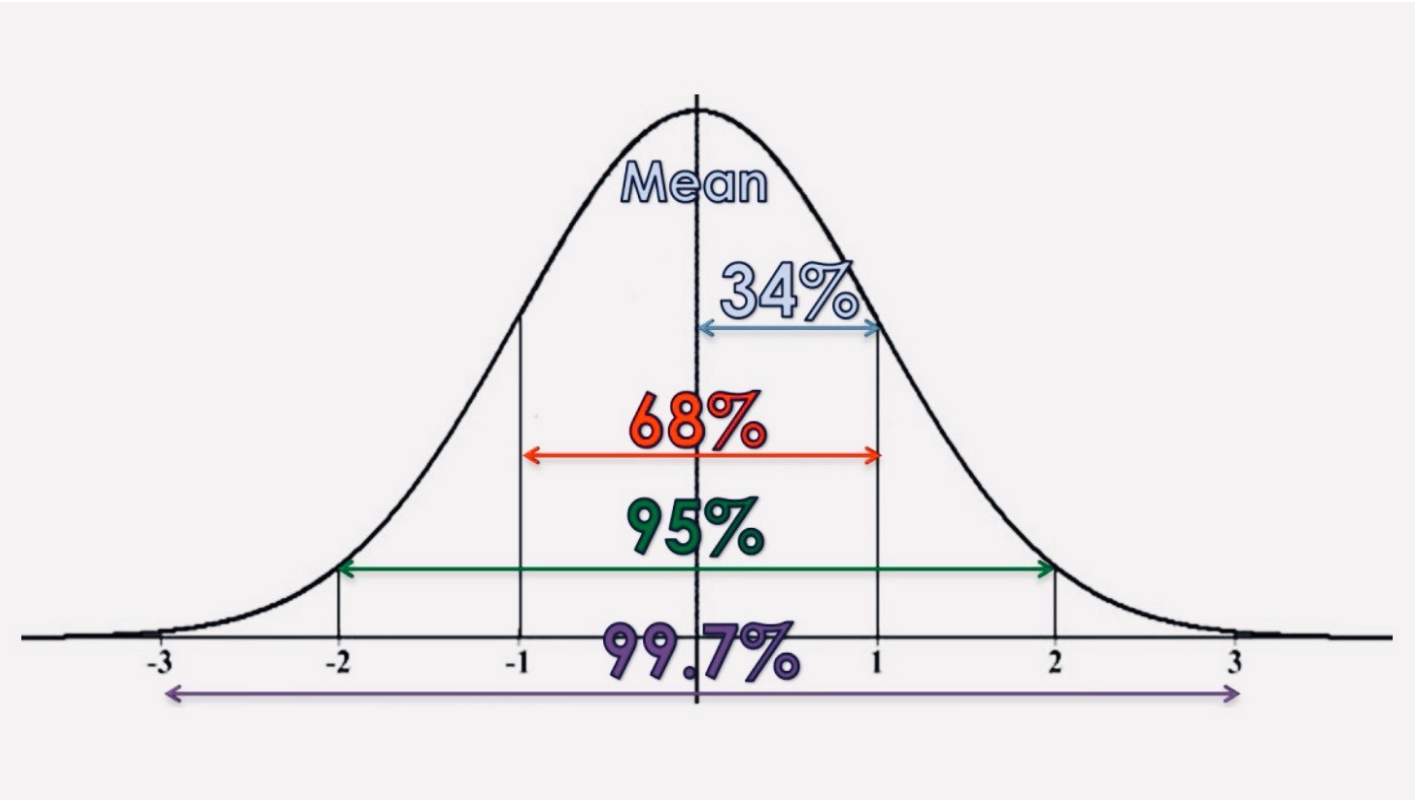
- Know thy data, know thyself
- All inferential statistics are based on assumptions
  - If you're data don't meet the assumptions, then you're conclusions may be false
- Skew, kurtosis, variance, and distribution

Variance

$$s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

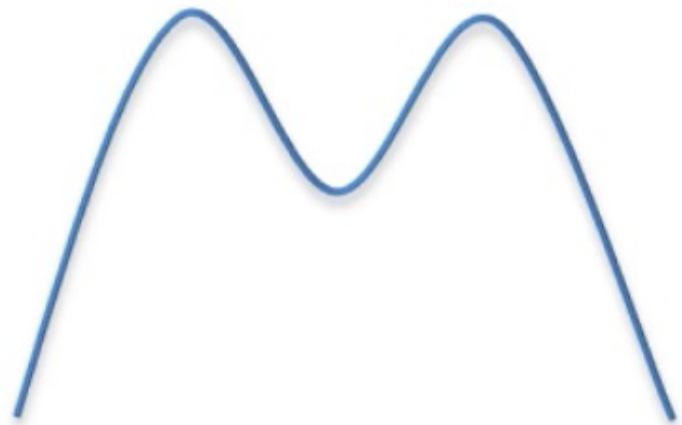
Standard Deviation

$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

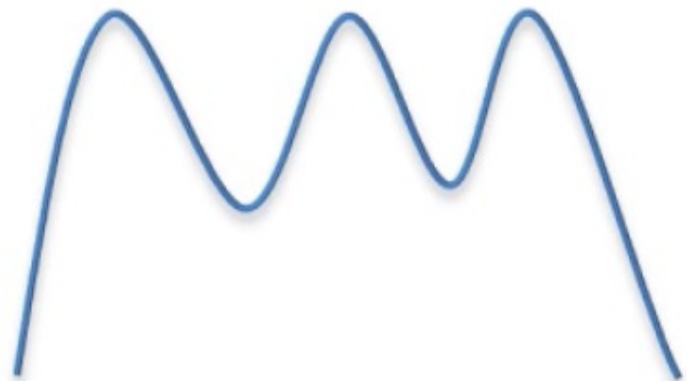


Unimodal

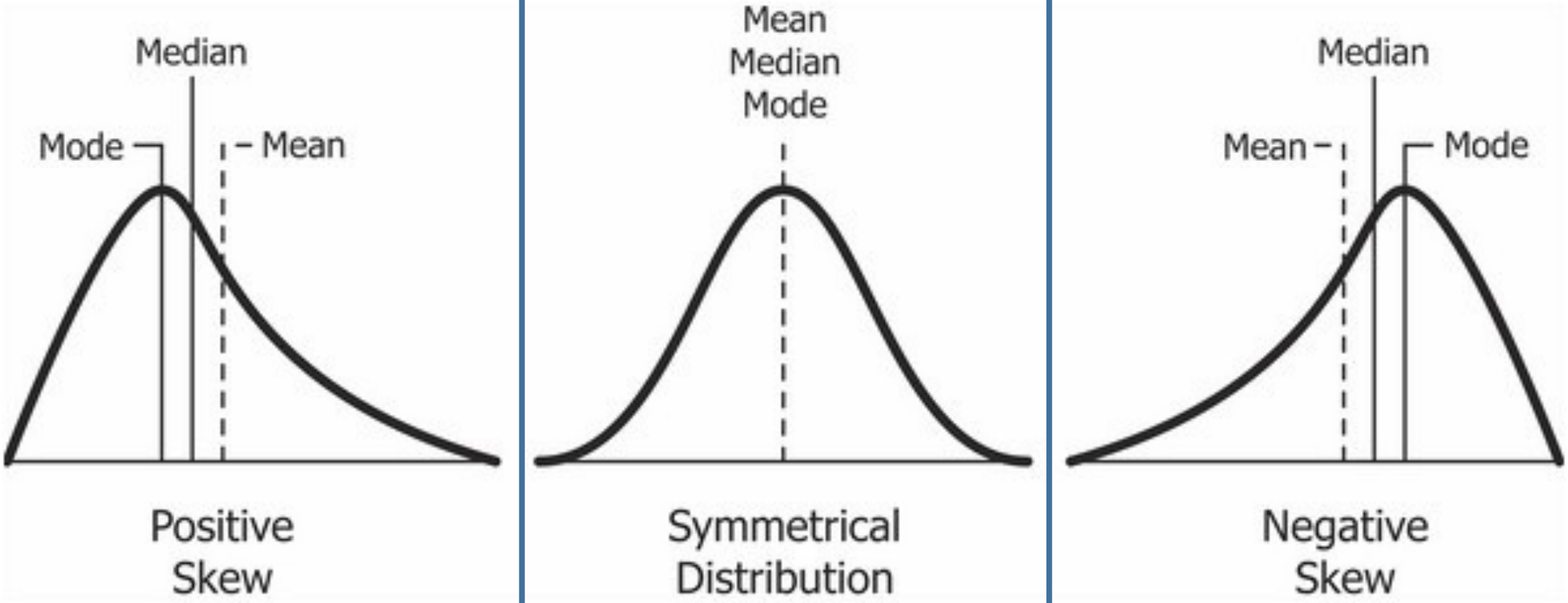
Bimodal

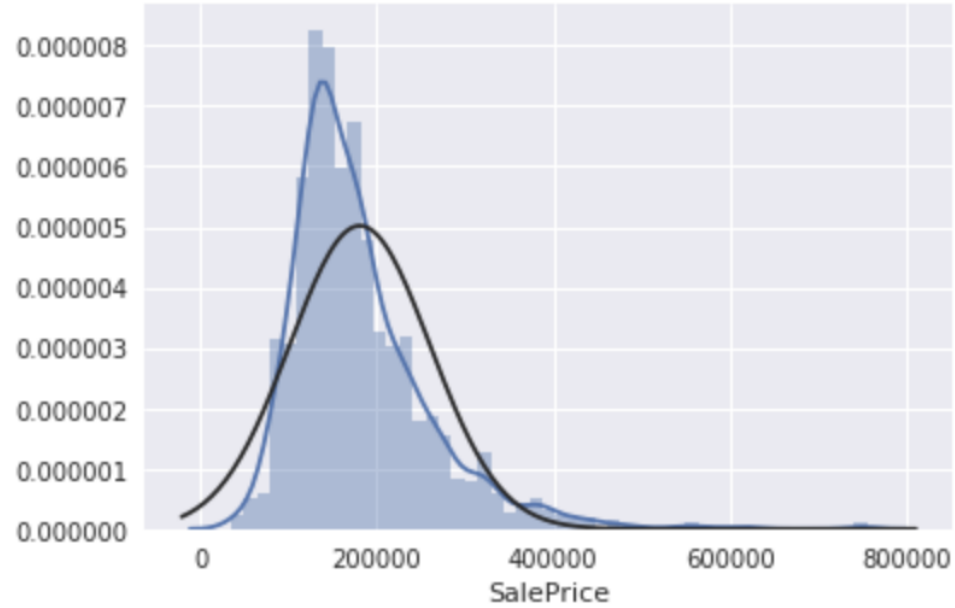


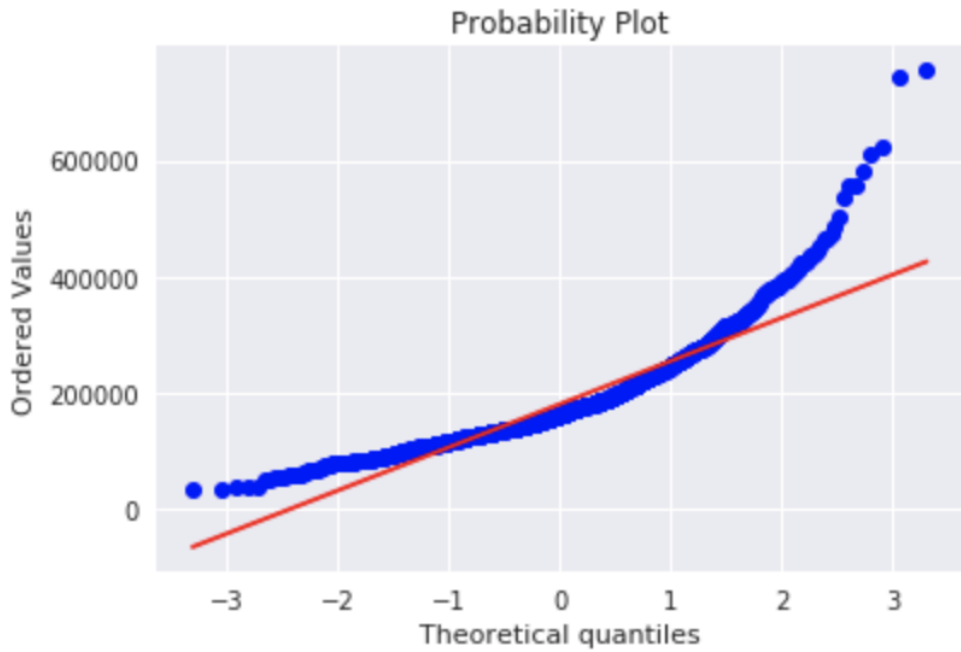
Multimodal



AKA *The Dream Distribution*




0.000008
0.000007
0.000006
0.000005
0.000004
0.000003
0.000002
0.000001
0.000000
0
200000
400000
600000
800000
SalePrice


Probability Plot
600000
400000
200000
0
Ordered Values
−3
−2
−1
0
1
2
3
Theoretical quantiles

Summary

- Descriptive statistics are not very interesting
- Not looking at descriptive statistics is bad
- Ryan Gosling stats memes are great ways to end a PowerPoint