Python for Data Science Comprehensive Workshop

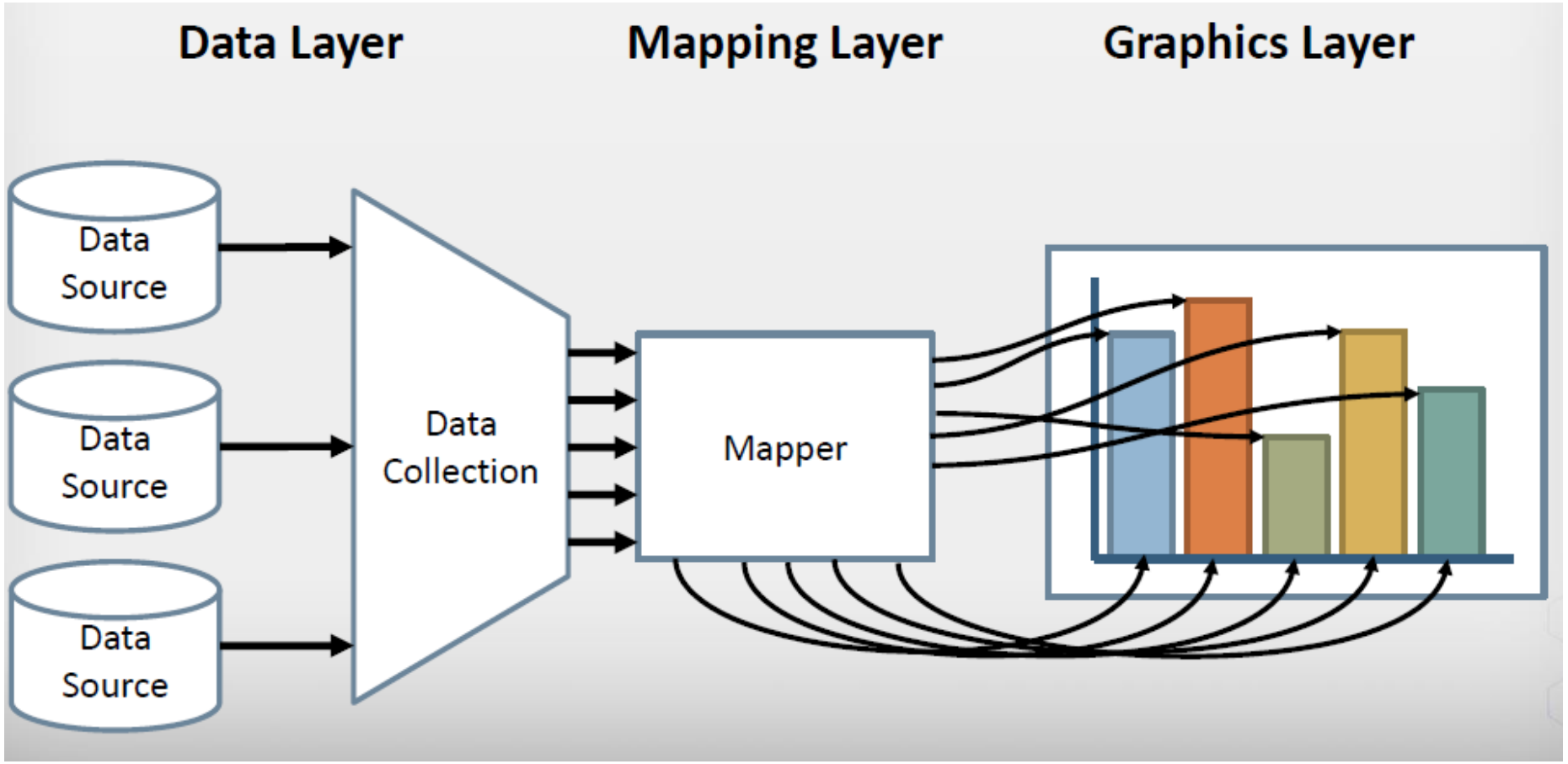
Part 03 – Data Visualization Using Matplotlib, Pandas & Seaborn

H.M. Samadhi Chathuranga Rathnayake

B.Sc(Hons).Special in Industrial Statistics (1st Class) (UOC),
B.Eng (Hons) in Software Engineering (LMU),
CLSSWB, Dip SE, Dip IT, Dip IT & E-Com, Dip B.Mgt, Dip HRM, Dip Eng

Data Visualization

Process of converting data into a graphical representation.



Data Visualization

Why Data Visualization ?

- Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs.
- This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets.

Data Visualization

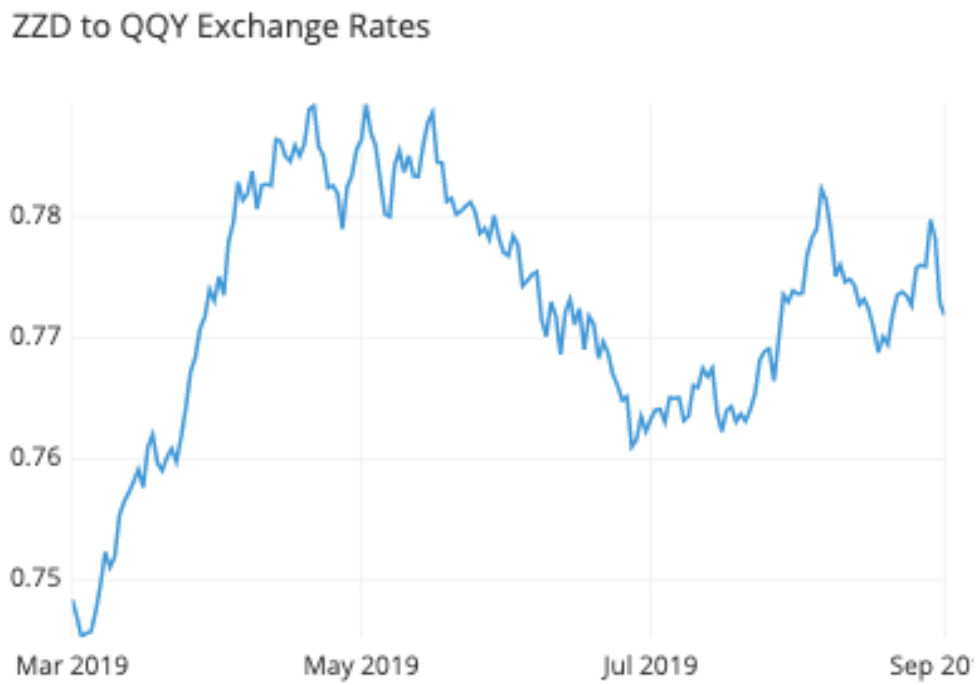
Type of variables is one of the main factors to be considered when visualizing

- Numerical variables (Ex- Age, Height, Weight)
- Categorical variables (Ex- Gender, Color, District, Country)

Data Visualization : Numerical Data

Line Charts:

A line chart or line plot or line graph or curve chart is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments.

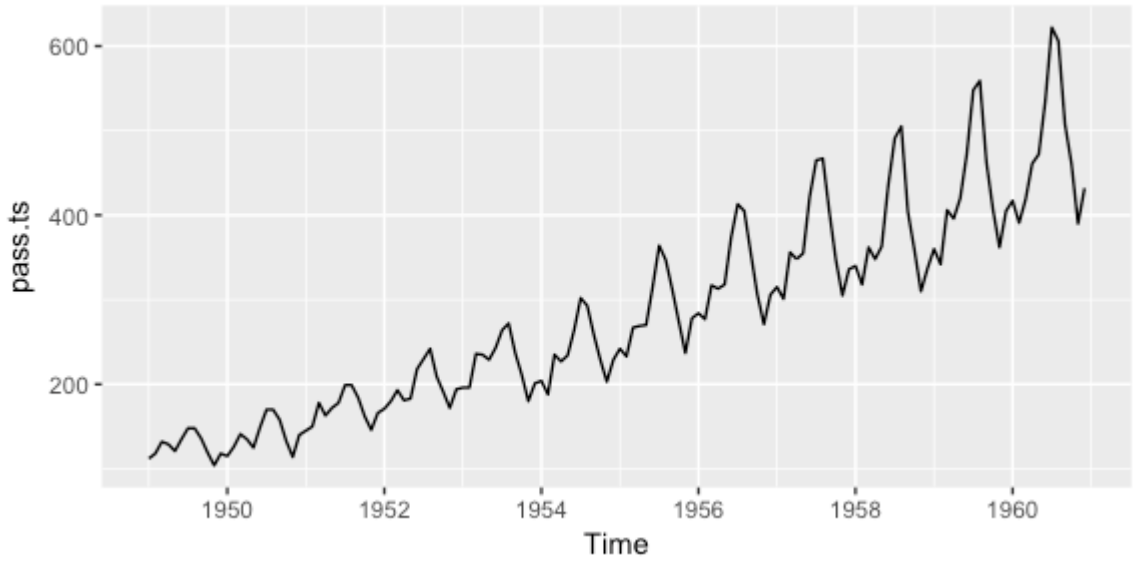


Data Visualization : Numerical Data

Line Charts:

Important patterns which can be observed

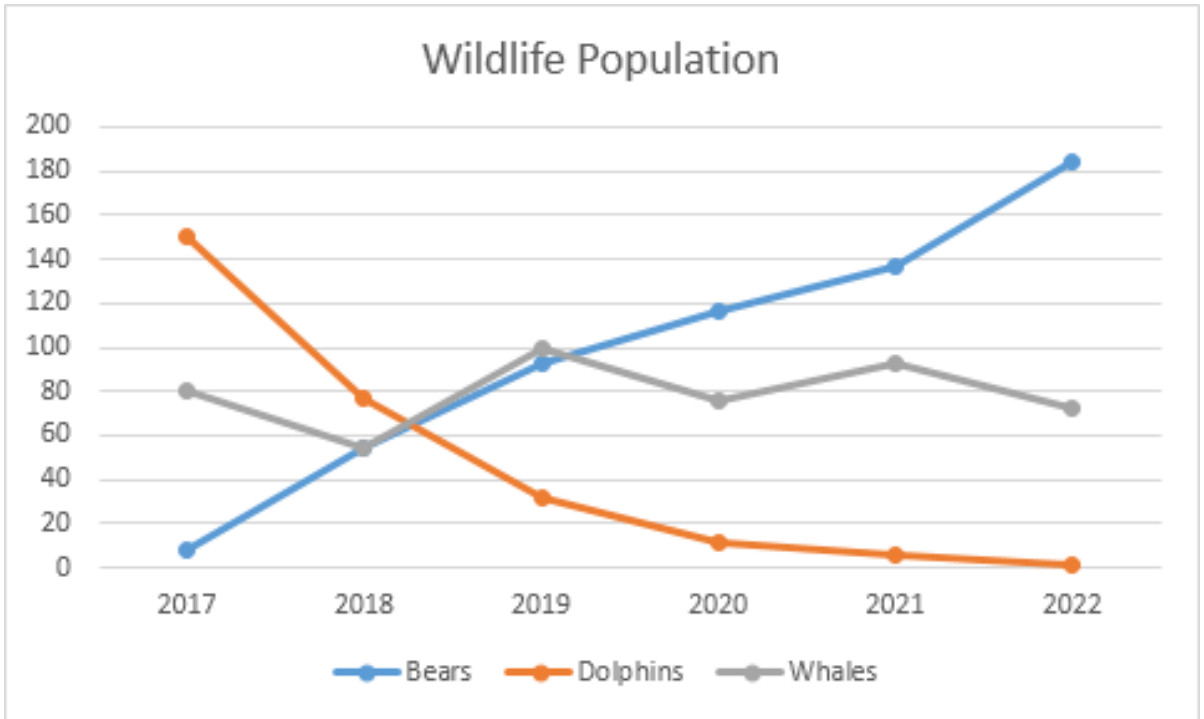
Seasonality



Data Visualization : Numerical Data

Line Charts:

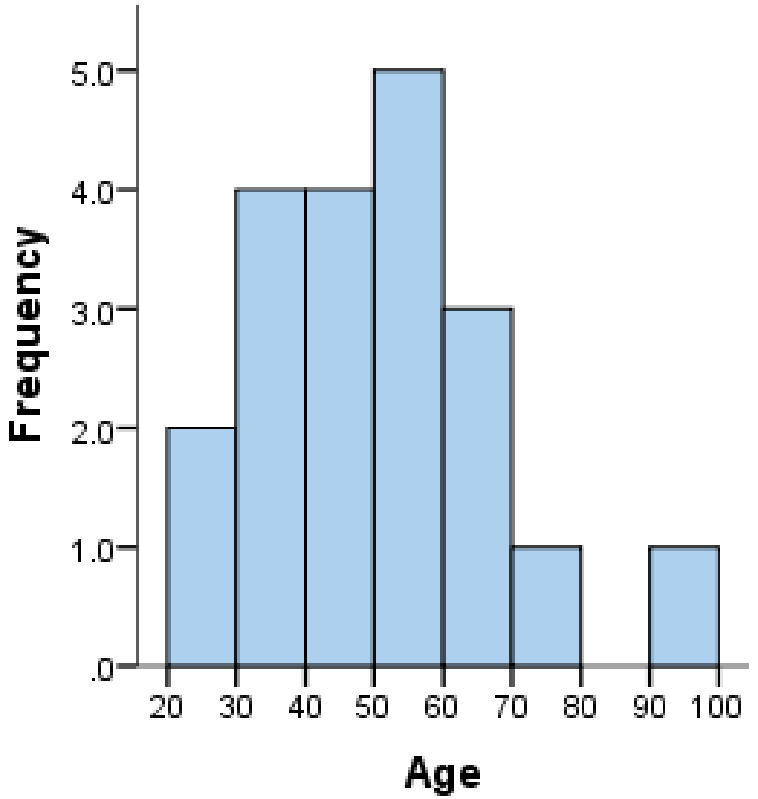
Line charts are important when comparing several charts.



Data Visualization : Numerical Data

Histograms:

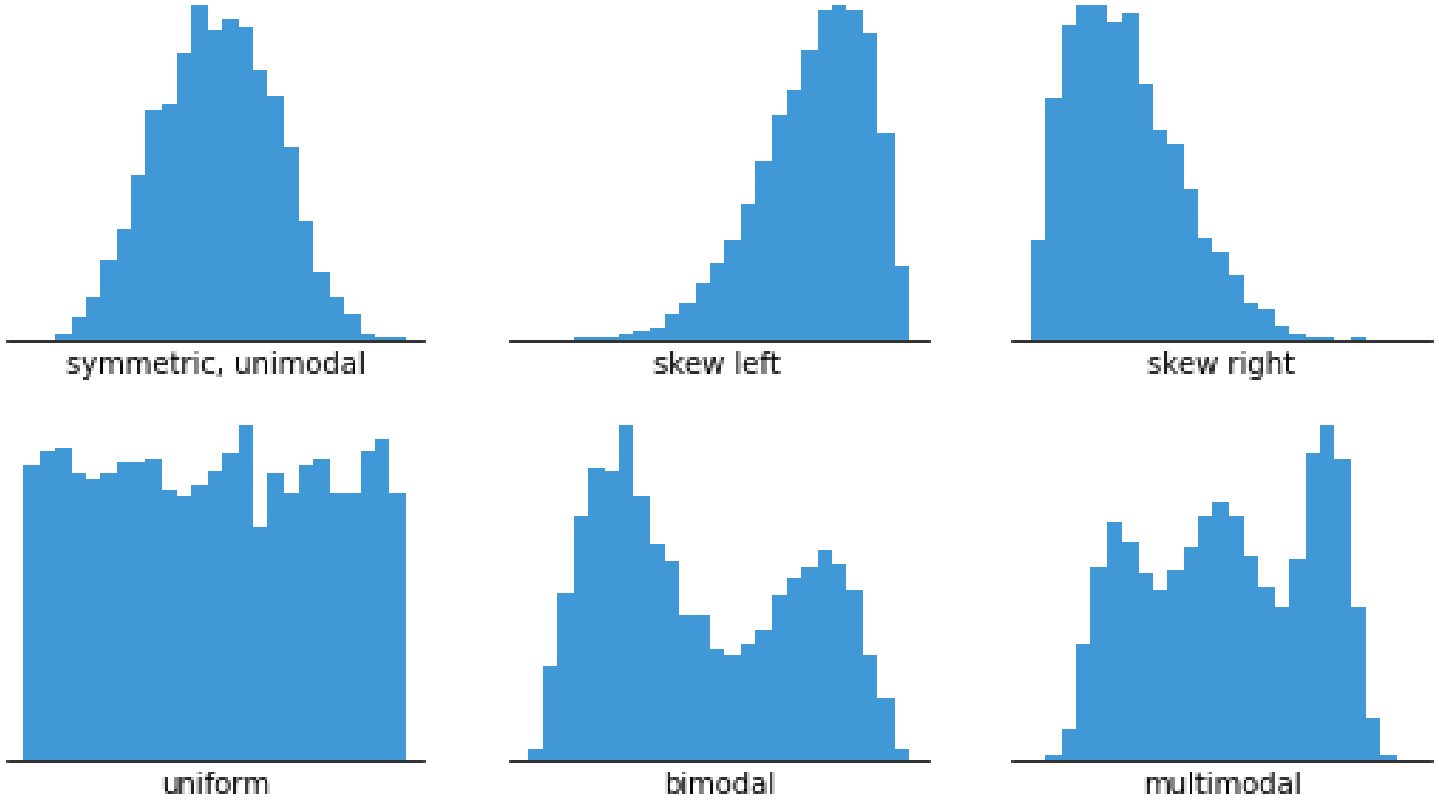
A histogram is an approximate representation of the distribution of numerical data.



Data Visualization : Numerical Data

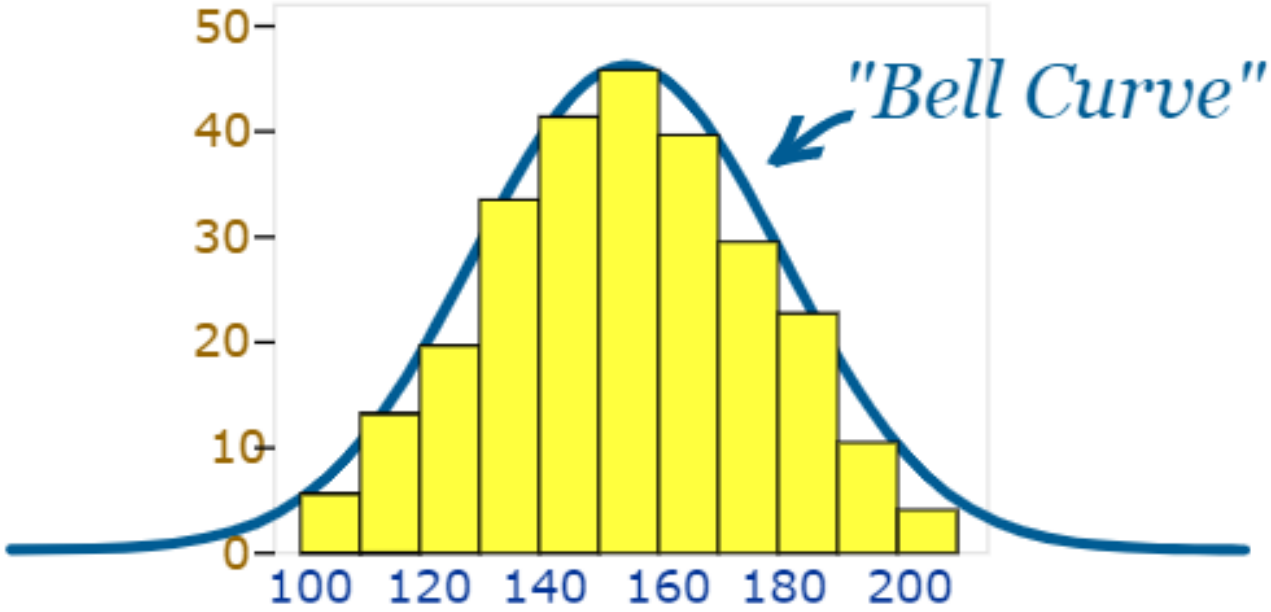
Histograms:

Different shapes can be observed when we draw histograms.

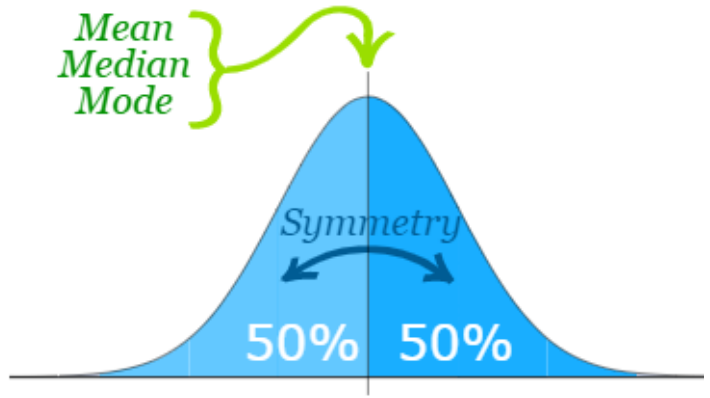


Data Visualization : Numerical Data

Histograms: Normal distribution



$$X \sim N(\mu, \sigma^2)$$

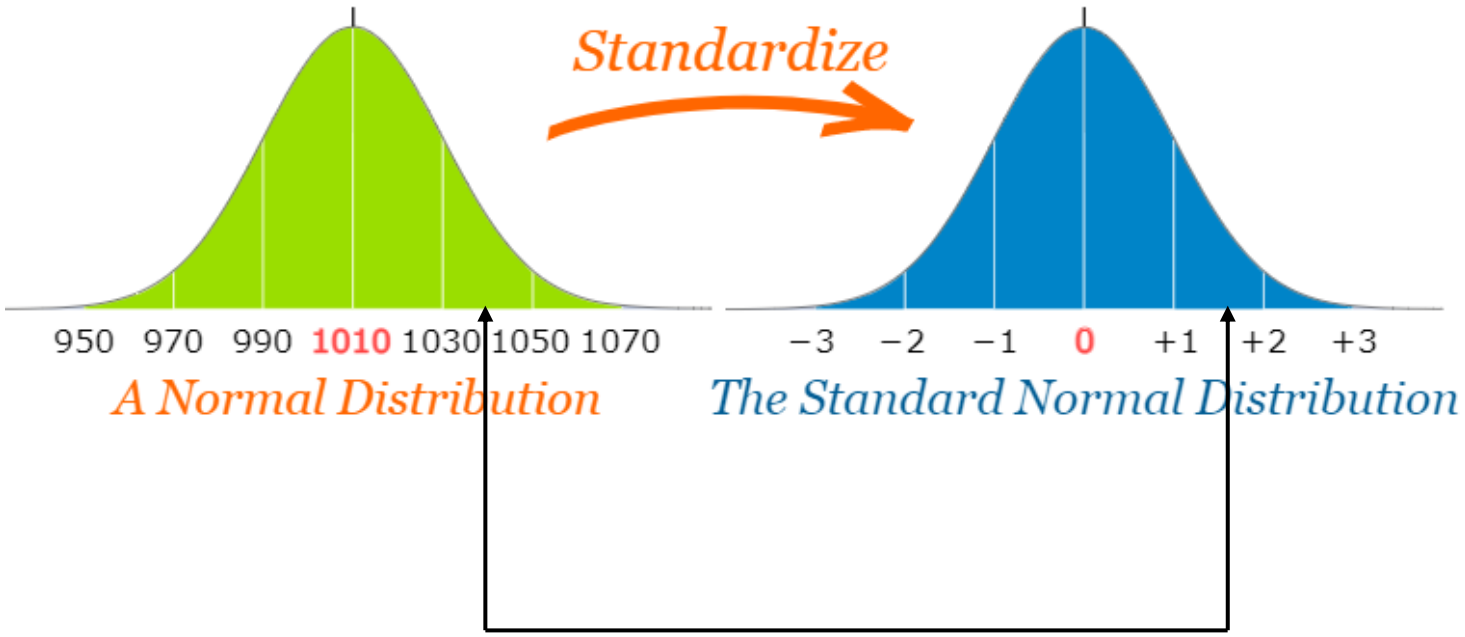


Important parameters are the **mean**$(\boldsymbol{\mu})$ and the **variance** $(\boldsymbol{\sigma^2})$**.**

Standard deviation $(\boldsymbol{\sigma})$ is the square root of variance.

Data Visualization : Numerical Data

Histograms: Normal distribution

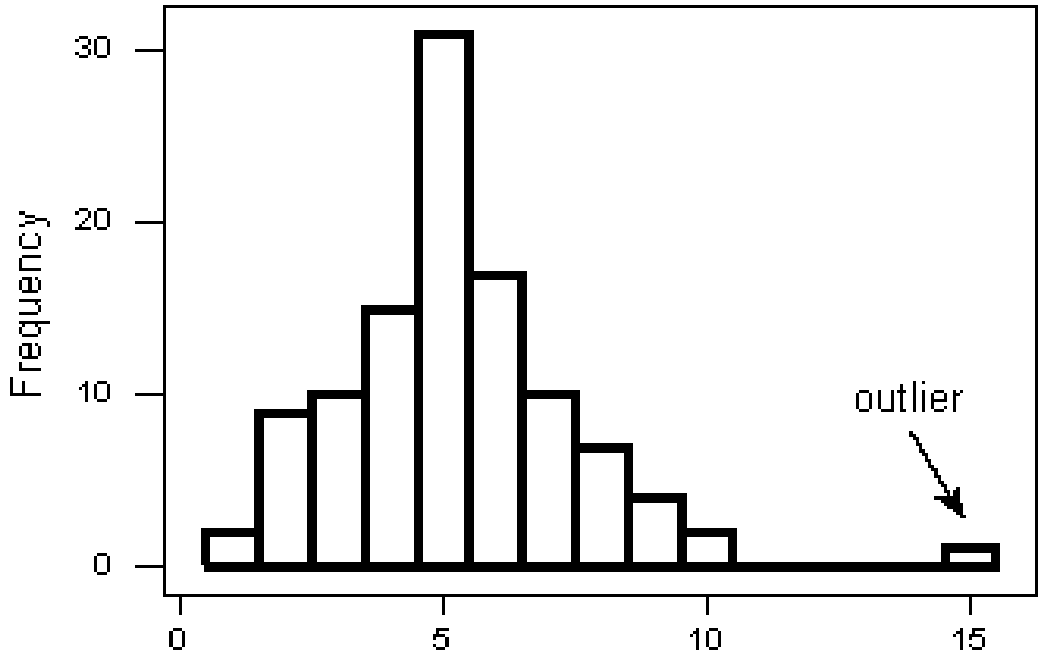


$$z = \frac{x - \mu}{\sigma}$$

Data Visualization : Numerical Data

Histograms:

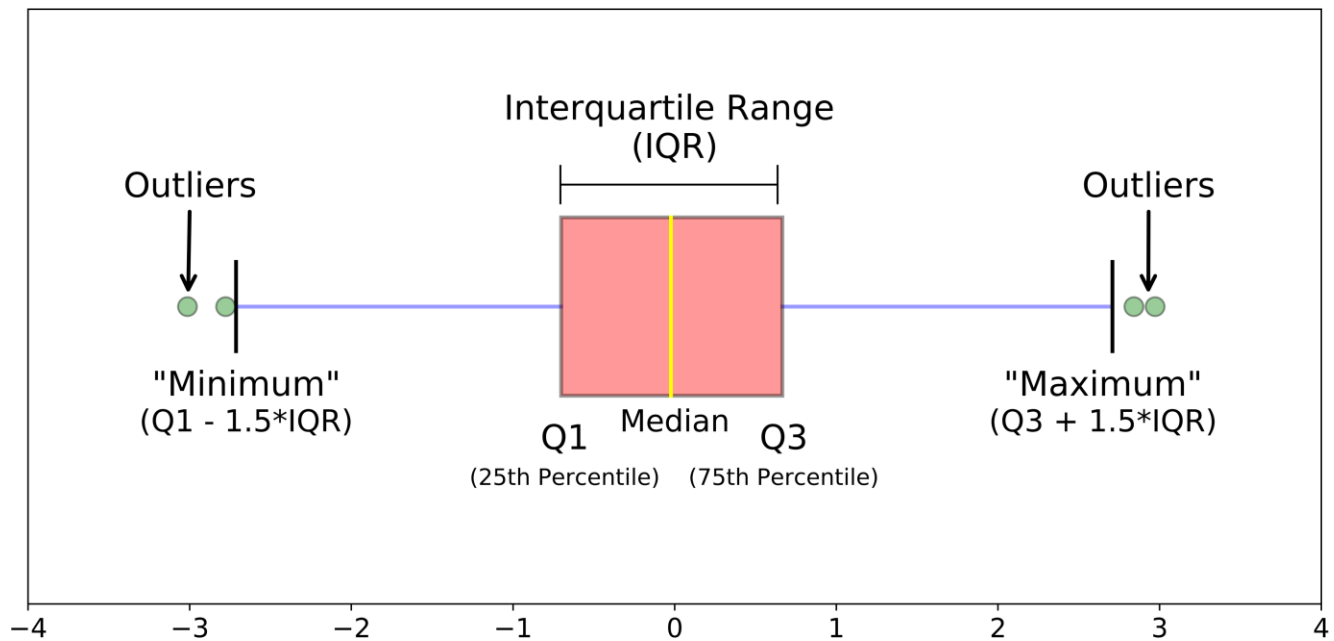
Histograms show the outliers as well.



Data Visualization : Numerical Data

Boxplots:

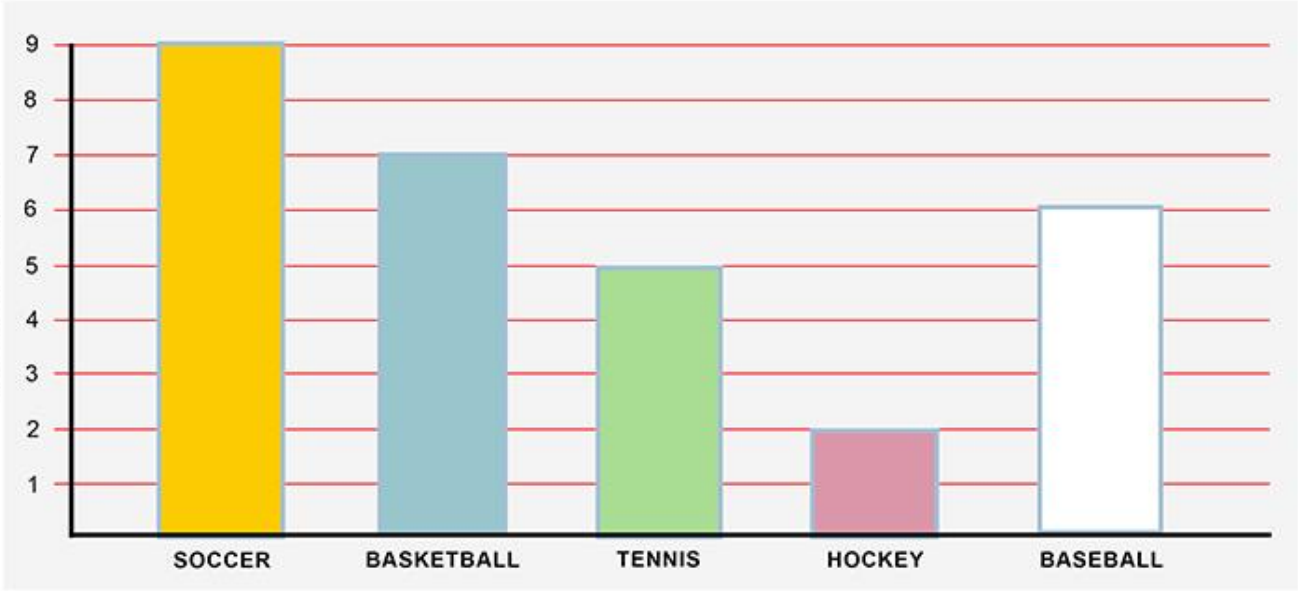
In descriptive statistics, a box plot or boxplot is a method for graphically depicting groups of numerical data through their quartiles. They are showing outliers as well.



Data Visualization : Categorical Data

Bar Graphs:

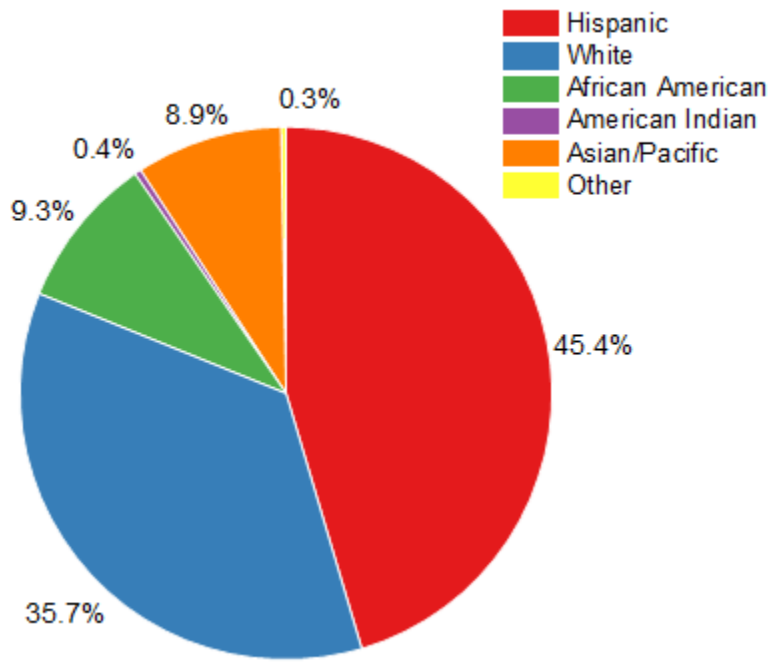
A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.



Data Visualization : Categorical Data

Pie Charts:

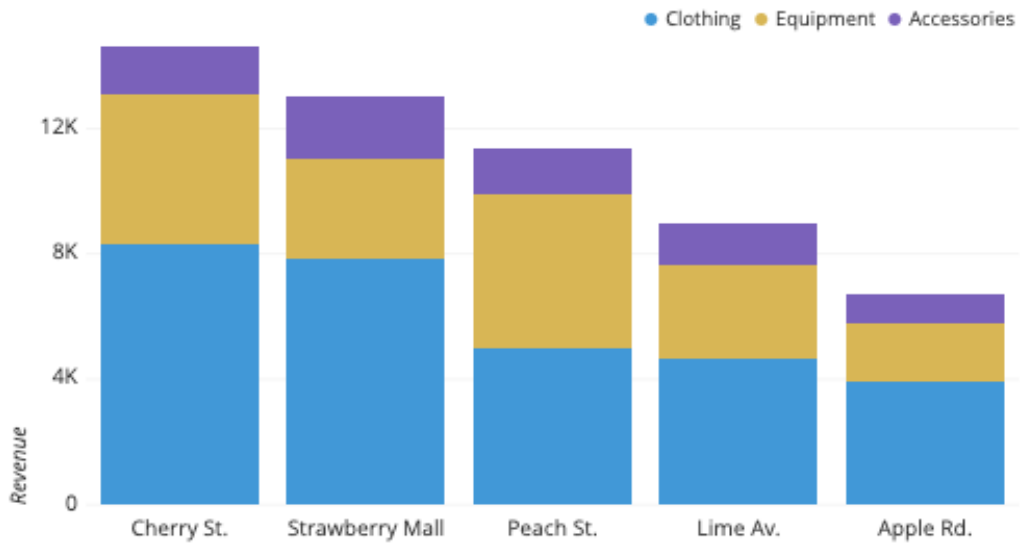
A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion.

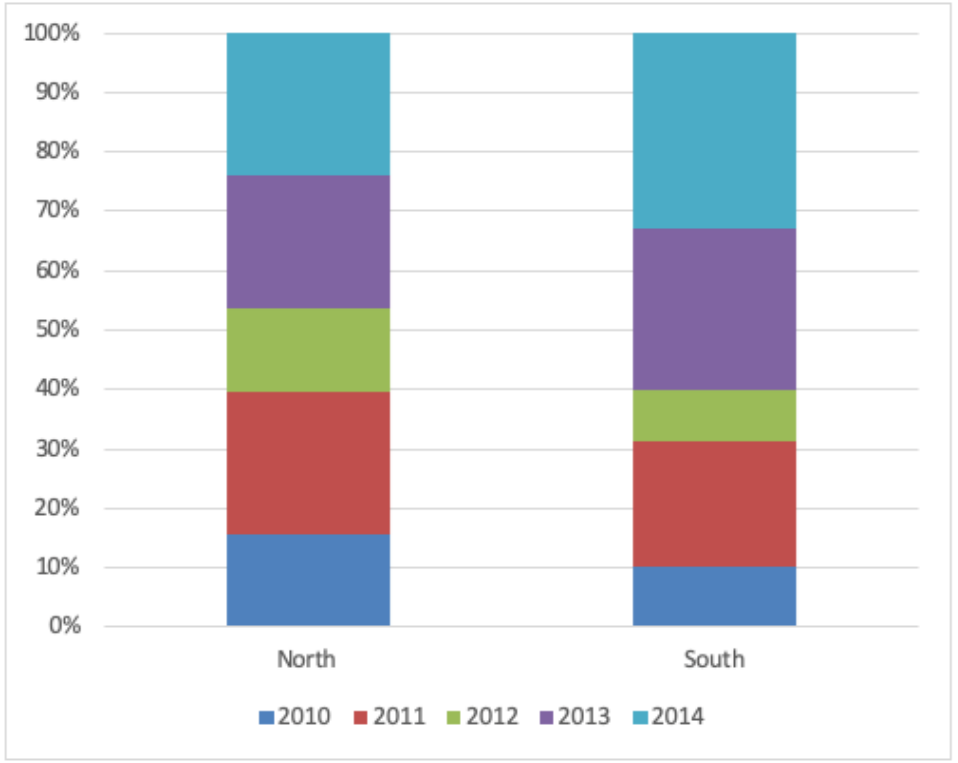


Data Visualization : Categorical Data VS Categorical Data

Stacked Bar Graphs:

In a stacked bar chart, parts of the data are adjacent or stacked. Each bar displays a total amount, broken down into sub-amounts.

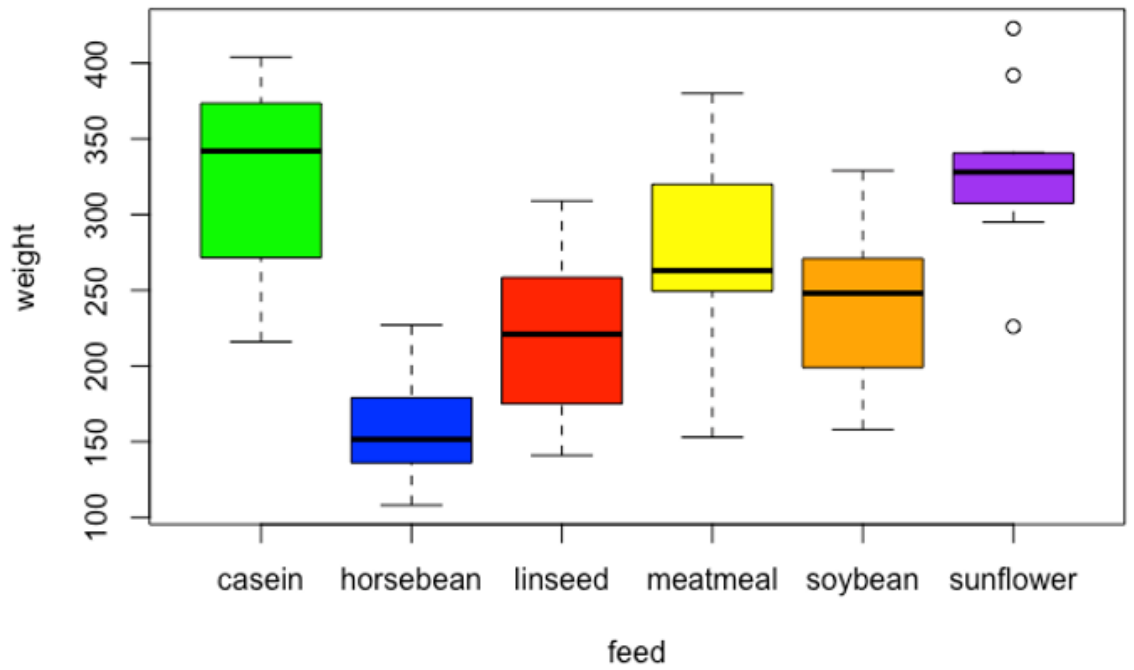




Data Visualization : Categorical Data VS Numerical Data

Side by Side Boxplots:

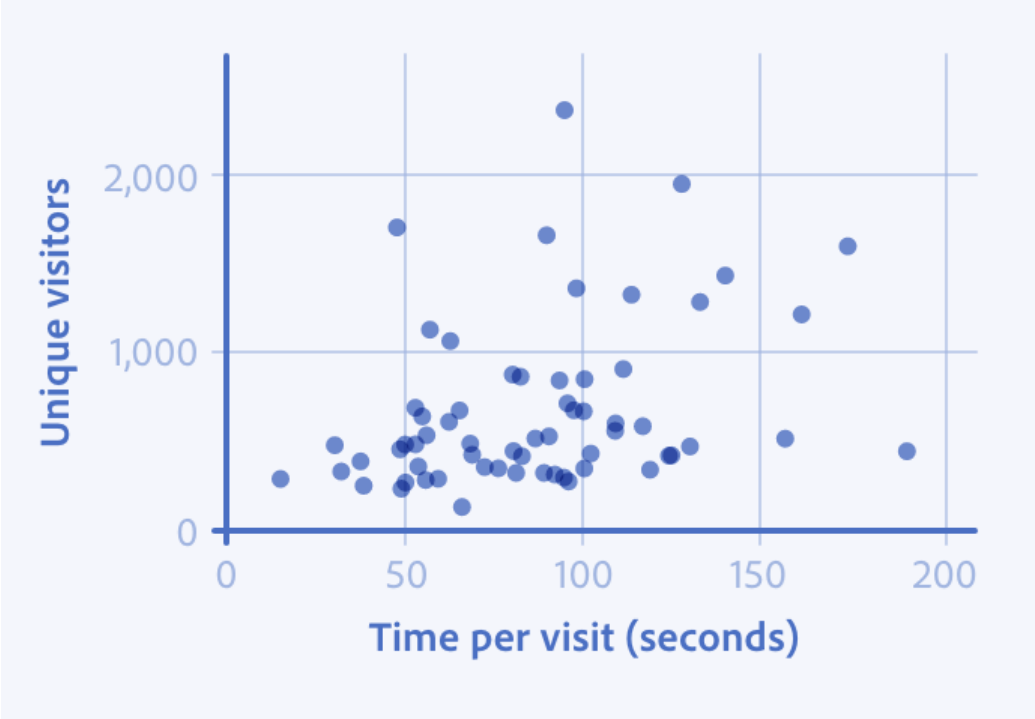
Side-By-Side boxplots are used to display the distribution of several quantitative variables or a single quantitative variable along with a categorical variable.



Data Visualization : Numerical Data VS Numerical Data

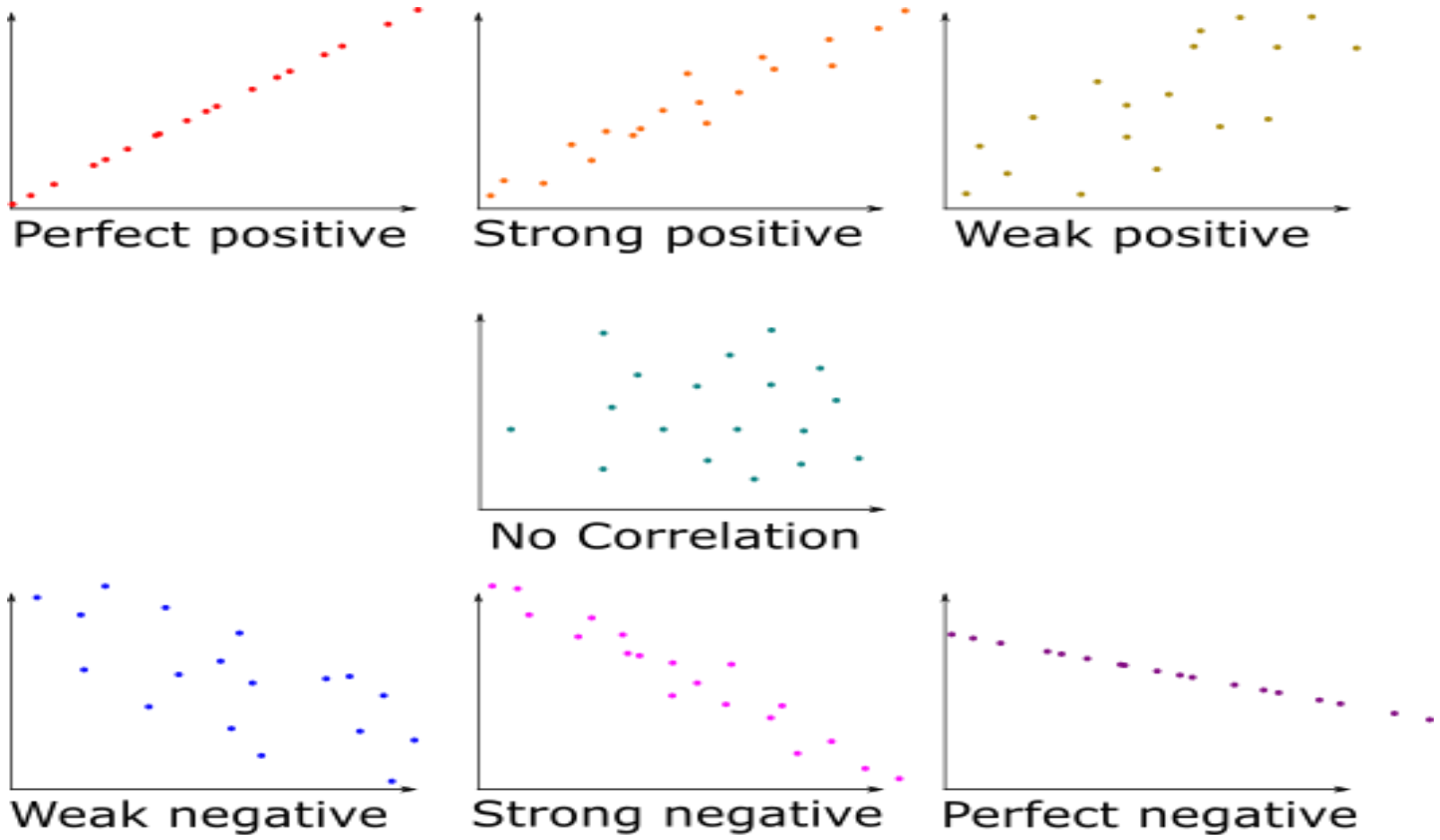
Scatter Plots:

A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data.



Data Visualization : Numerical Data VS Numerical Data

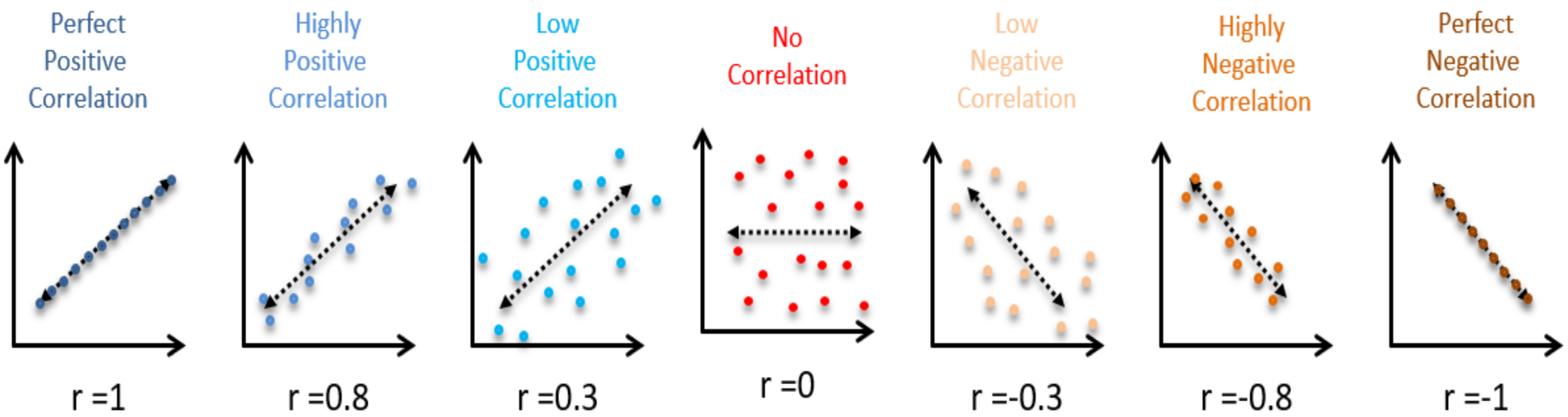
Scatter Plots: Describing scatter plots.



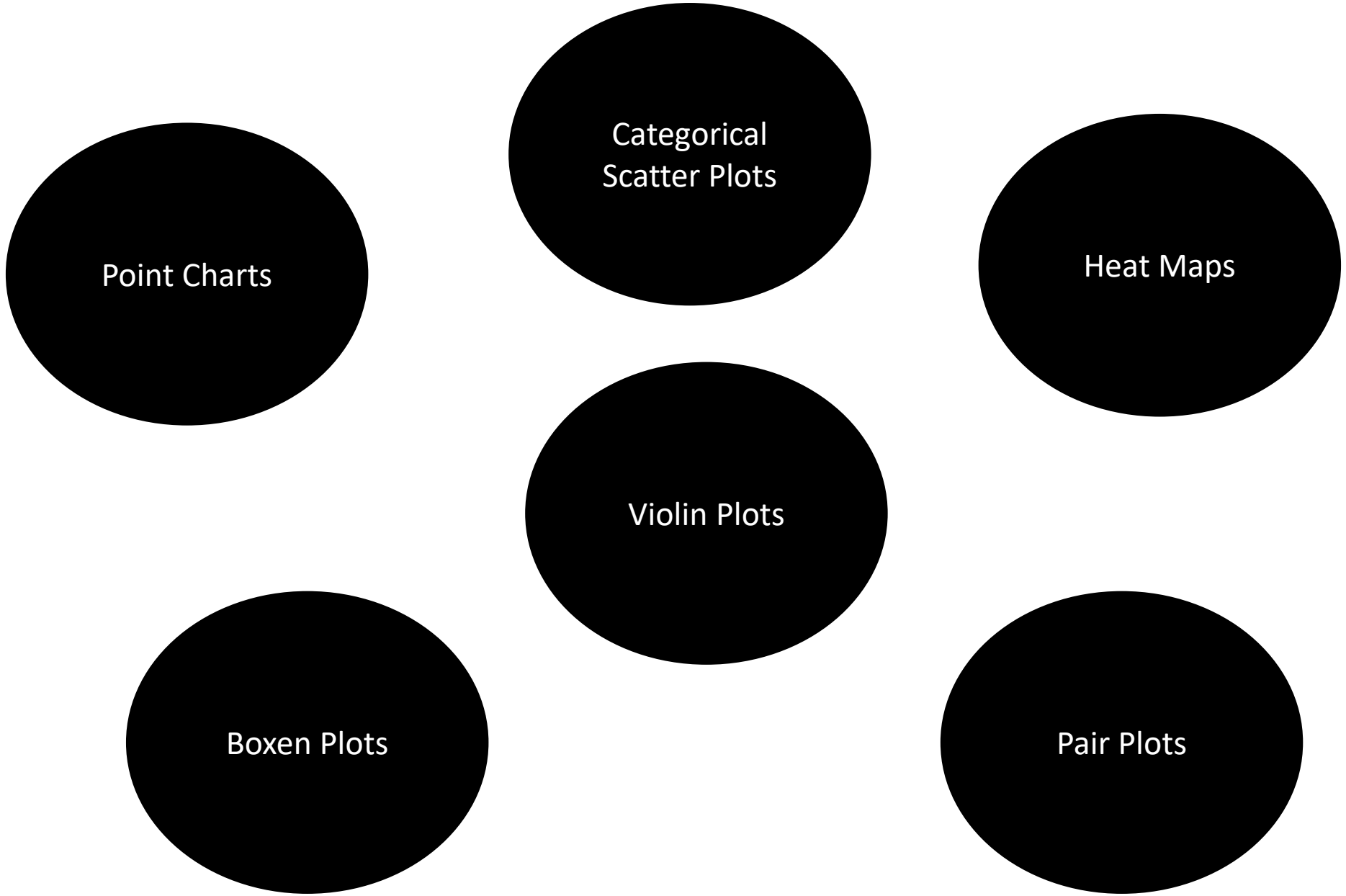
Data Visualization : Numerical Data VS Numerical Data

Scatter Plots: Correlation

Correlation coefficients are used to measure how strong a relationship is between two variables.



Data Visualization : Other Chart Types



Visualization Libraries in Python