

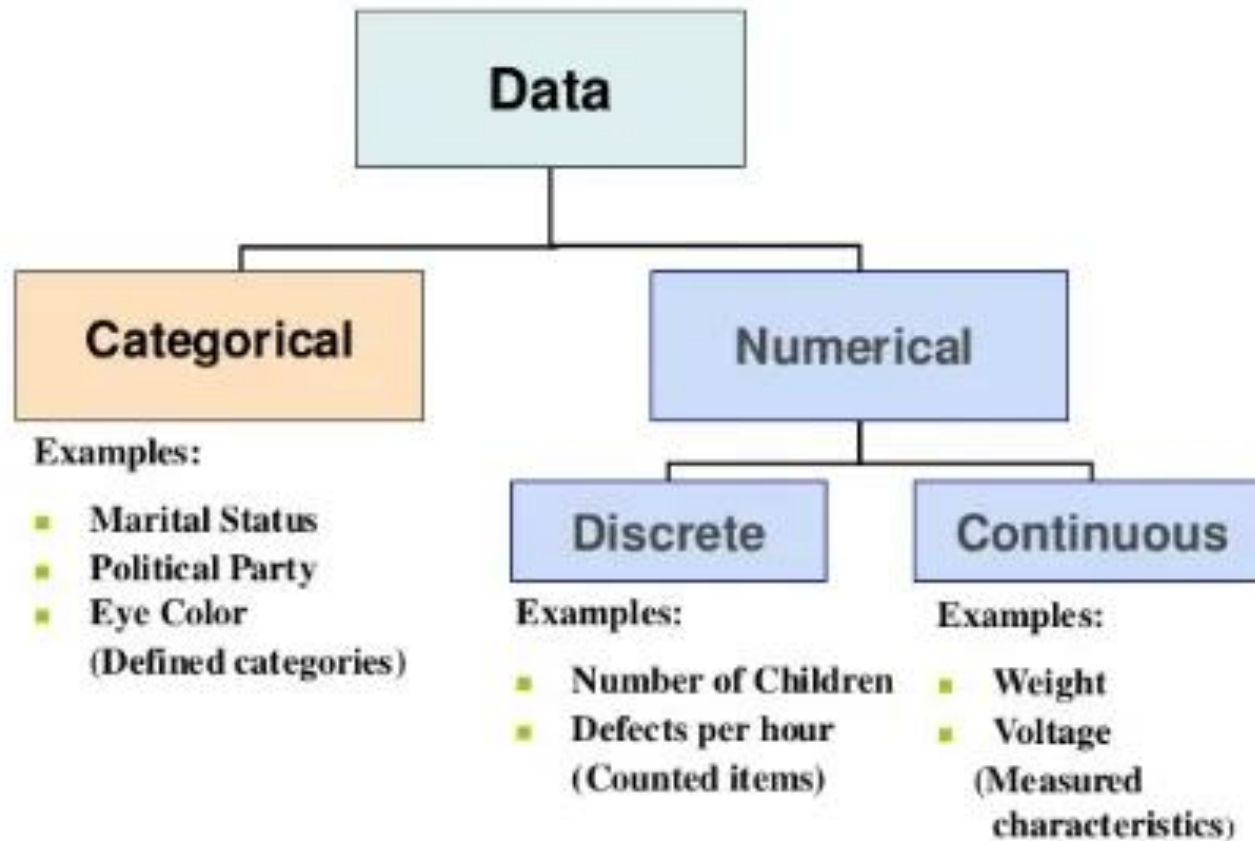
Exploratory Data Analysis

CS 459 Business Intelligence

Univariate Analysis



Data Types

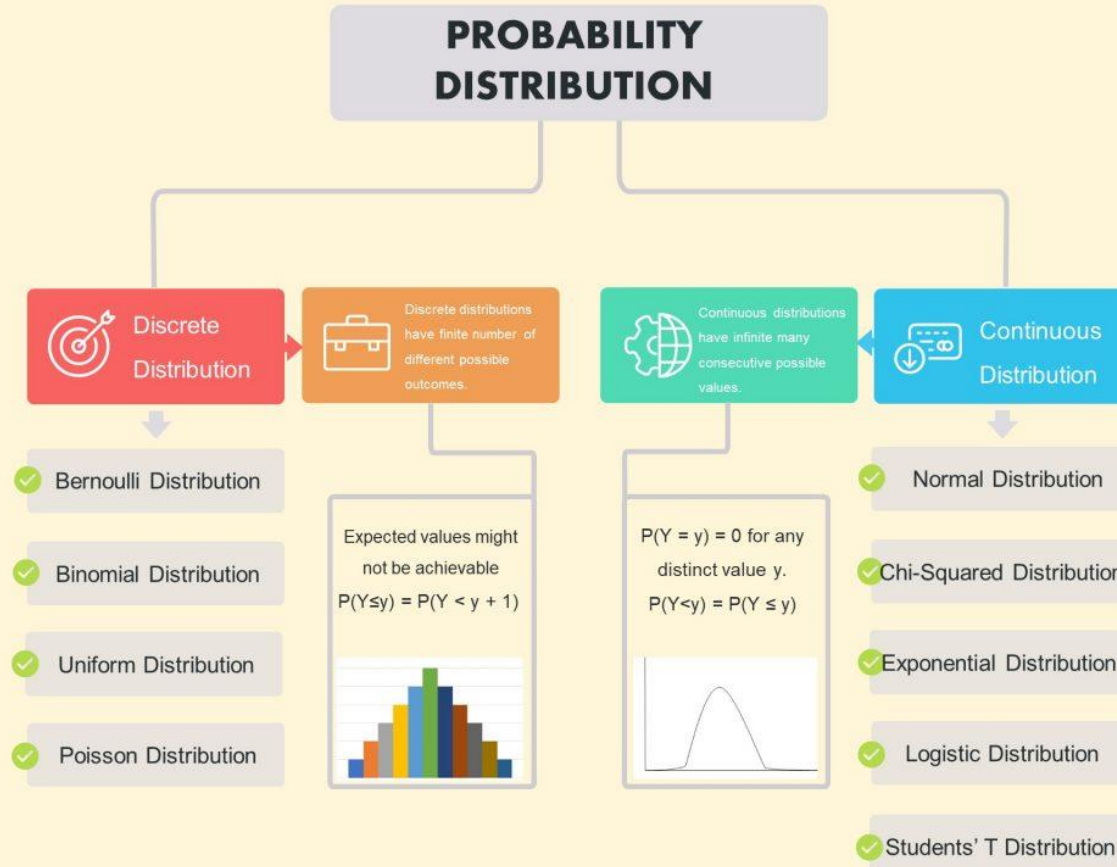


- Note: Categorical Data can be visualized by Bar Plot, Pie Chart, Pareto Chart.
- Numerical Data can be visualized by Histogram, Line Plot, Scatter Plot

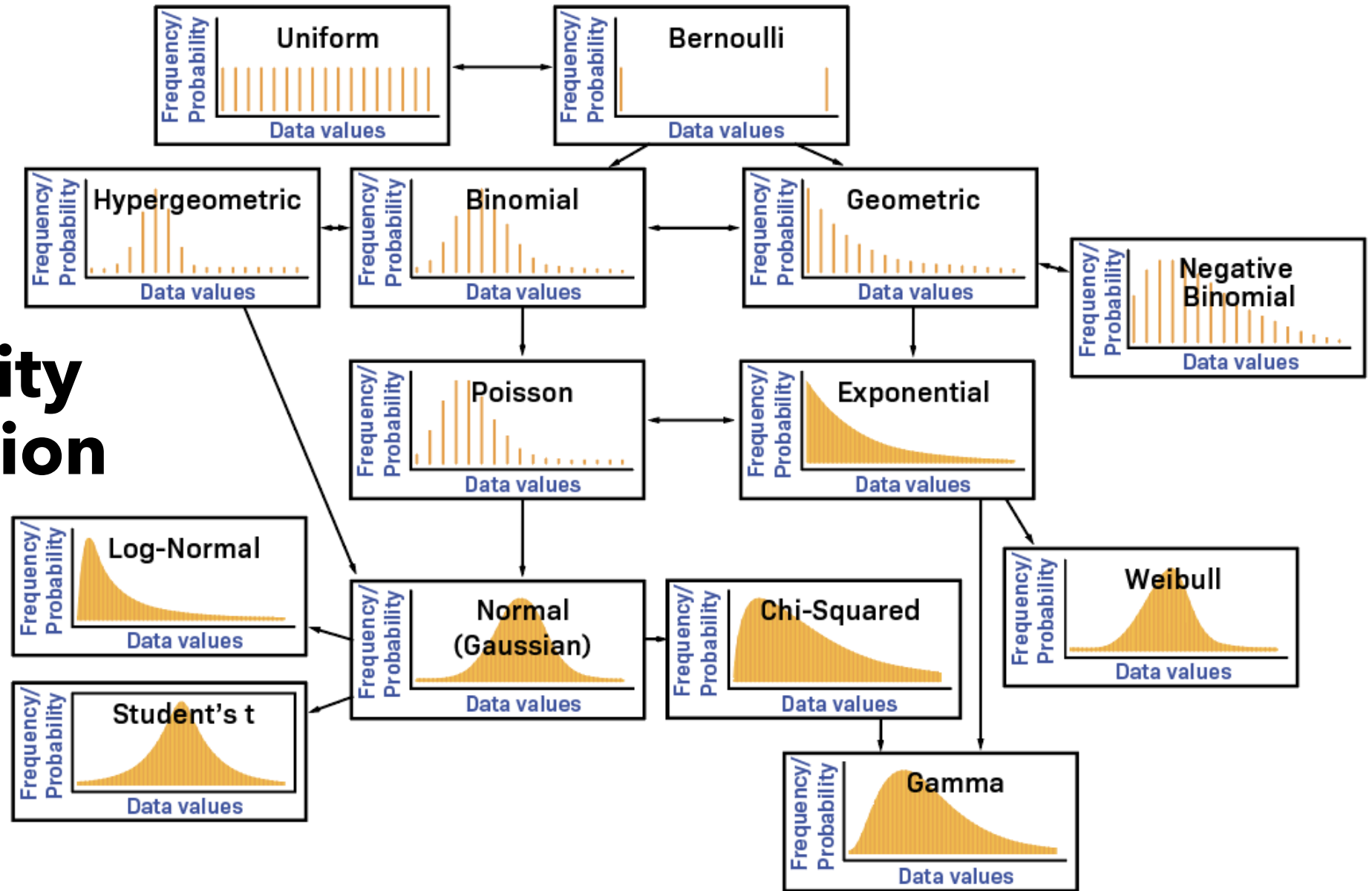
Probability Distributions

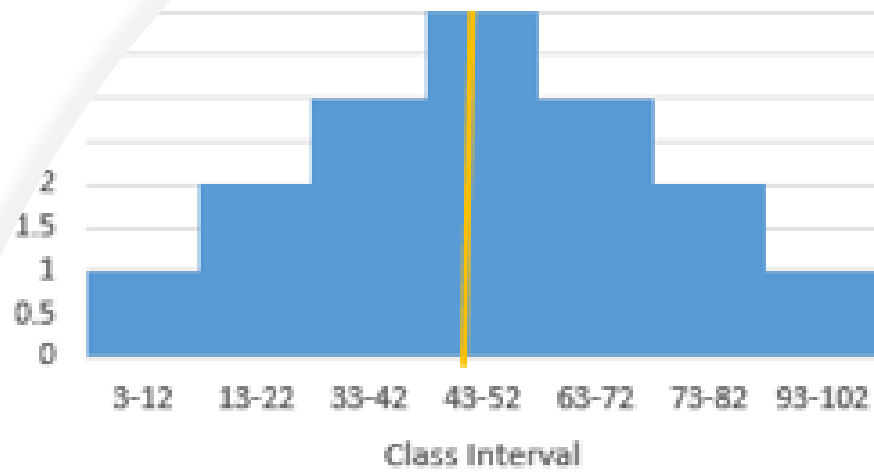
- A probability distribution is a mathematical function that describes the probability of different possible values of a variable.

Types of Probability Distribution Characteristics, Examples, & Graph

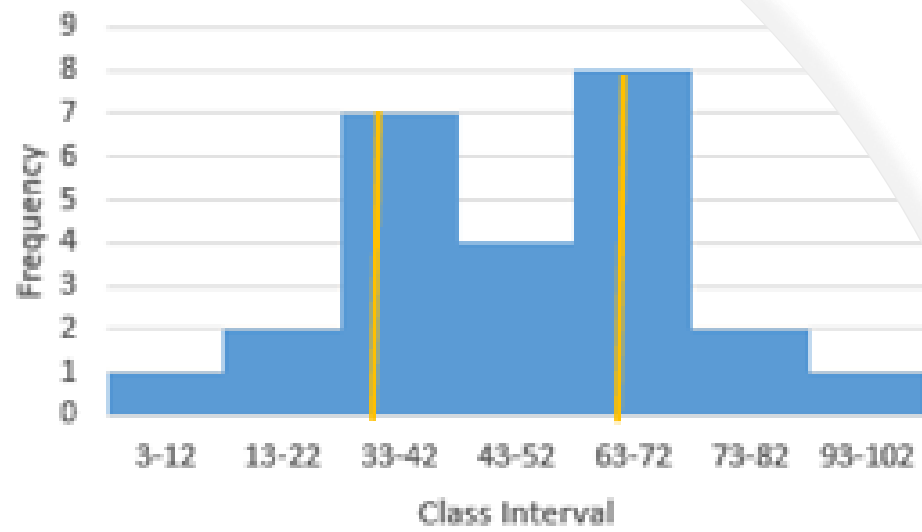


Probability distribution



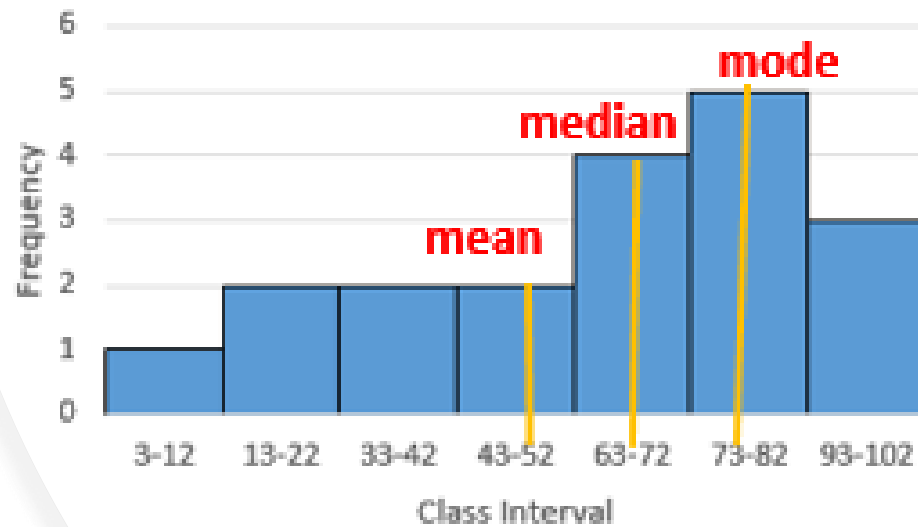


Mean=median=mode



2 Modes

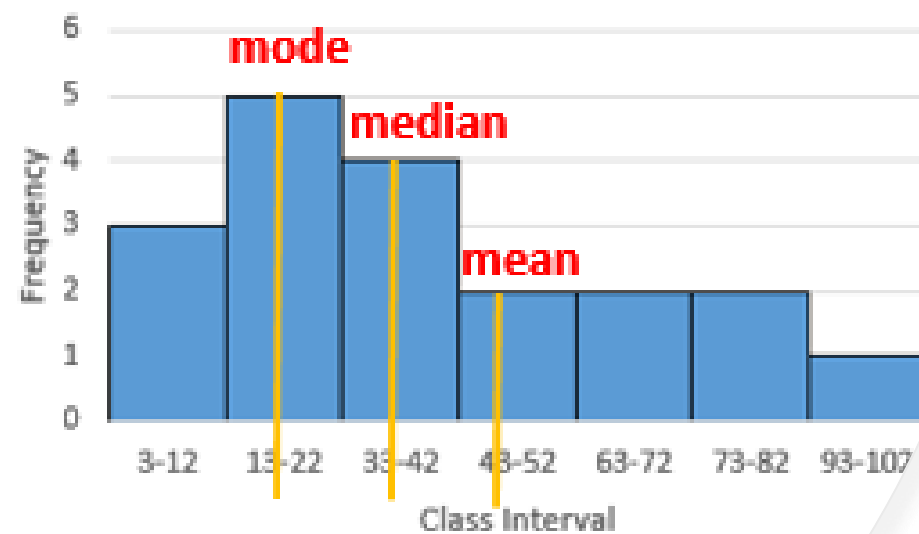
NEGATIVELY SKEWED



Mean<median<mode

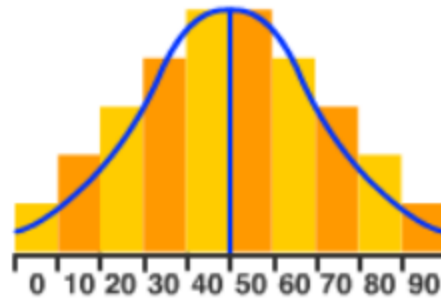


POSITIVELY SKEWED

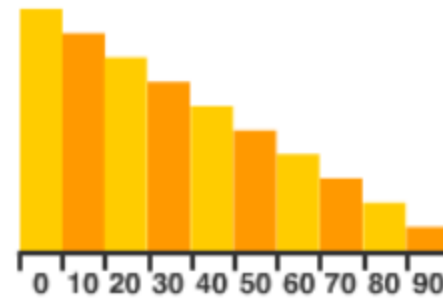


Mean>median>mode

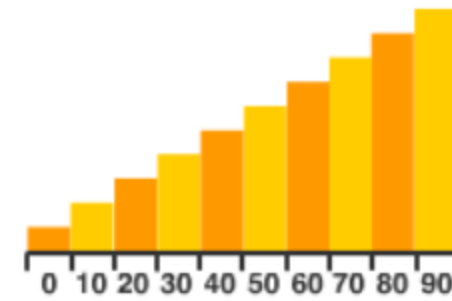




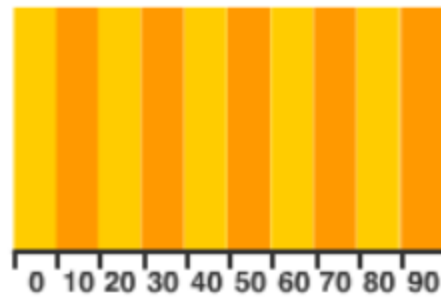
normal distribution
unimodal, symmetric,
aka 'bell curve'



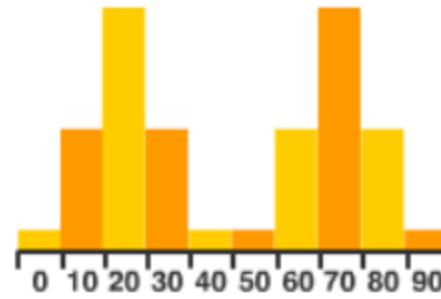
skewed distribution
positively skewed,
skewed right



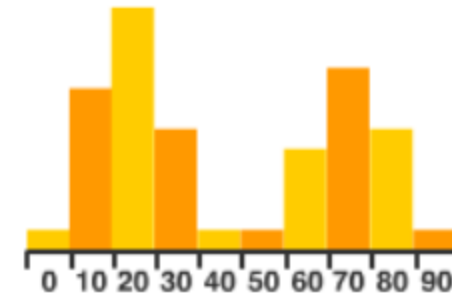
skewed distribution
negatively skewed,
skewed left



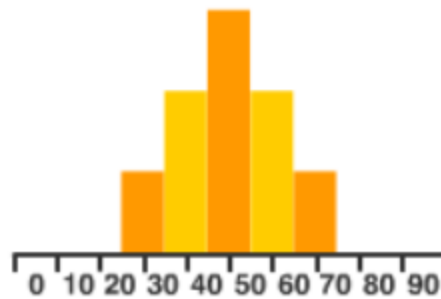
uniform distribution
equally spread,
no peaks



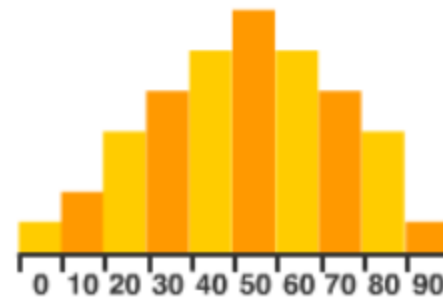
bimodal distribution
two modes,
symmetric



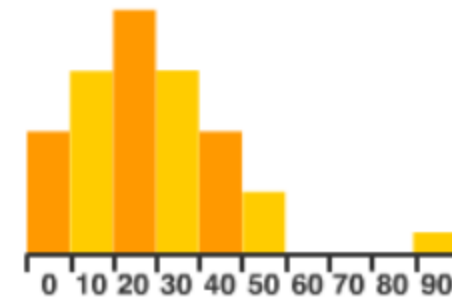
bimodal distribution
two modes,
non-symmetric



spread
narrow range

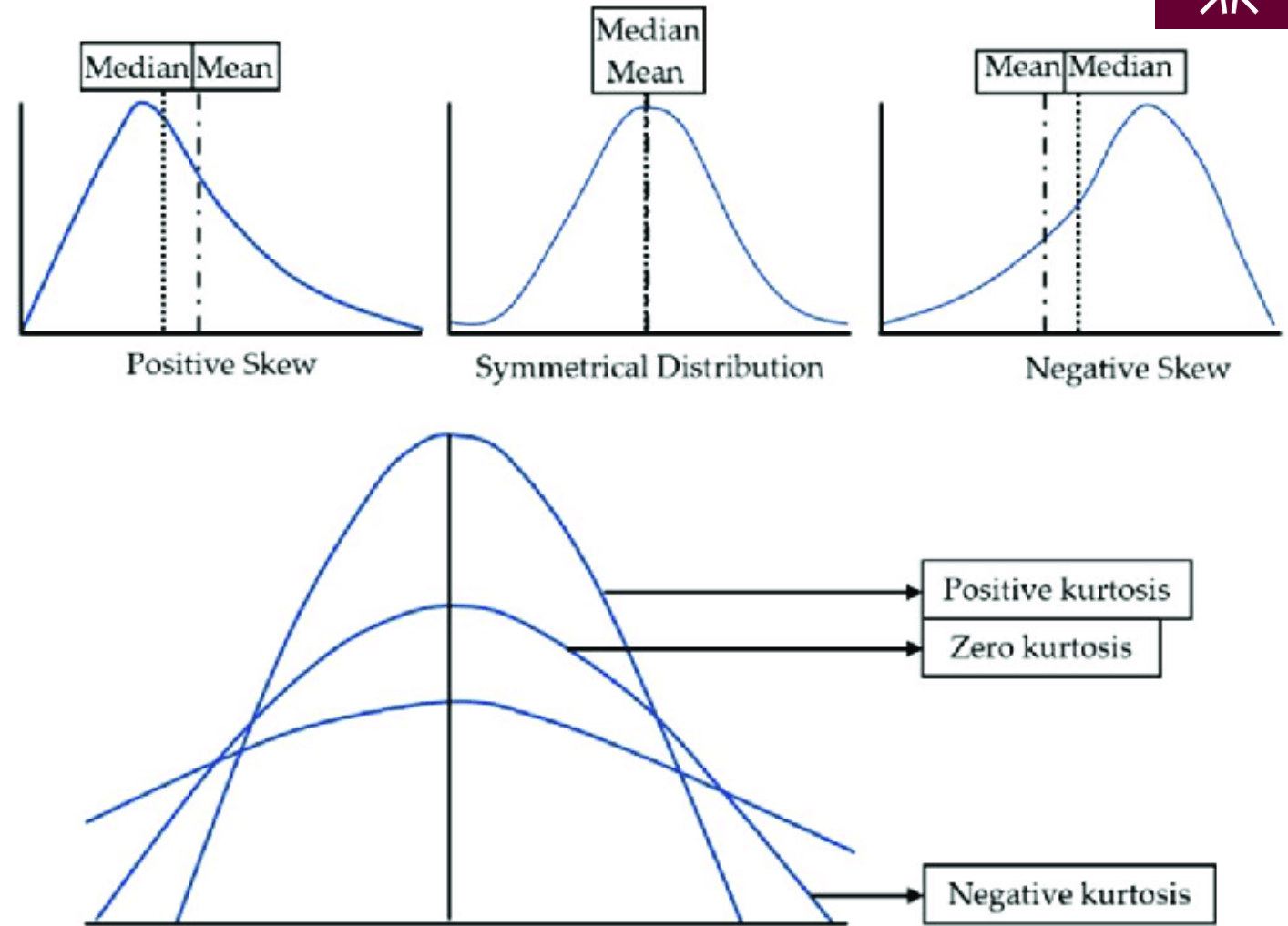


spread
wide range

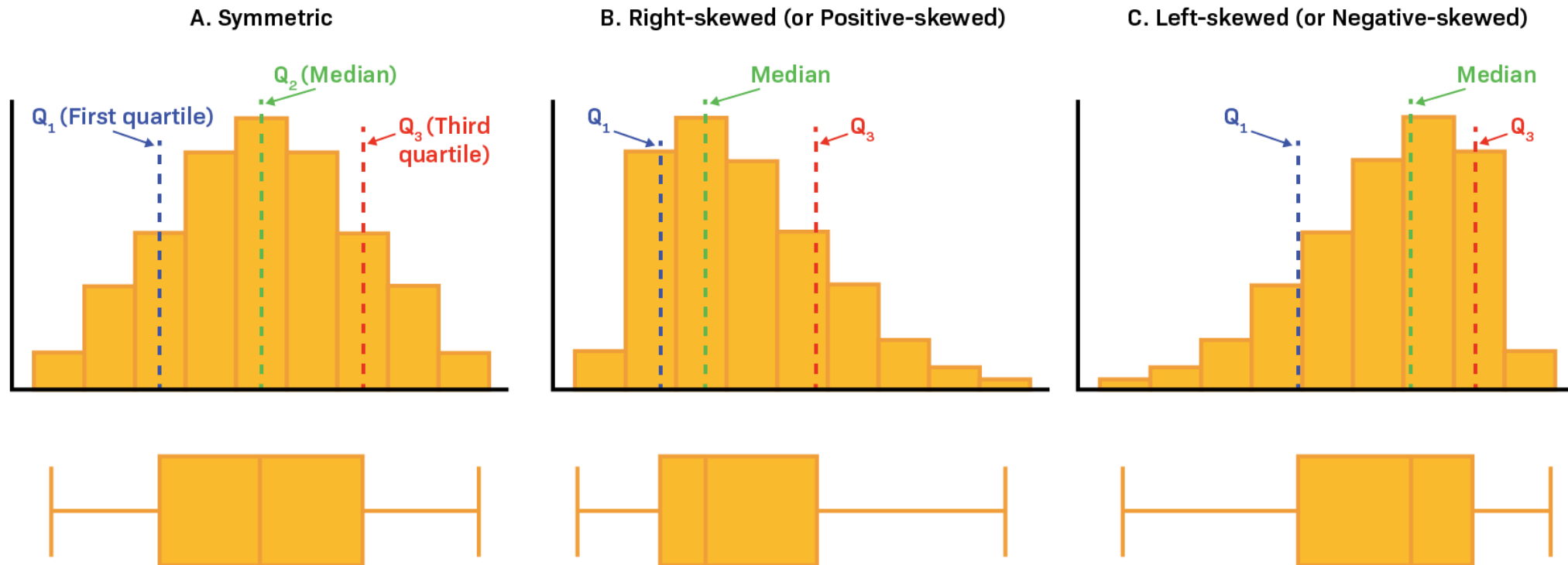


spread
outlier

Histograms: Shape, Skew and Kurtosis



Histograms and Box Plots



Python Notebooks:

Descriptive Statistics

Outlier Analysis