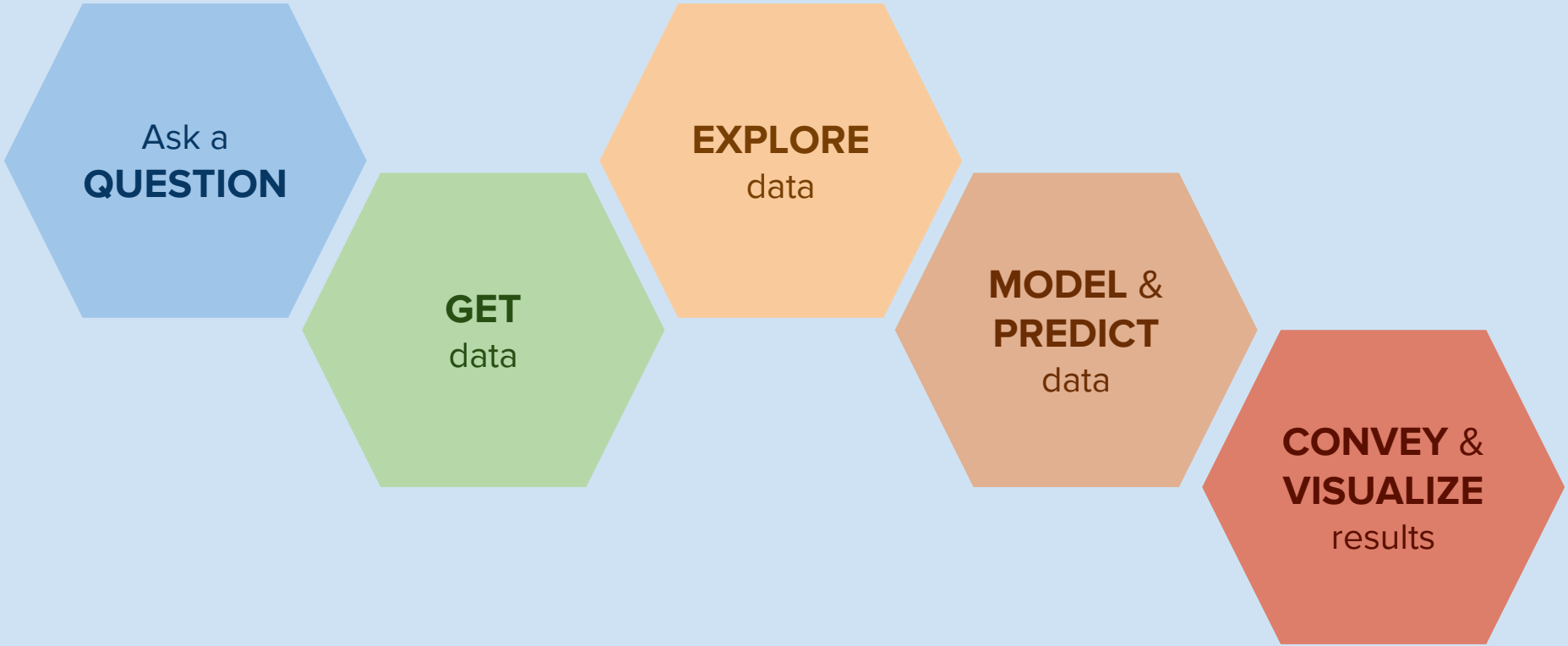
DS 1.1: Data Analysis & Visualization

The Normal Distribution!

Data Science Process

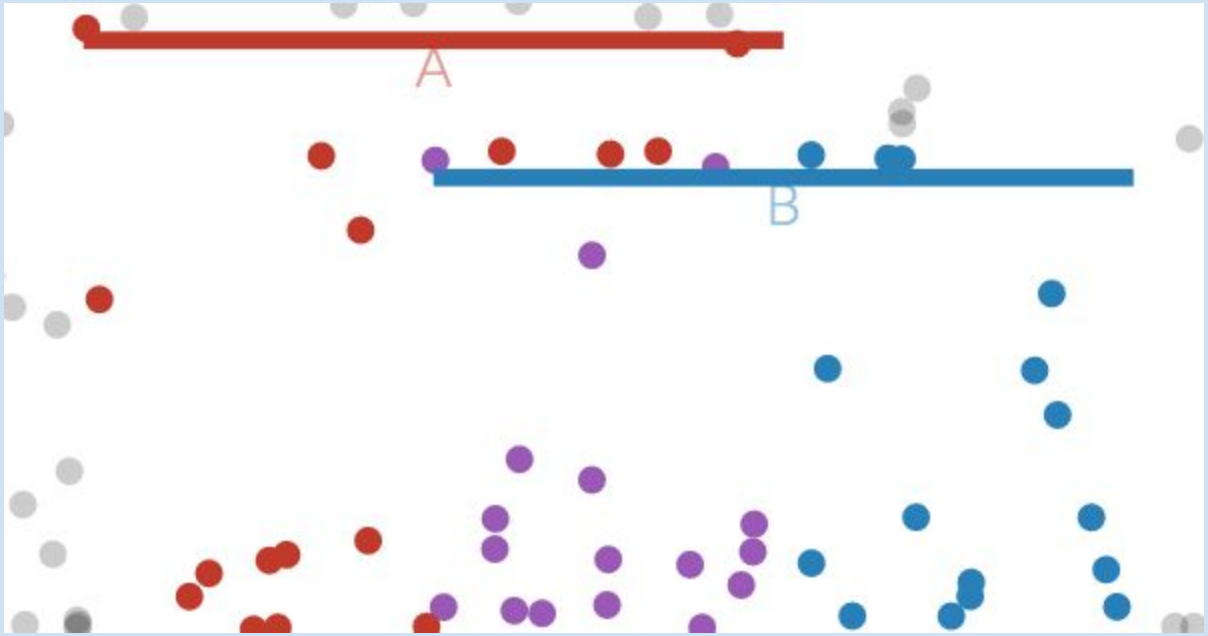


KEY LANGUAGES

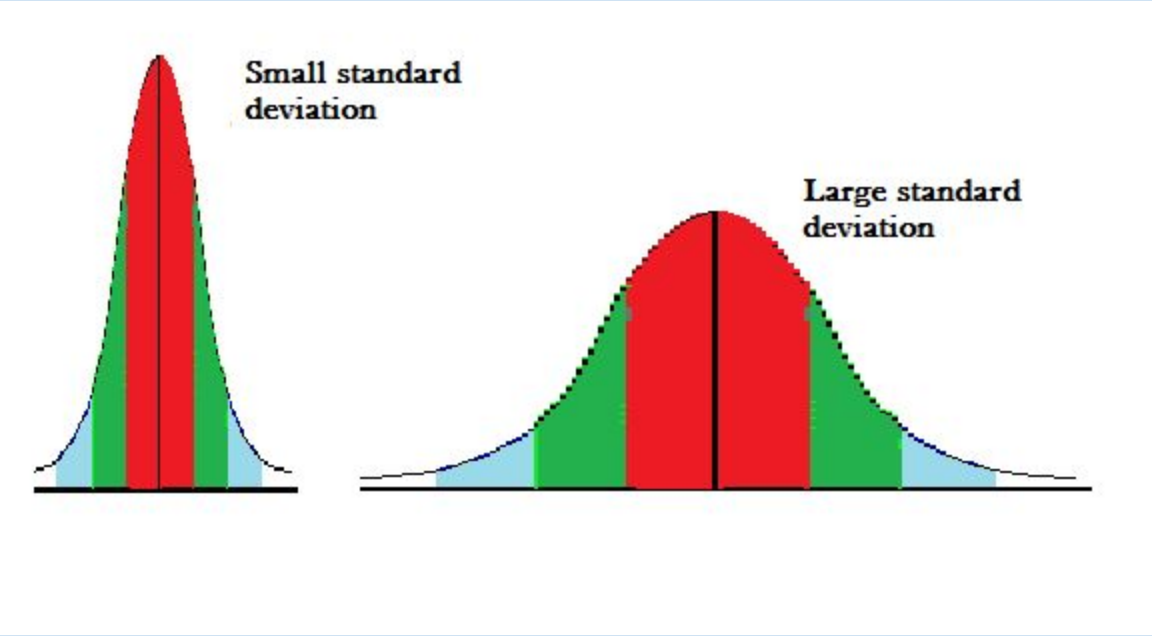
DATA ANALYSIS

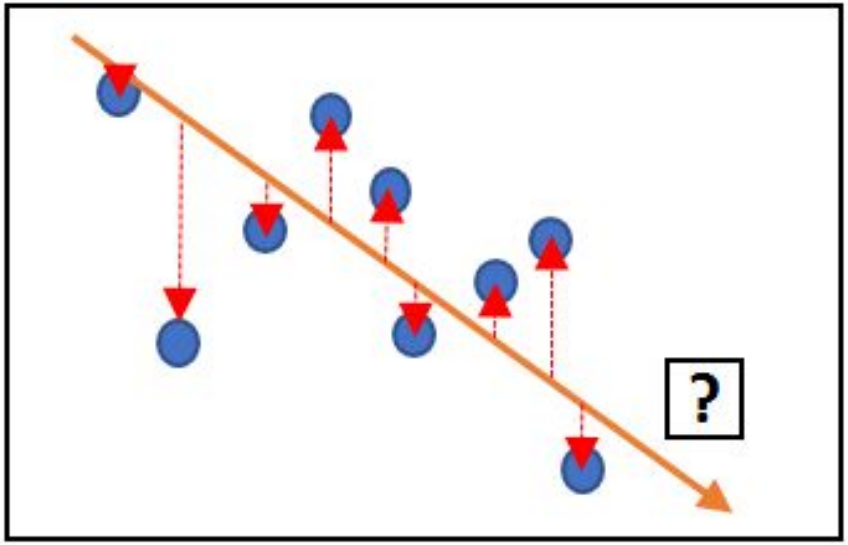
VISUALIZATION TOOLS

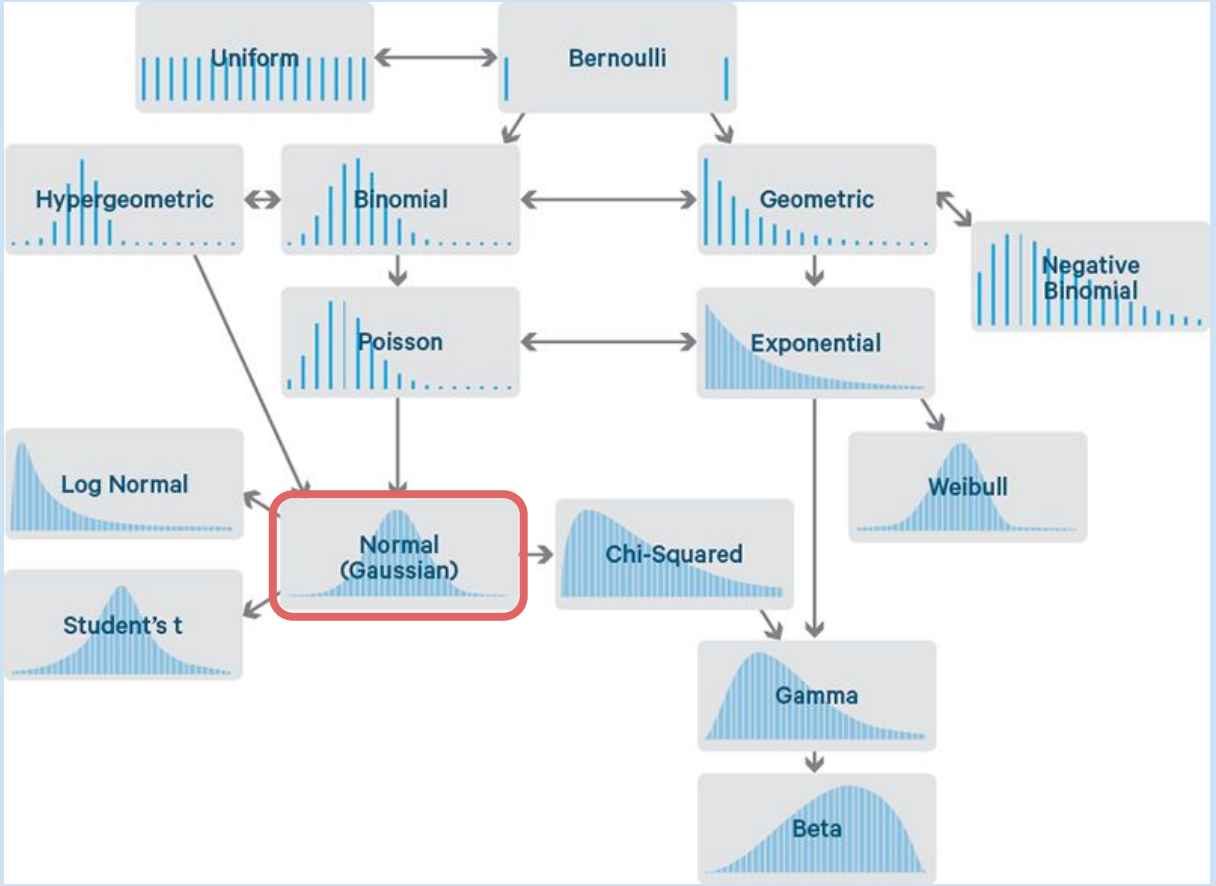
ESSENTIAL RESOURCES

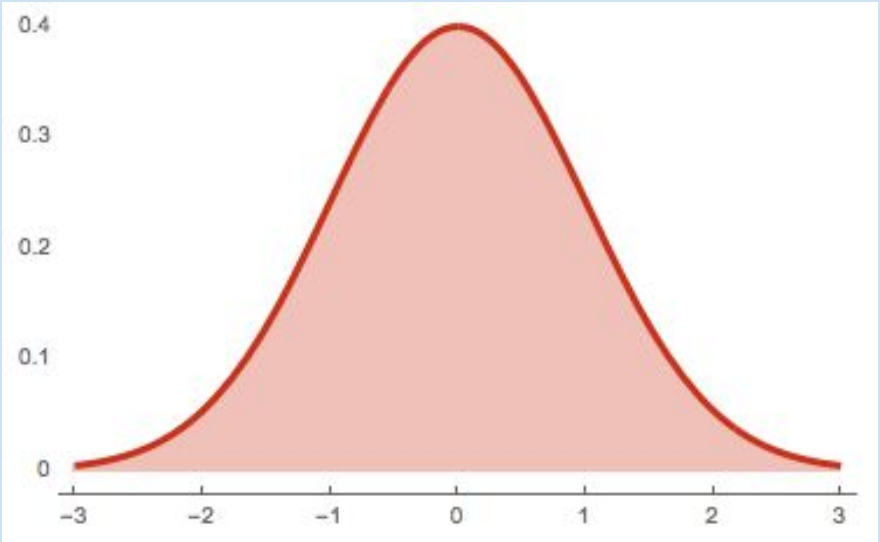


Click here!

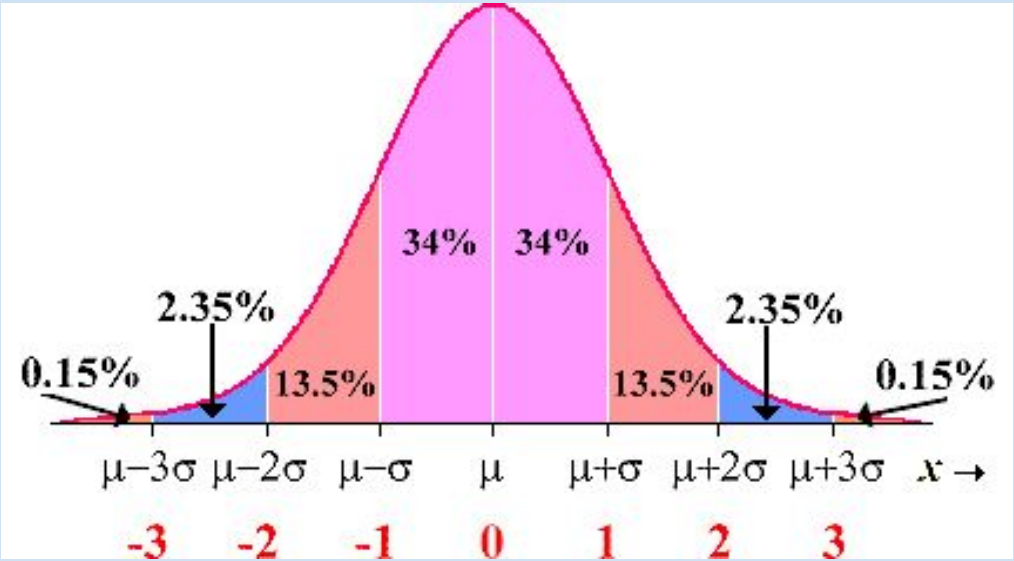

Small standard deviation
Large standard deviation




Uniform
Bernoulli
Hypergeometric
Binomial
Geometric
Negative Binomial
Poisson
Exponential
Log Normal
Normal (Gaussian)
Chi-Squared
Weibull
Student's t
Gamma
Beta


0.4
0.3
0.2
0.1
0
−3
−2
−1
0
1
2
3

M = M = M


34%
34%
2.35%
2.35%
0.15%
13.5%
13.5%
0.15%
μ−3σ μ−2σ μ−σ μ μ+σ μ+2σ μ+3σ x →
-3 -2 -1 0 1 2 3

Z-Score!

$$z = \frac{X - \overline{X}}{S}$$

where z is the standard score,
S = the standard deviation of a sample,
X = each value in the data set,
$\overline{X}$ = mean of all values in the data set.

The ***Central Limit Theorem***