

# Statistical Distributions

Presented by David John Baker  
December 2019

 **FLATIRON SCHOOL**

# Statistical Distributions ?

- Distributions are to statistics as data structures are to programming
- Reflect assumptions about the underlying processes
- Allow us to make generalizations about data that we have not yet seen

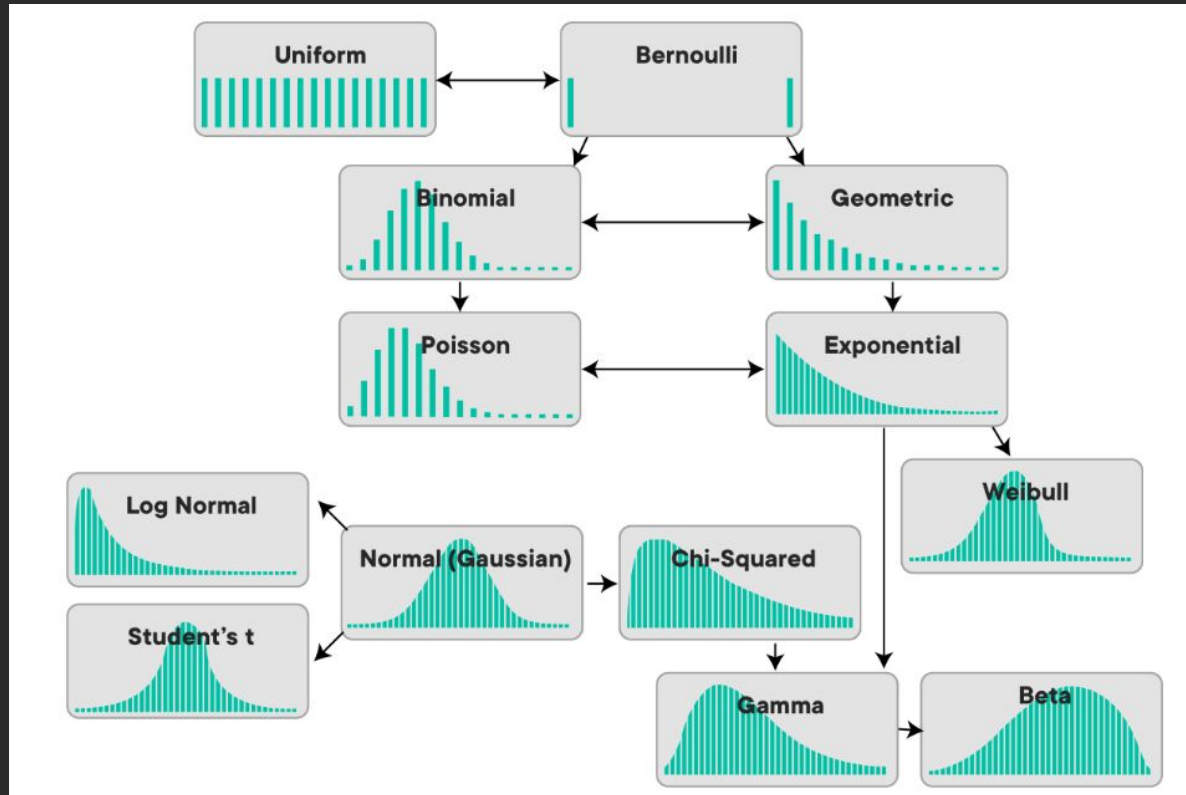


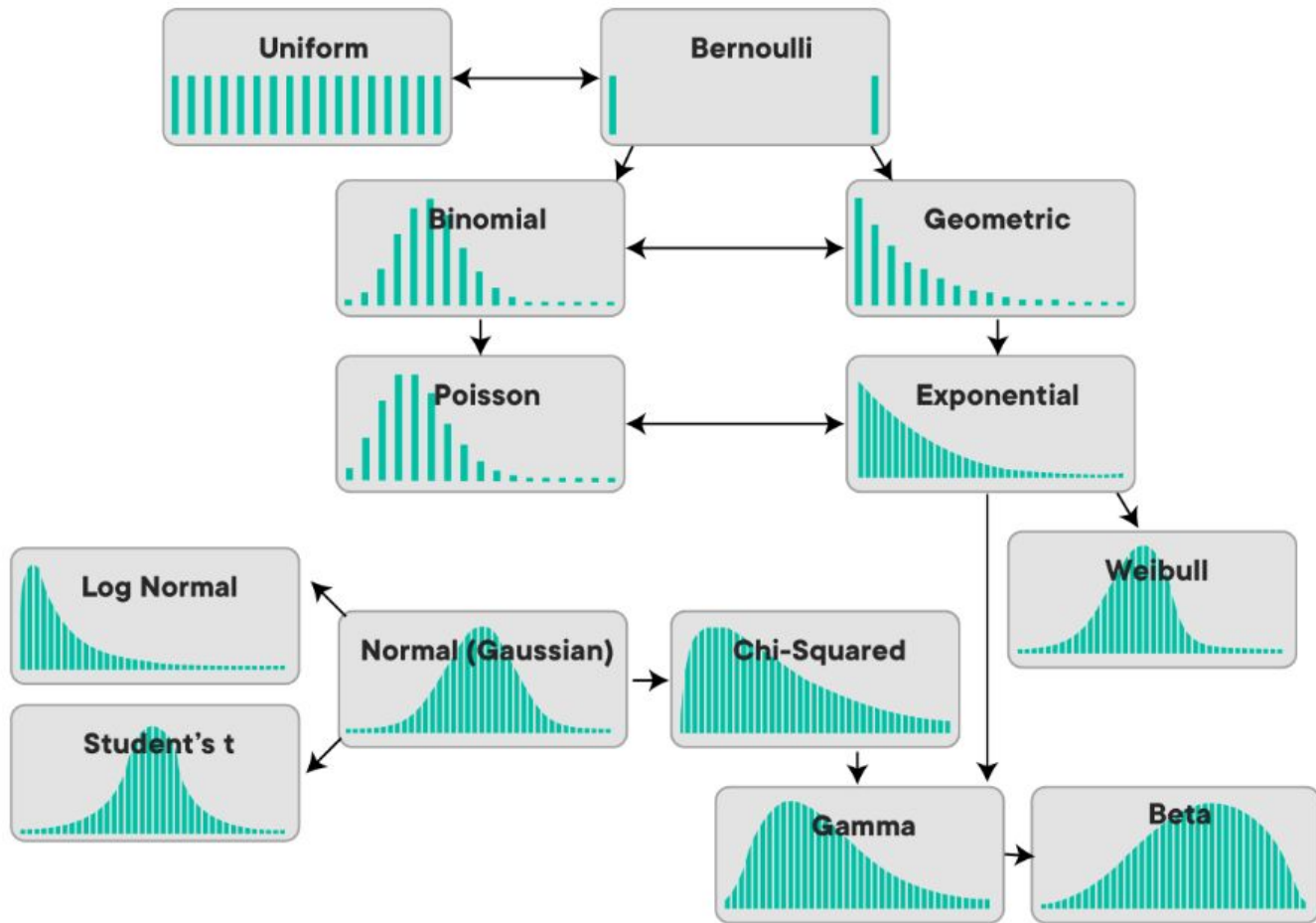
## Lesson Goals

- Name and Describe Distributions
- Explain Difference between Continuous and Discrete Distributions
- Know PMF, PDF, CDF
- Describe Normal Distribution + (Standard Normal)
- Give examples of distributions



# Distributions → Known Shapes



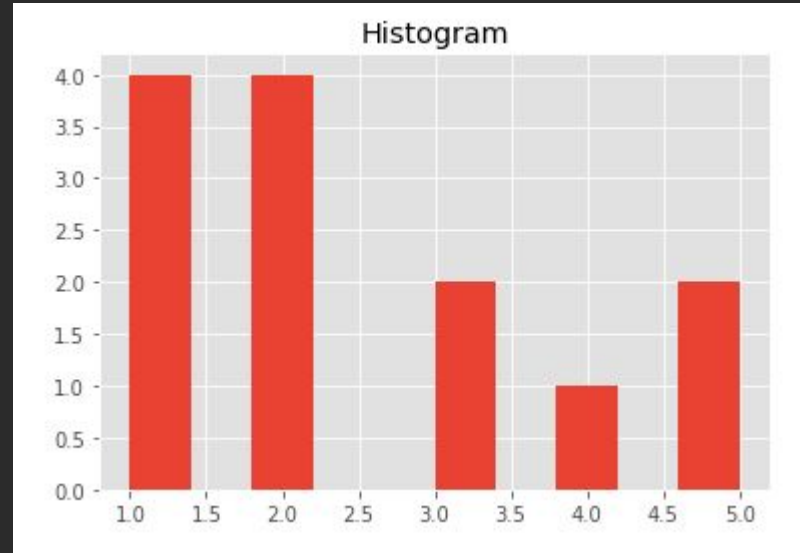
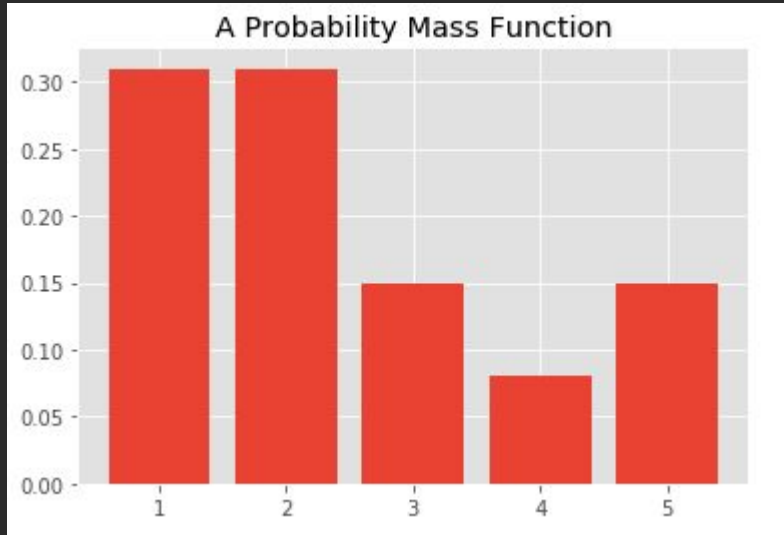


# Discrete vs Continuous

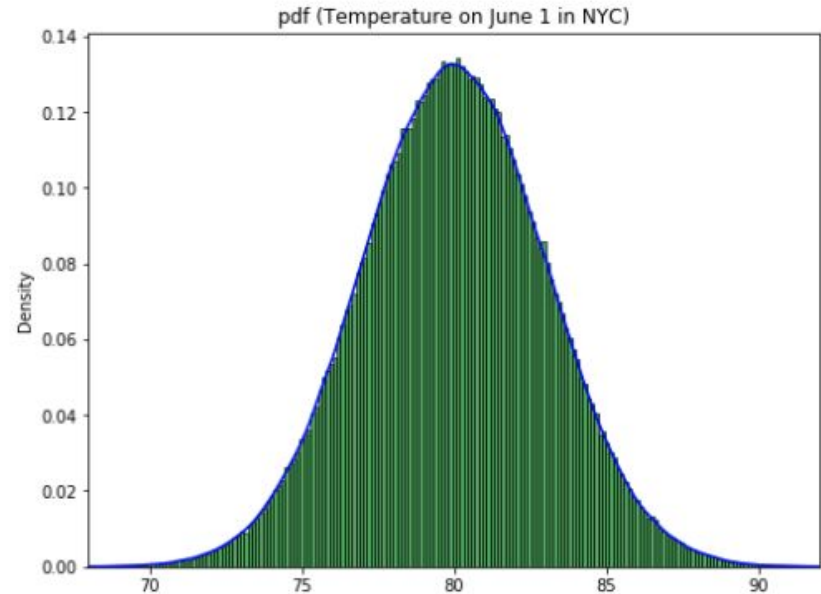
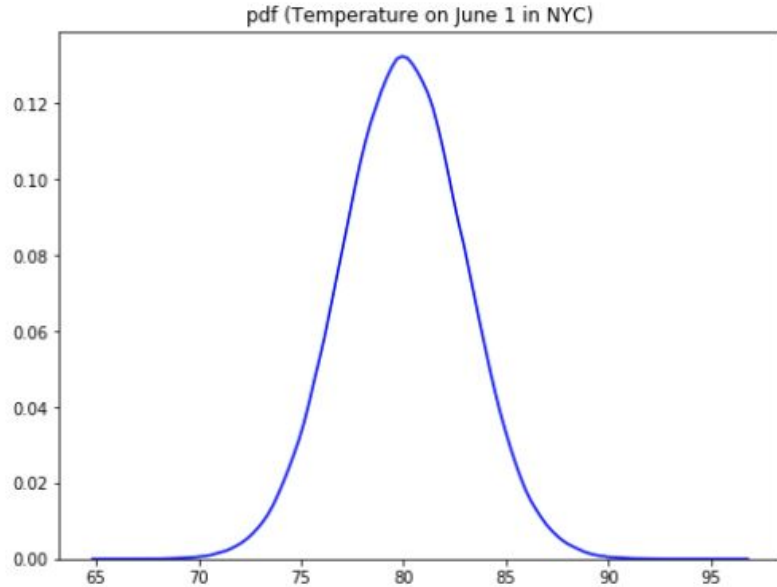
- Discrete distributions have finite sample space (ex: dice, categories of drawing balls out of an urn, species of animals)
  - Bernoulli
  - Poisson
  - Uniform
- Continuous distributions have infinite event space (ex: heights, time, financial data)
  - // • Normal (Standard Normal)

# Probability Mass Functions (PMF)

(~A Normalized Barchart~)

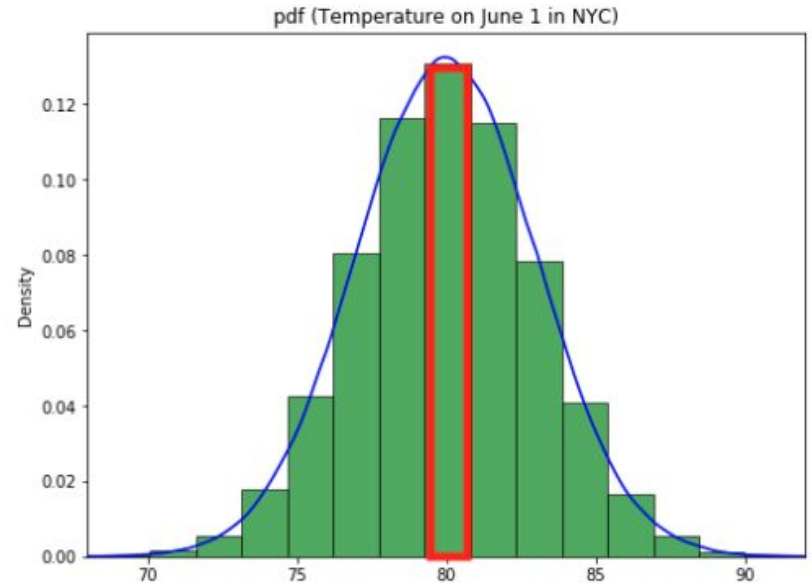
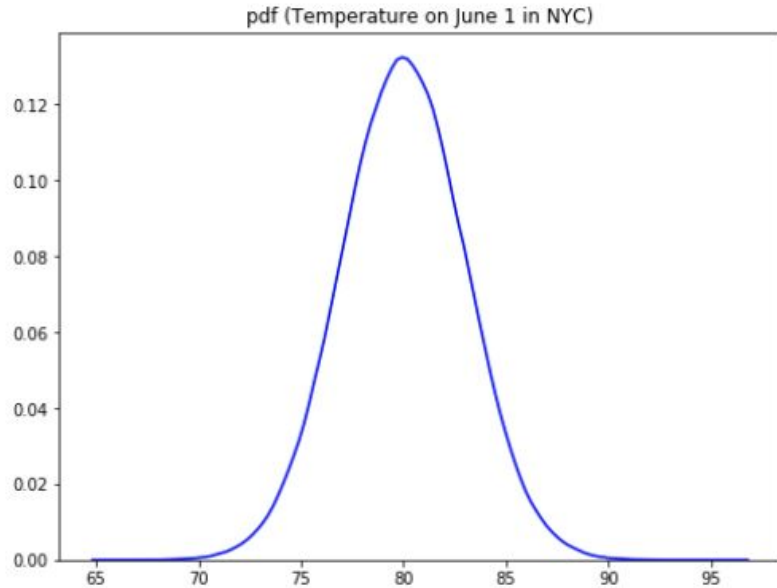


# Probability Density Functions (PDF) (A Normalized Histogram)

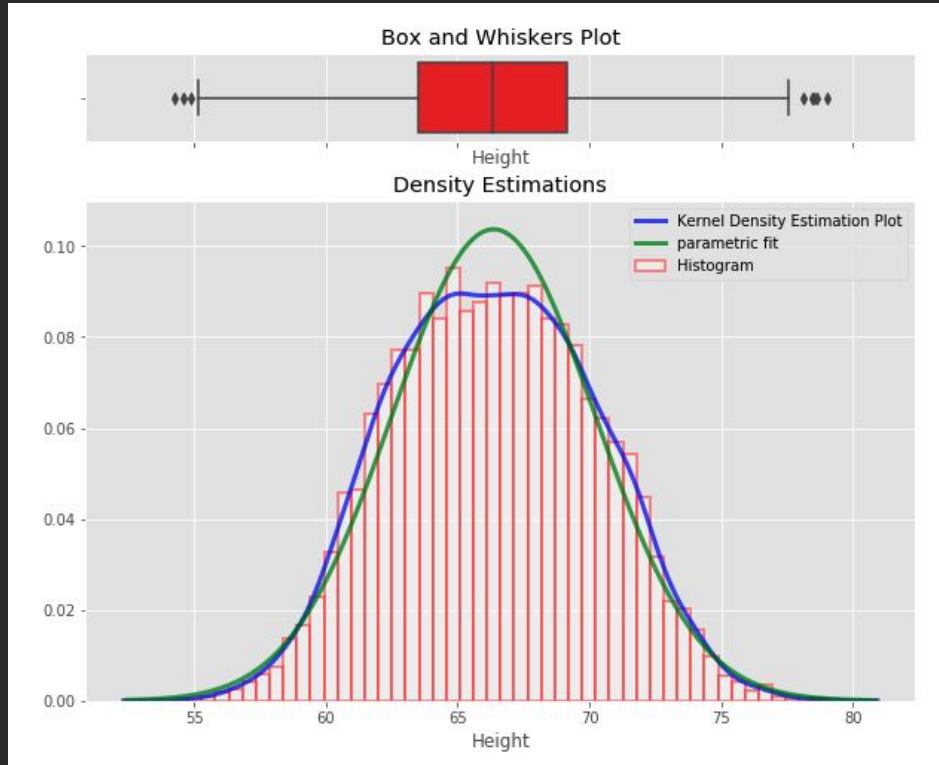




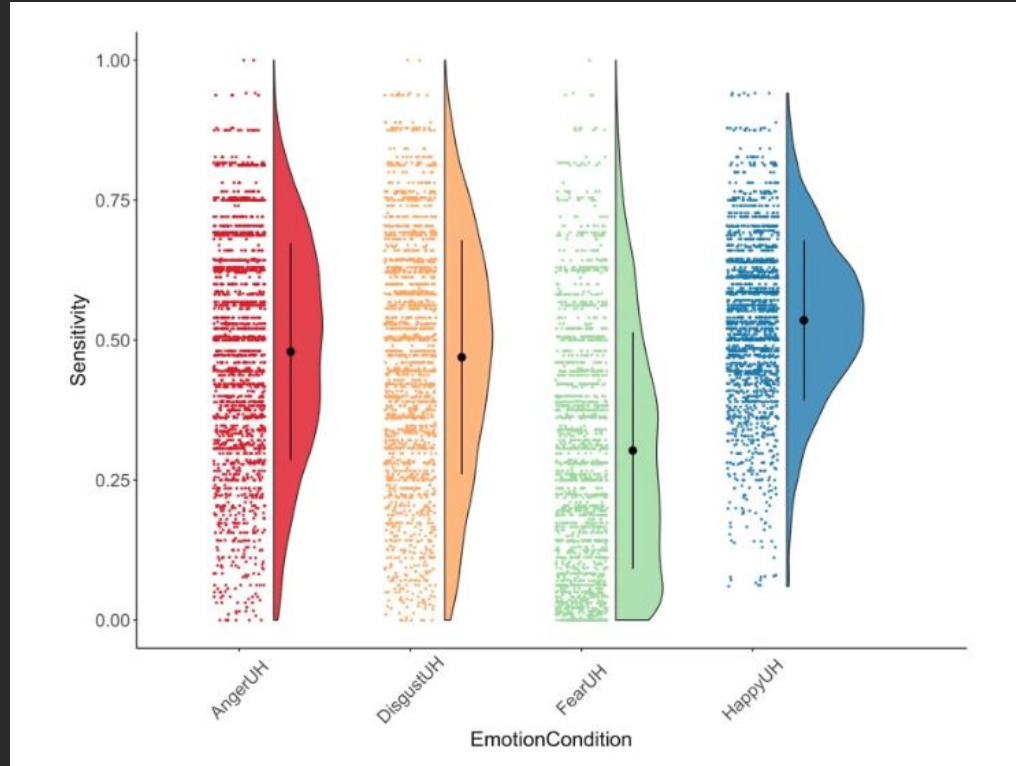
# Probability Density Functions (PDF) (A Normalized Histogram)



# Probability Density Functions (PDF) (A Normalized Histogram)

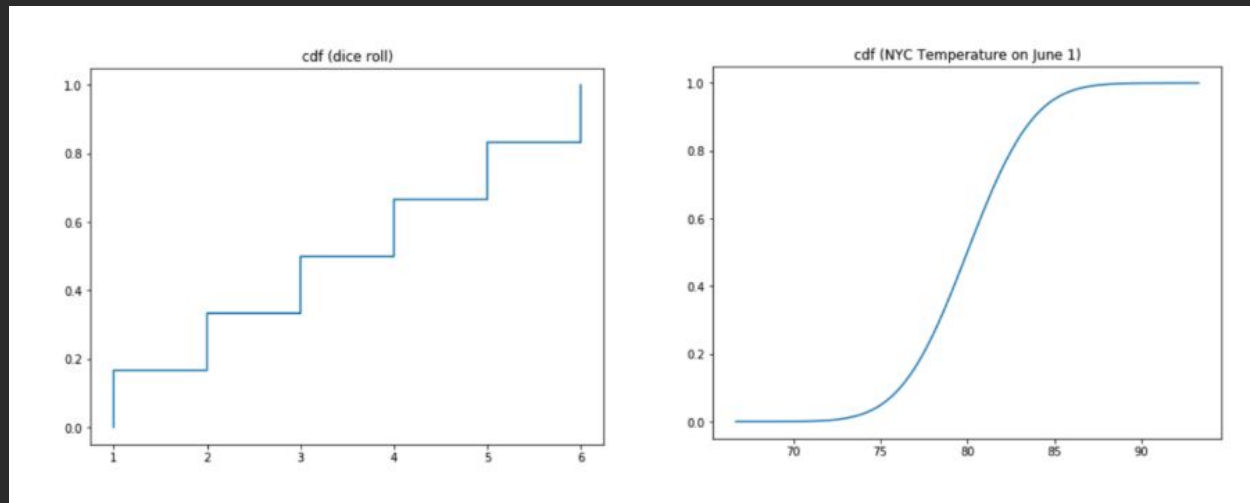


# Probability Density Functions (PDF) (A Normalized Histogram)



# Problems with PMF and PDF?

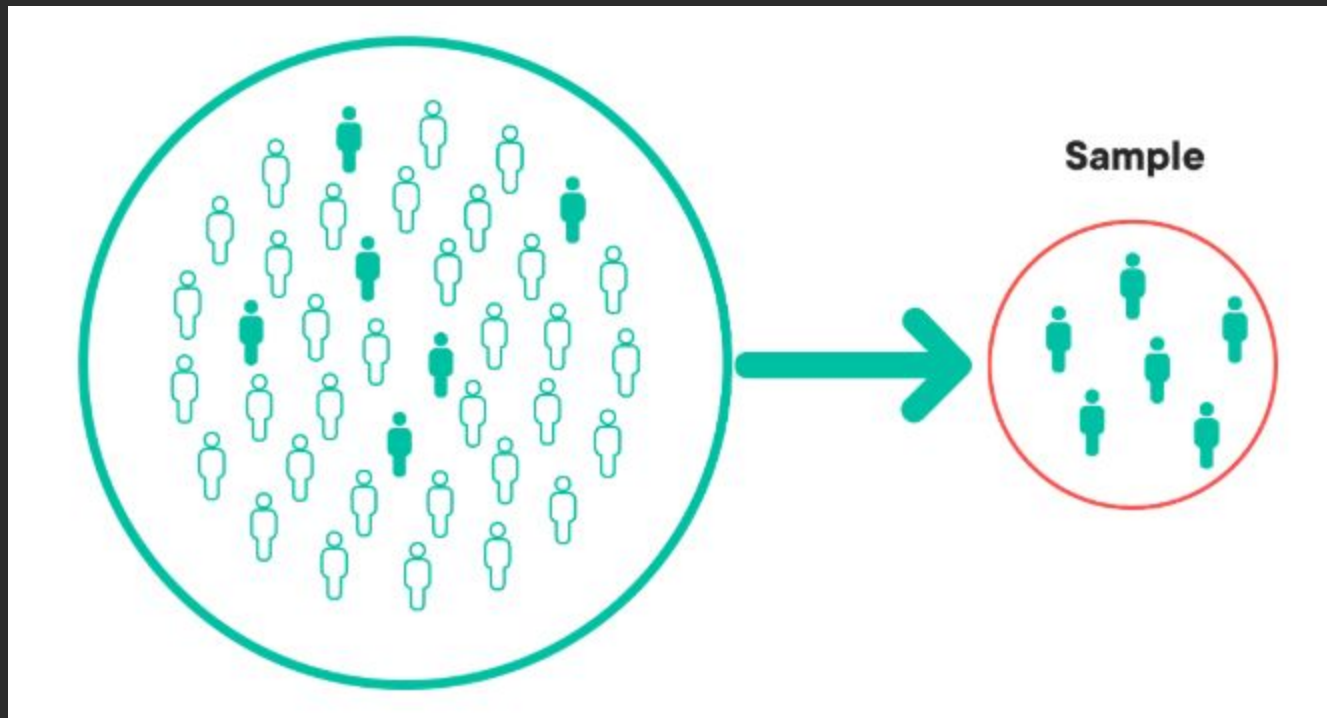
- Interpreting PMF, not to so bad?
- But what about if you wanted to say what is the probability of getting an exact value from a PDF? ( a point probability)
- CDF!



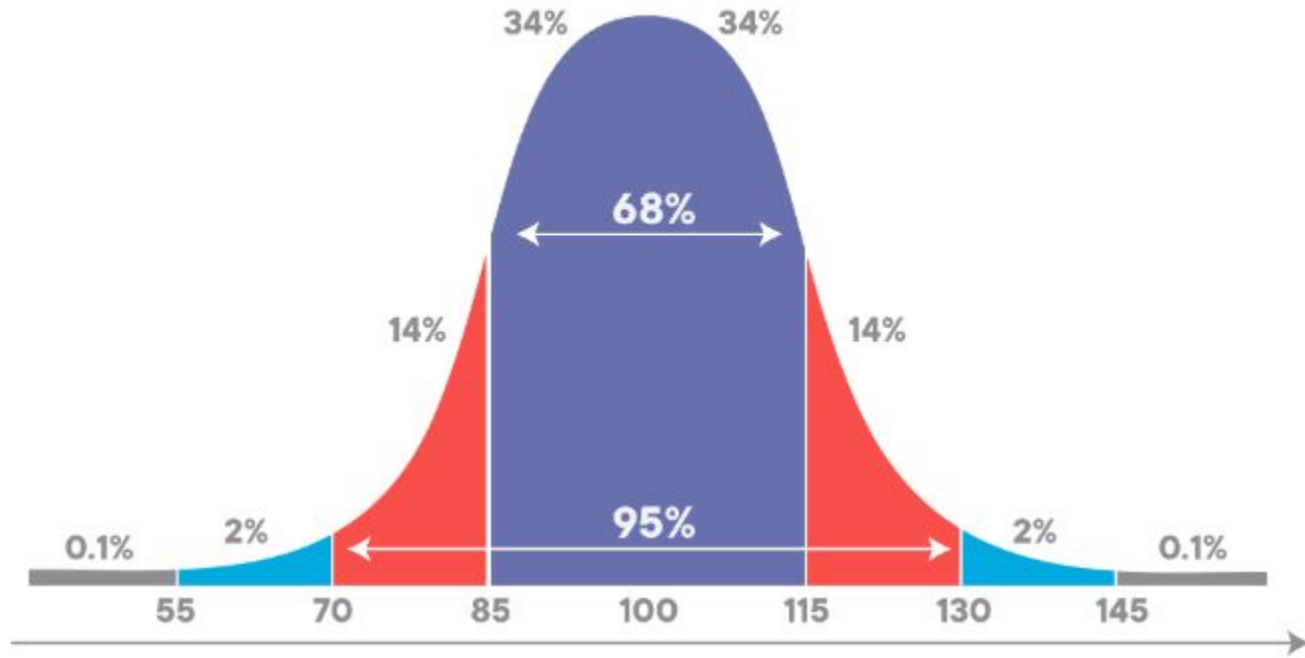
# Normal Distributions

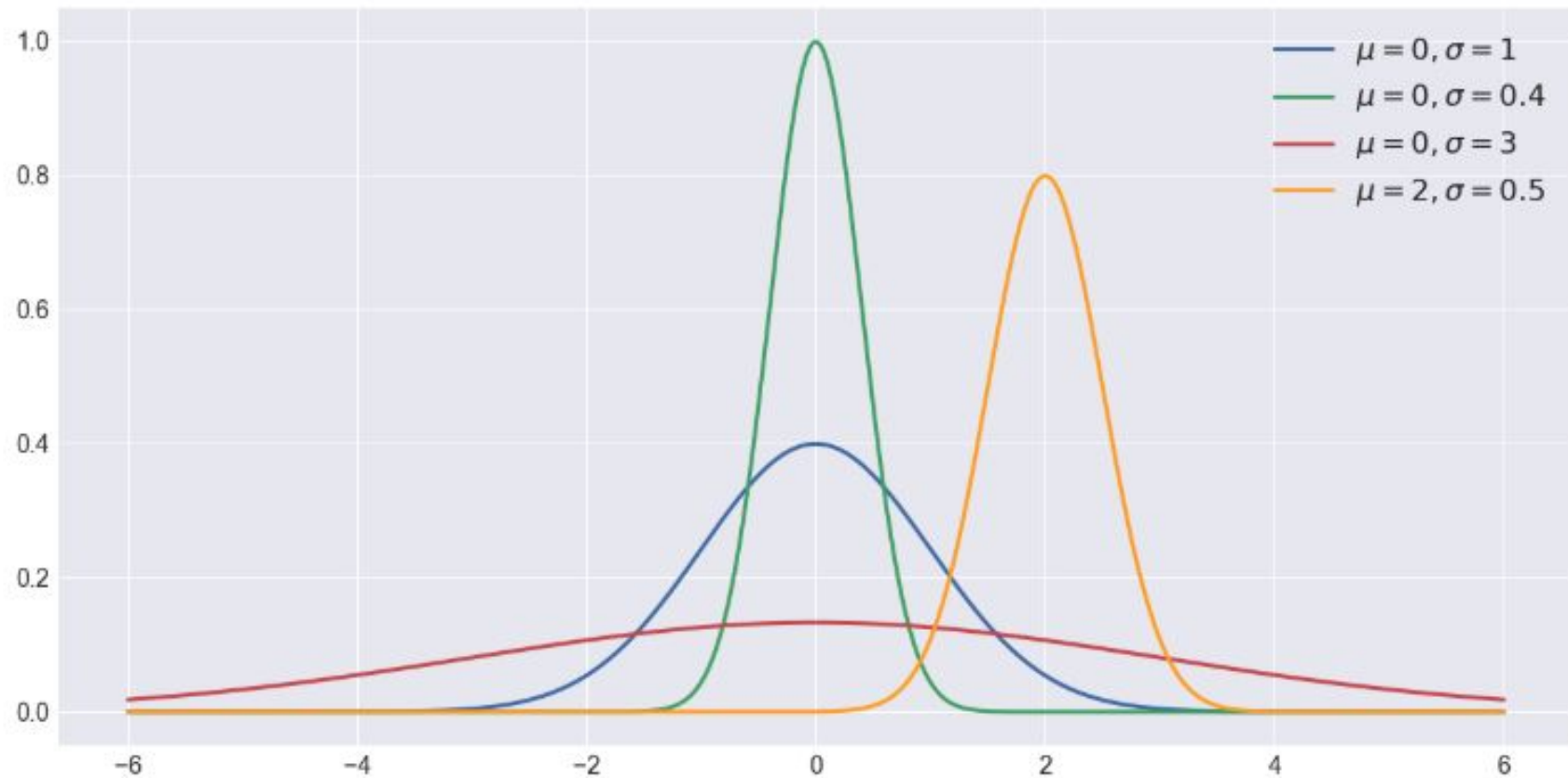
- One of most common distributions found in natural world
- Normal is not a value judgment
- Defined by a mean ( $\mu$ ) and standard deviation ( $\sigma$ ) and general symmetry
- Understanding the properties of the normal distribution will allow us to generalize about data we have not yet seen!
- Reflect on relationship between population, sample, and sampling distribution.



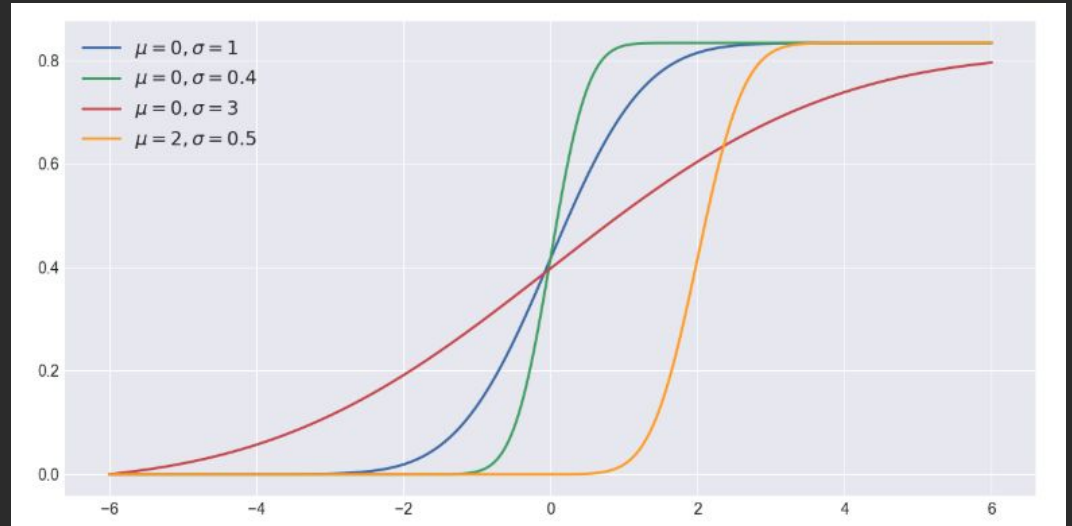
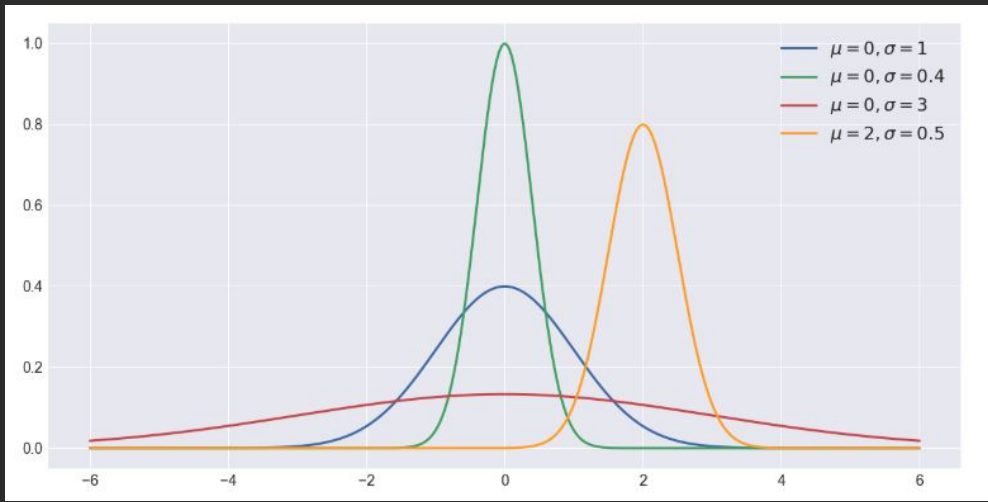


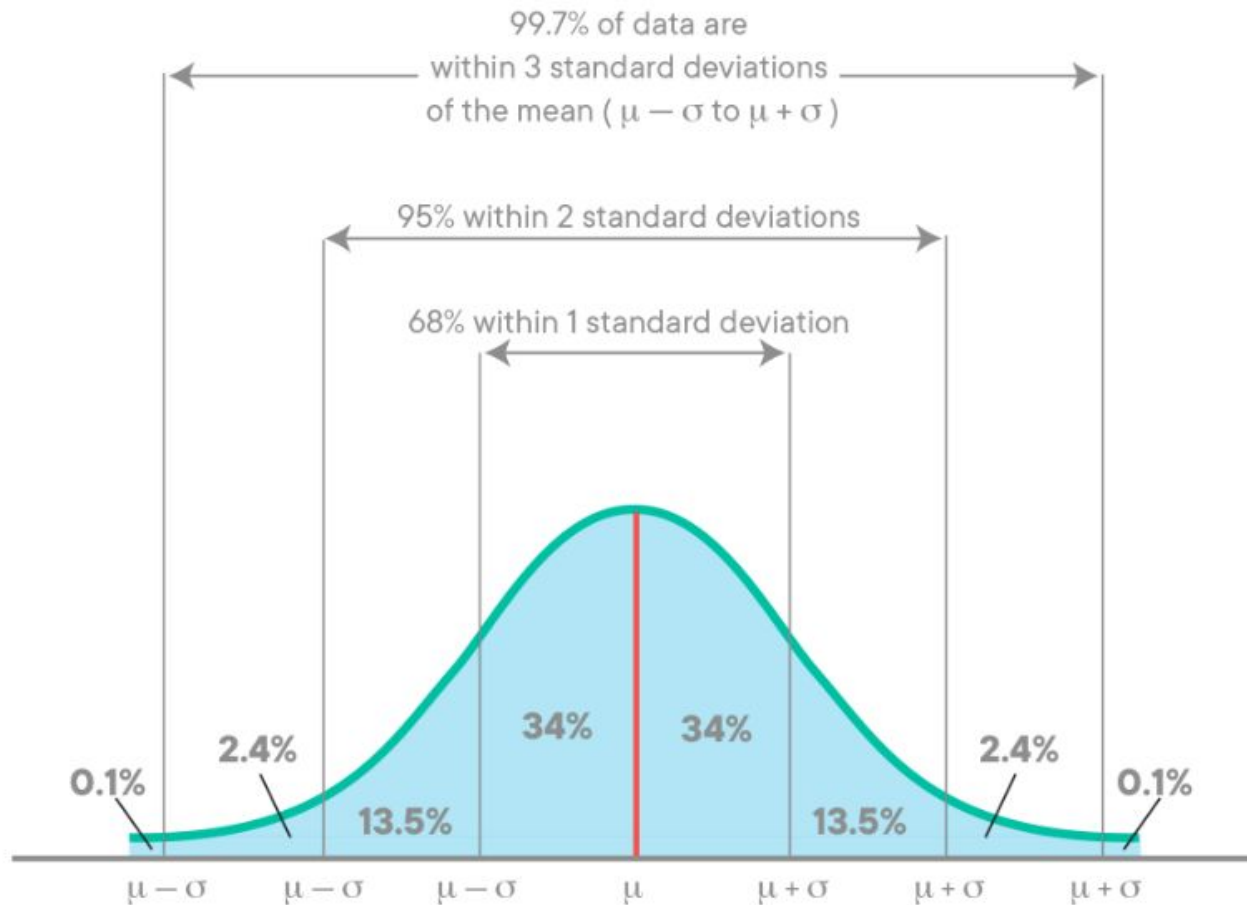
## IQ Score Distribution









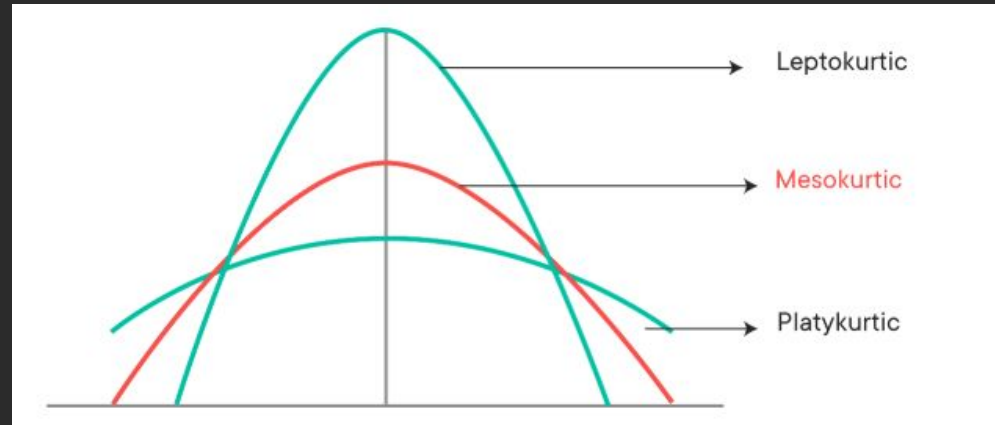
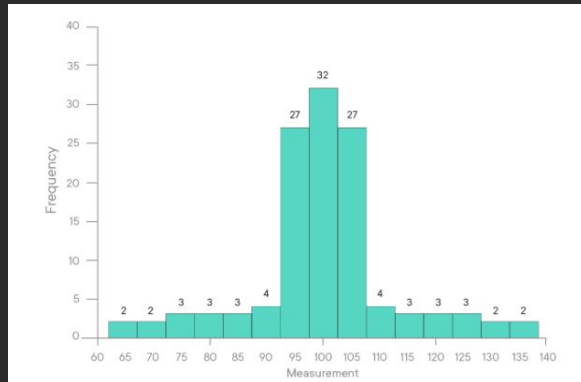
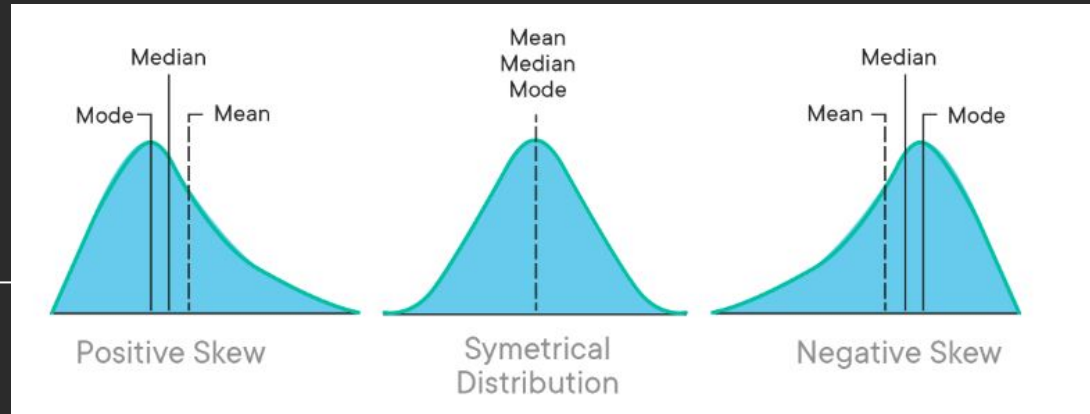


## Central Limit Theorem

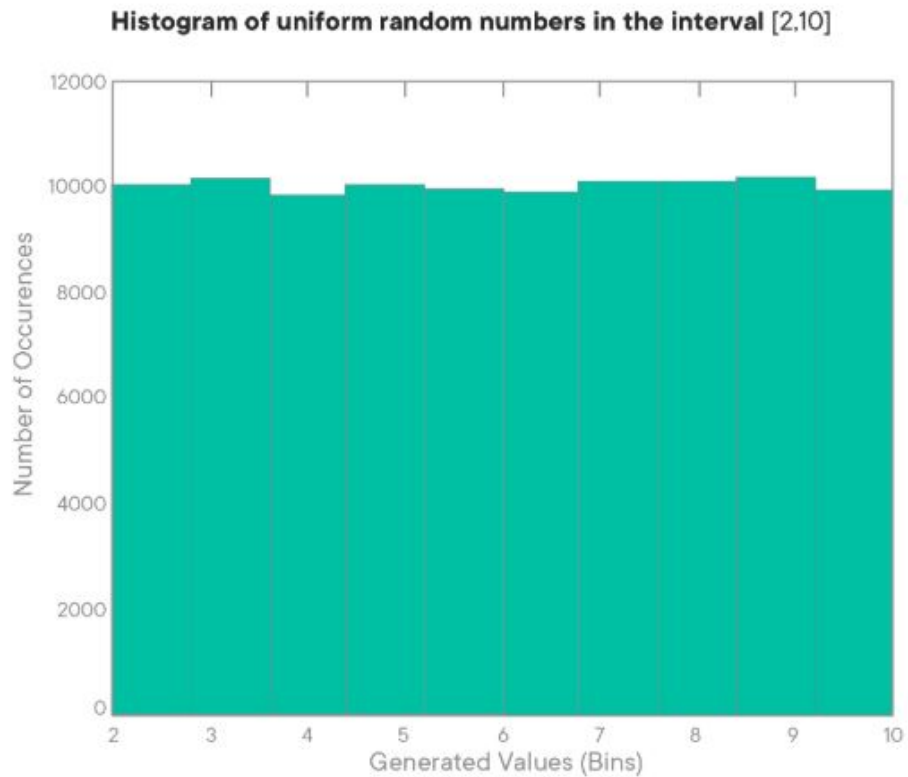
- When you add a large number of independent random variables, irrespective of the original distribution of these variables, their sum tends towards a normal distribution.



# Skew and Kurtosis



# Uniform Distribution



# Practice

- In the last portion of class, want to do a class activity
  - Collect heights of all people in class in CM
  - Create small analysis script that is going to import the data
  - First create a plot that plots the raw data
  - Check if that data is normal (visual inspection)
  - Convert all scores to z scores

$$z = \frac{x - \mu}{\sigma}$$

$\mu$  = Mean  
 $\sigma$  = Standard Deviation



# Checking For Understanding

- Explain discrete vs continuous distributions
- Explain difference between PDF, PMF, CDF
- Explain normal distributions
- Standard Normal Distributions
- Z Scores

