SESSIONS 4

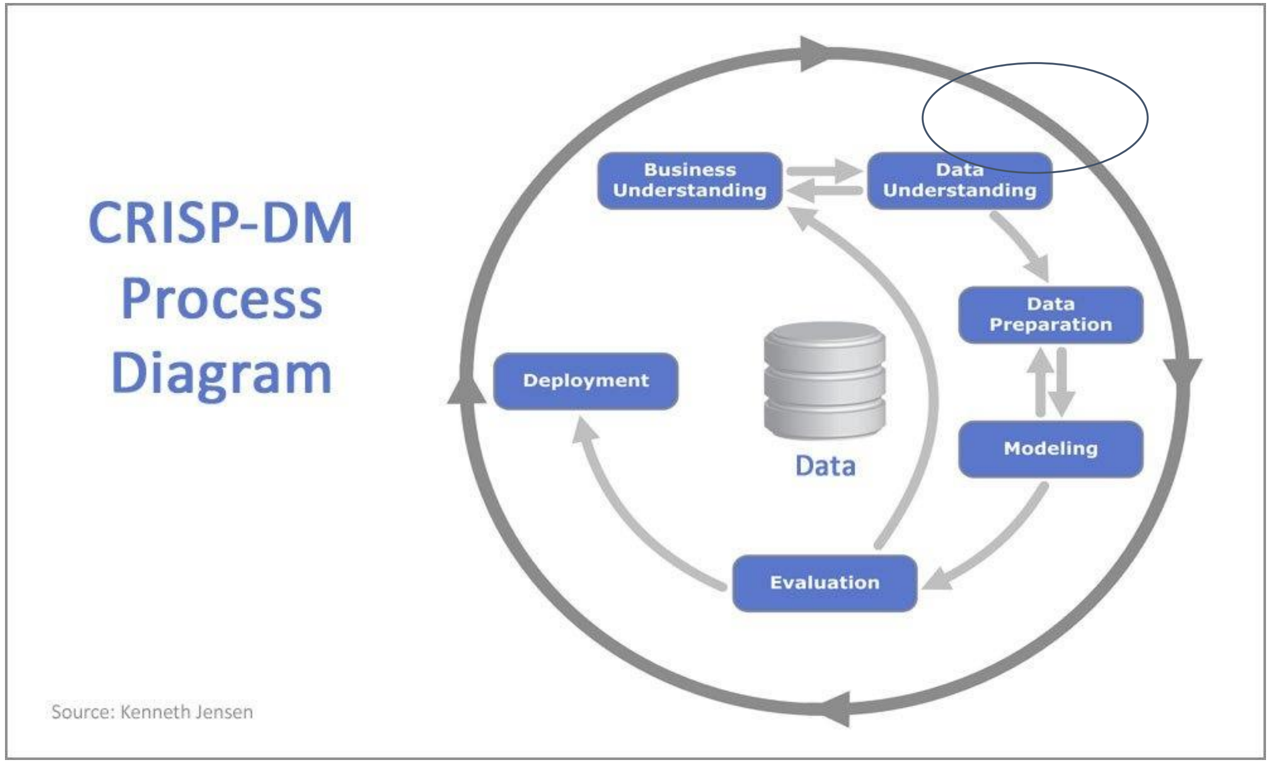
Descriptive Statistics: Graphical Summary

Data Science Program

Purwadhika
Startup and Coding School

Outline

- Frequency table
  - For numerical
  - For categorical
- Cross tabulation
- Graphical Summary



Source: Kenneth Jensen

Purwadhika
Startup and Coding School

Frequency Table

Frequency Table for categorical variable

Day	Visitor Count
Saturday	87
Sunday	76
Thursday	62
Friday	19

Frequency Table for numerical variable

Tip Range ($)	Visitor Count
0 - 2.5	108
2.5 - 4	95
4 - 5.5	29
5.5 - 7	9

Cross Tabulation / Contingency Table

Cross Tabulation : Frequency

Day	Visitor Count (Male)	Visitor Count (Female)
Saturday	87	32
Sunday	76	9
Thursday	62	28
Friday	19	18

Cross Tabulation : Percentage

Day	Visitor Count (Male)	Visitor Count (Female)	Total
Saturday	73.1	26.9%	100%
Sunday	89.4%	10.6%	100%
Thursday	68.8%	31.1%	100%
Friday	51.3%	48.6%	100%

Purwadhika
Startup and Coding School

Graphical Summary

Numerical :

- Histogram
- Boxplot
- Scatterplot, etc

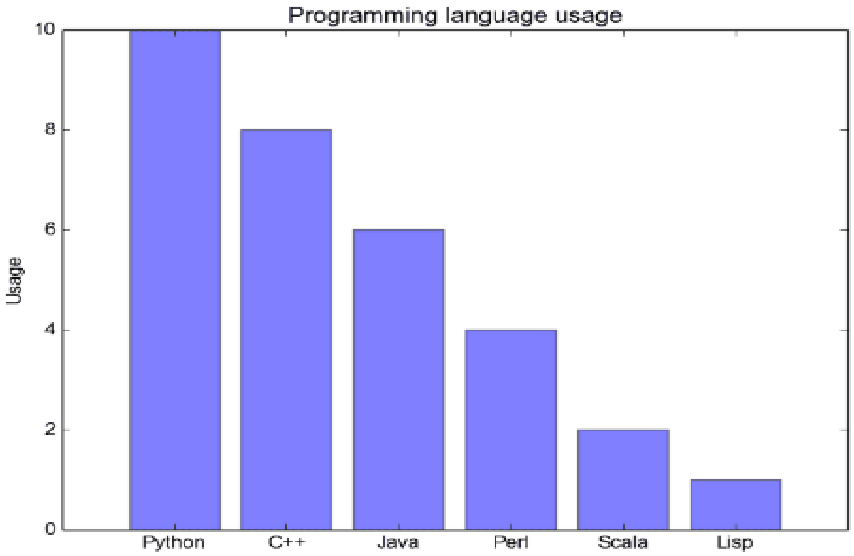
Categorical

- Pie chart
- Barchart, etc

Both numerical and Categorical:

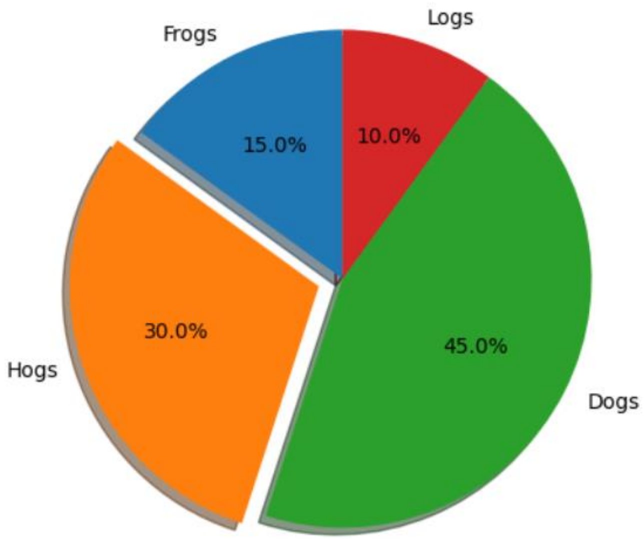
- Barplot
- Boxplot

Bar chart



- Represents **categorical data** with rectangular bars. Each bar has a height corresponds to the value it represents. It's useful when we want to **compare** a given numeric value on different **categories**.
- Each category can be consecutive and overlapping
- Can be used to see composition or comparison

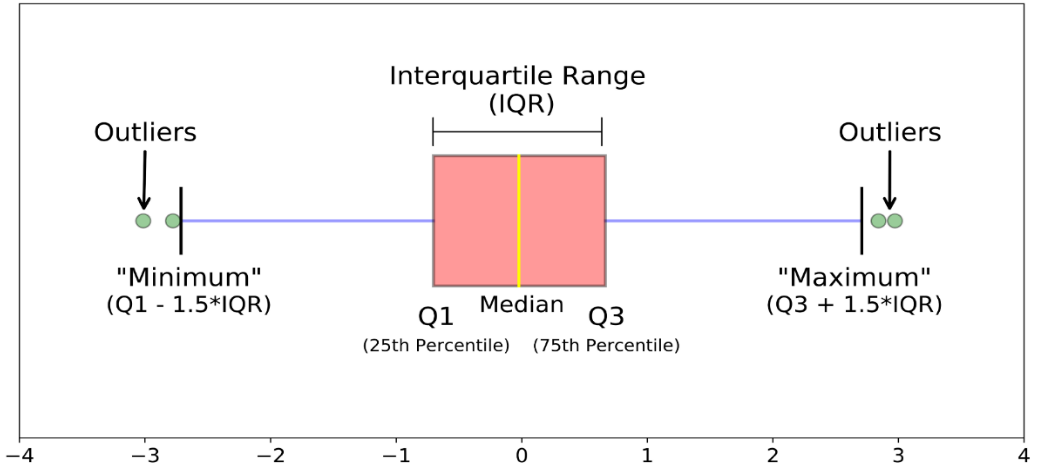
Pie chart



- A circular plot, divided into slices to show numerical proportion of the categorical data. They are widely used in the business world.
- Each category are consecutive and non-overlapping
- Main purpose is composition
- Not recommended if there are too many categories

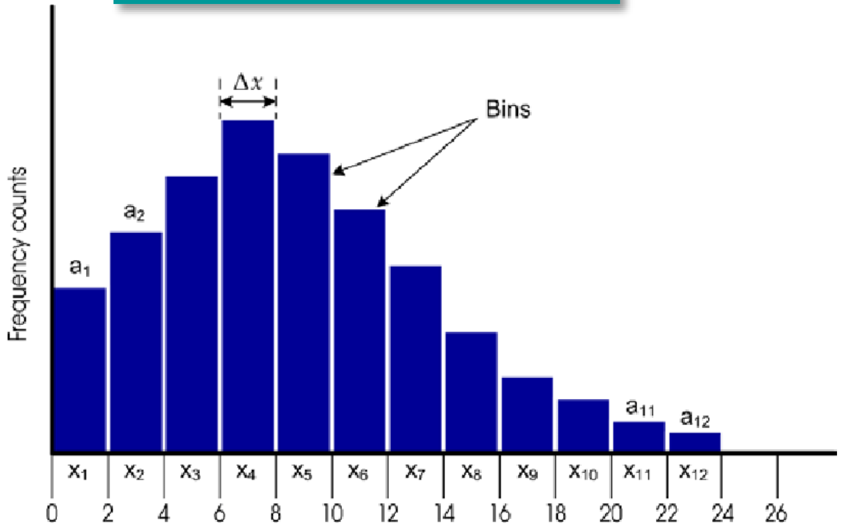
Purwadhika
Startup and Coding School

Boxplot



- Box plot, also called the box-and-whisker plot: a way to show the **distribution of values based on the five-number summary**: minimum, first quartile, median, third quartile, and maximum.
- Can be used to detect anomaly data/outliers

Histogram



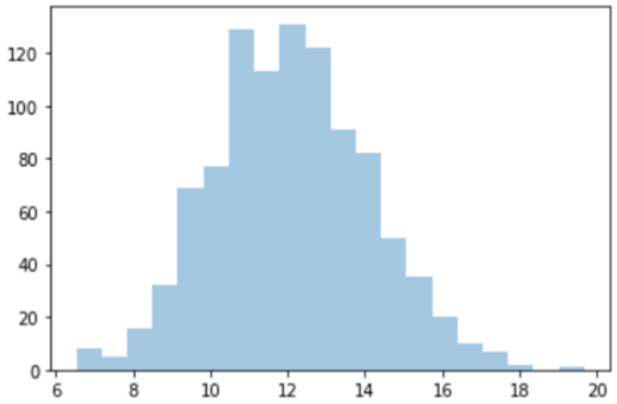
- **Histogram** is an accurate representation of the **distribution of numeric data**.
- A histogram is a graph that uses bars to portray the frequencies or the relative frequencies of the possible outcomes for a quantitative variable.

Purwadhika
Startup and Coding School

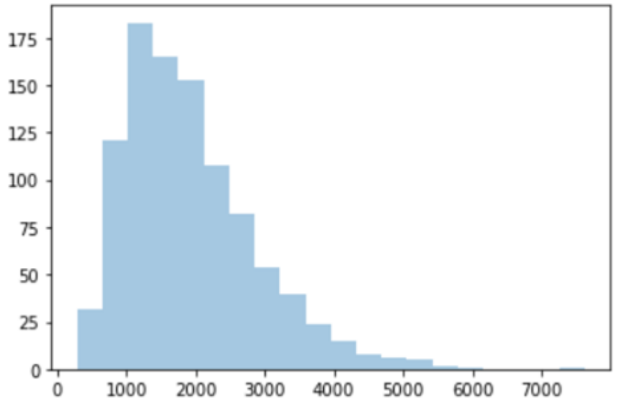
Histogram

Using histogram we can see how data spread.

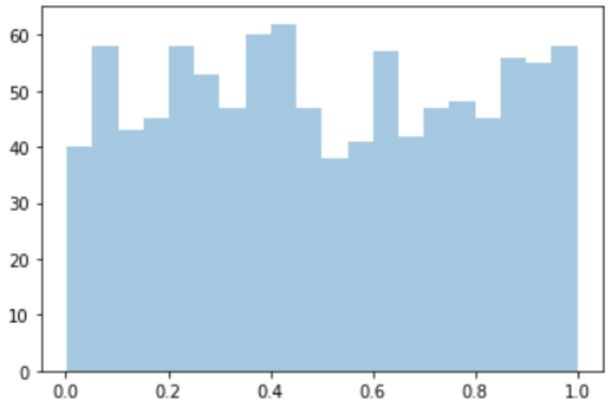
Symmetric or Normally Distributed



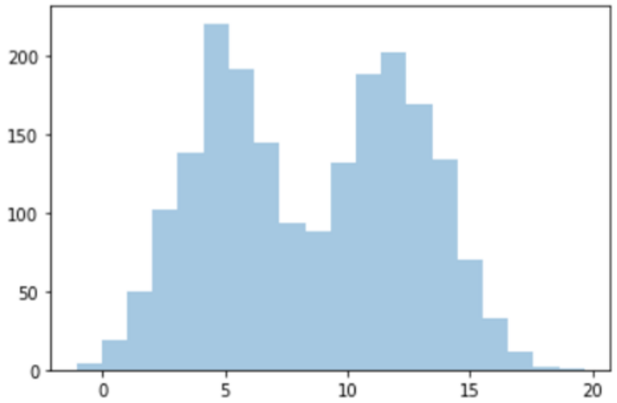
Right Skewed



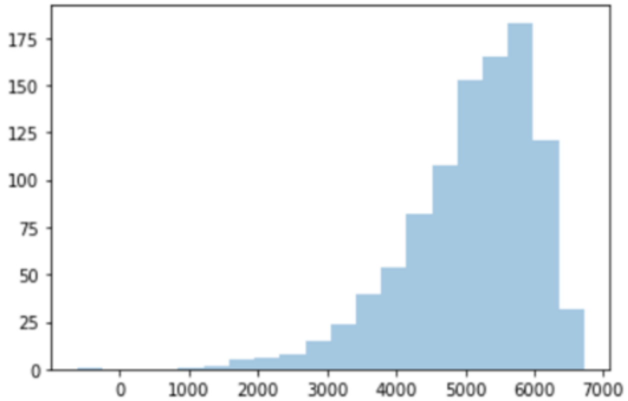
Uniform



Bimodal



Left Skewed

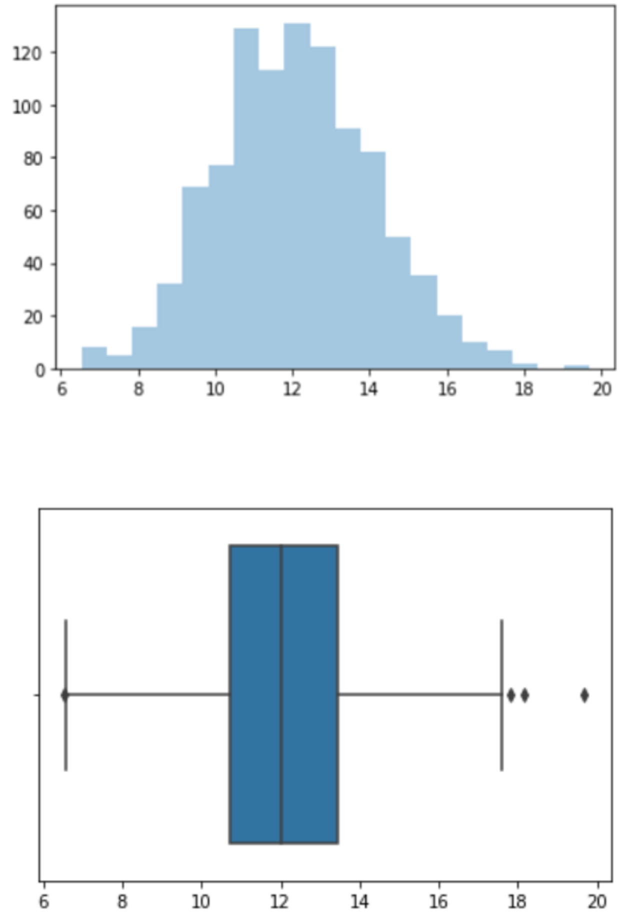


Purwadhika
Startup and Coding School

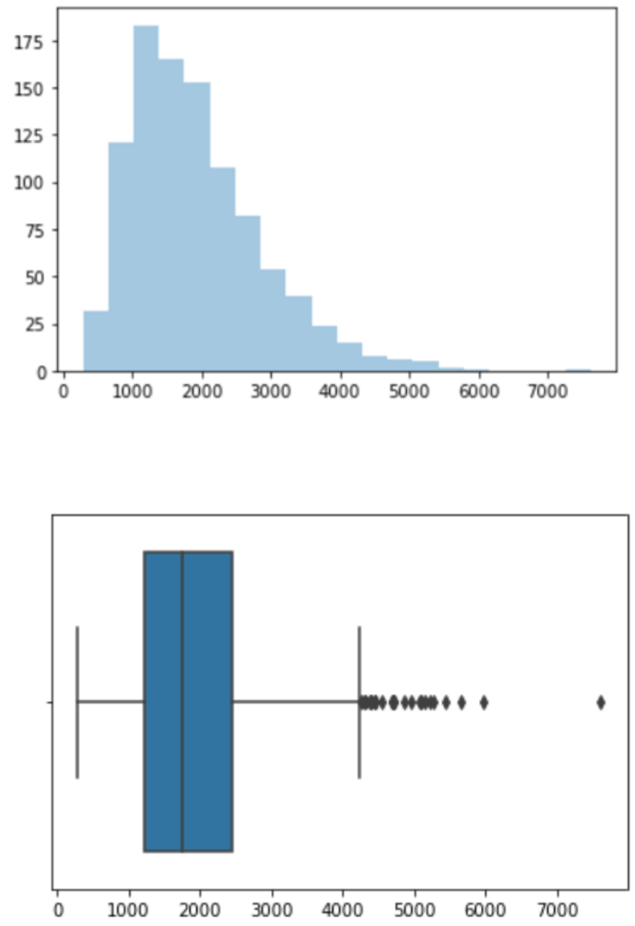
Boxplot

Using boxplot we can detect outliers

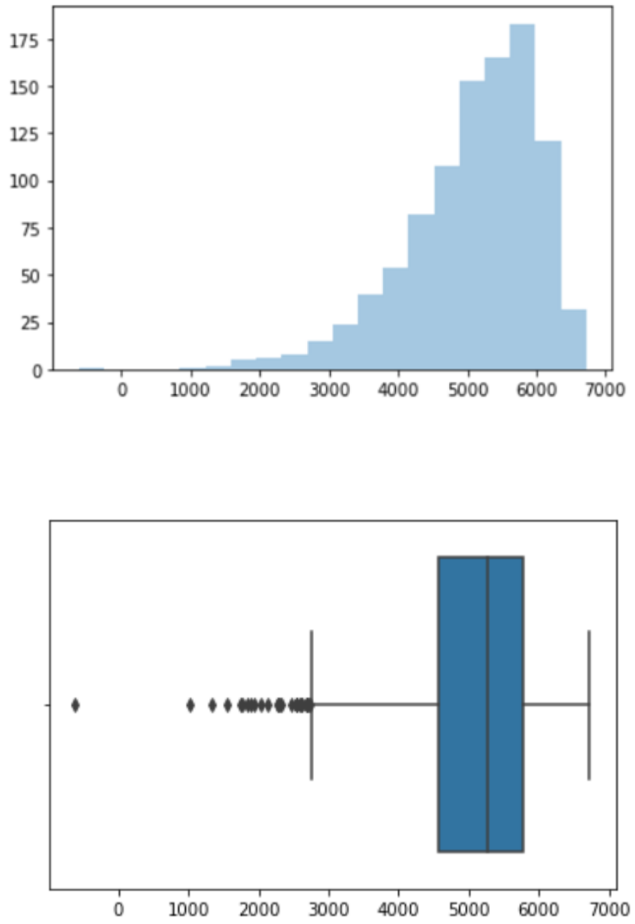
Symmetric or Normally Distributed



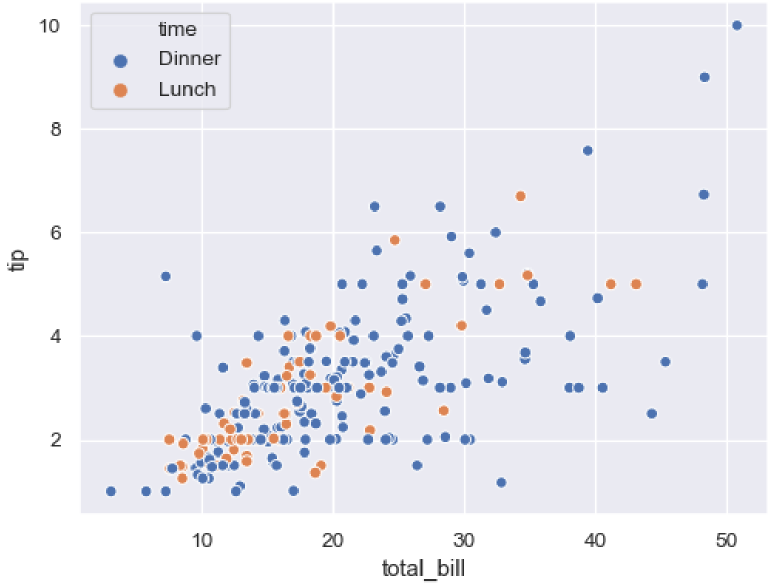
Right Skewed



Left Skewed

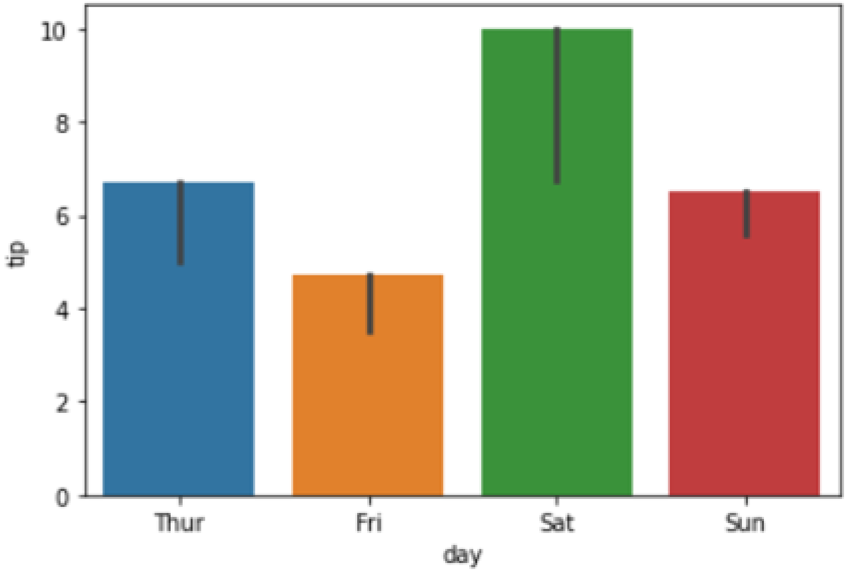


Scatterplot



- This type of plot shows **all individual data points**. Here, they aren't connected with lines.
- Each data point has the value of the x-axis value and the value from the y-axis values.
- This type of plot can be used to display **trends or correlations**.
- In data science, it shows relationship between **two numerical variables**.

Barplot

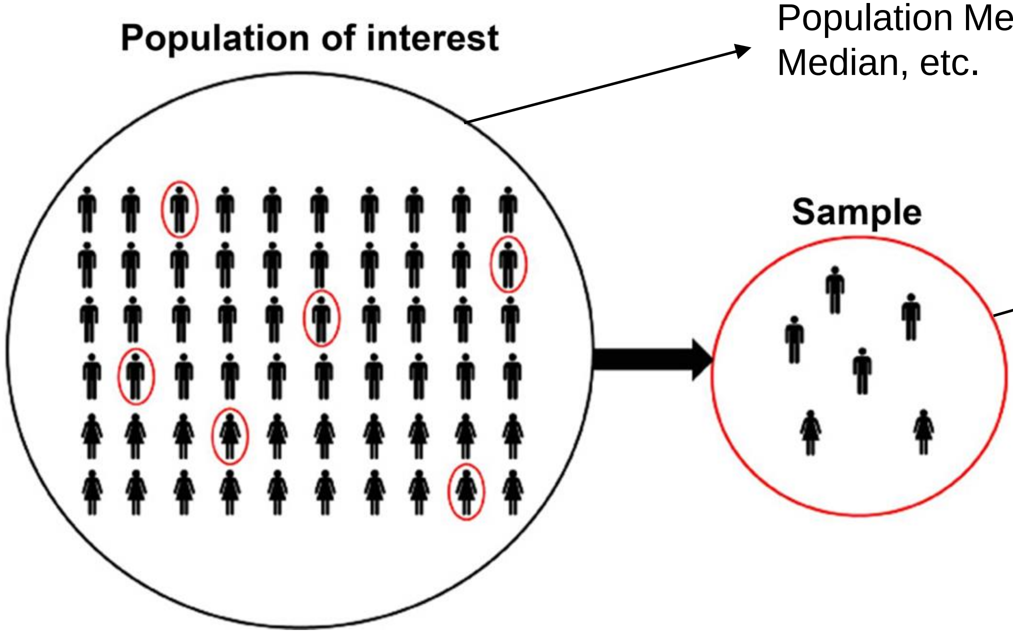


- **Barplot** is a general plot that allows you to aggregate some values in the categorical data based on some function (mean, sum, min, max, std, etc)
- In data science, it shows composition and relationship between **a numerical variables** and **a categorical variables**.

Purwadhika
Startup and Coding School

Statistics and Parameter

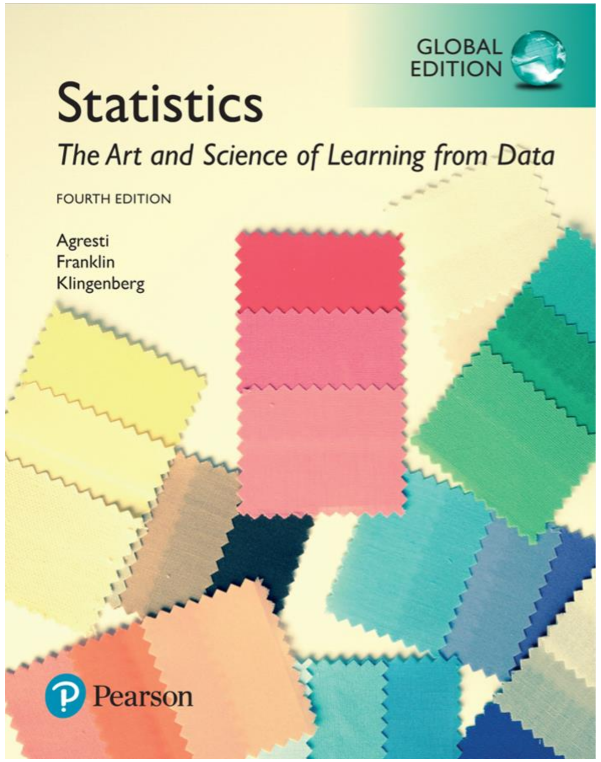
- **A parameter** is a numerical summary of the population. **A statistic** is a numerical summary of a sample taken from the population.
- Population parameter are unknown and sample statistic used to make inference about it

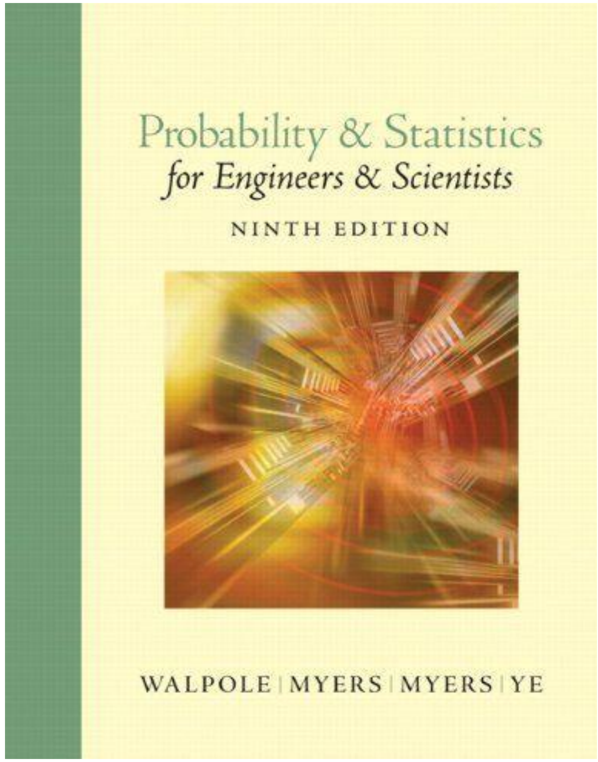


Population Parameter such as Population Mean, Population Median, etc.

Statistic sample such as Sample Mean, Sample Median, etc.

Reference





Purwadhika
Startup and Coding School

Reference

https://towardsdatascience.com/data-science-you-need-to-know-a-b-testing-f2f12aff619a

https://towardsdatascience.com/data-science-fundamentals-a-b-testing-cb371ceecc27

https://www.niagahoster.co.id/blog/ab-testing-adalah/

https://vwo.com/blog/ab-testing-examples/

https://www.scribbr.com/methodology/sampling-methods/