

Data Visualization

- Data visualization allows us to quickly interpret the data and adjust different variables to see their effect
- Technology is increasingly making it easier for us to do so

Why visualize data?

- Observe the patterns
- Identify extreme values that could be anomalies
- Easy interpretation

Popular plotting libraries in Python

Python offers multiple graphing libraries that offers diverse features

- | | |
|-------------------------------|---|
| • matplotlib | • to create 2D graphs and plots |
| • pandas visualization | • easy to use interface, built on Matplotlib |
| • seaborn | • provides a high-level interface for drawing attractive and informative statistical graphics |
| • ggplot | • based on R's ggplot2, uses Grammar of Graphics |
| • plotly | • can create interactive plots |

Matplotlib

- Matplotlib is a 2D plotting library which produces good quality figures
- Although it has its origins in emulating the MATLAB graphics commands, it is independent of MATLAB
- It makes heavy use of NumPy and other extension code to provide good performance even for large arrays

Scatter Plot

What is a scatter plot?

- A scatter plot is a set of points that represents the values obtained for two different variables plotted on a horizontal and vertical axes

When to use scatter plots?

- Scatter plots are used to convey the relationship between two numerical variables
- Scatter plots are sometimes called correlation plots because they show how two variables are correlated

Histogram

What is a histogram?

- It is a graphical representation of data using bars of different heights
- Histogram groups numbers into ranges and the height of each bar depicts the frequency of each range or bin

When to use histograms?

- To represent the frequency distribution of numerical variables

Bar plot

What is a bar plot?

- A bar plot is a plot that presents categorical data with rectangular bars with lengths proportional to the counts that they represent

When to use bar plot?

- To represent the frequency distribution of categorical variables
- A bar diagram makes it easy to compare sets of data between different groups