Data Visualisation

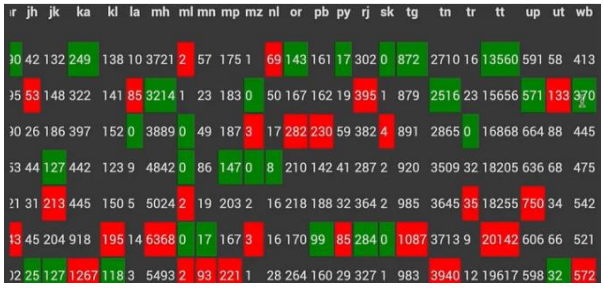
Reading Complex JSON Files:

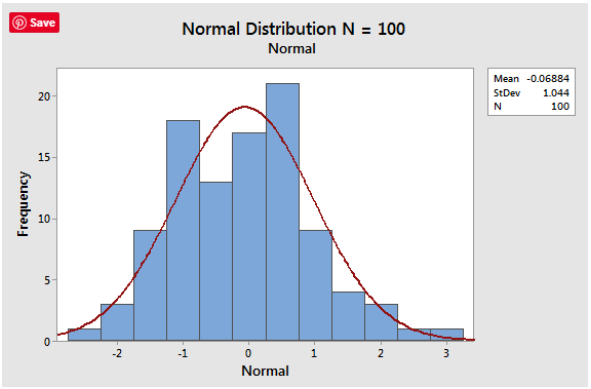
JSON (JavaScript Object Notation) is a popular data format for storing and exchanging data. Python provides libraries like json to parse and manipulate JSON data. Reading complex JSON files involves navigating through nested structures and extracting relevant information.

Styling Tabulation:

Tabulation refers to the presentation of data in a tabular format. Styling tabulation involves customizing the appearance of tables to enhance readability and aesthetics. Libraries like Pandas provide functions to style data tables in various ways, such as applying colors, formatting text, and adding visual elements.

https://prvnk10.medium.com/styling-tabulation-a5b4045e5189





Distribution of Data - Histogram:

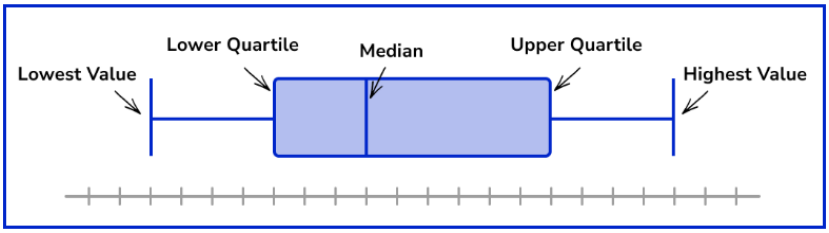
A histogram is a graphical representation of the distribution of numerical data. It divides the data into intervals (bins) and displays the frequency of values within each interval using bars. Histograms provide insights into the underlying distribution of data, including central tendency, variability, and skewness.

https://statisticsbyjim.com/basics/assessing-normality-histograms-probability-plots/

Box Plot:

A box plot, also known as a box-and-whisker plot, is a graphical summary of the distribution of a dataset. It displays key statistics such as the median, quartiles, and outliers. Box plots are useful for comparing distributions and identifying potential outliers or anomalies in the data.

https://thirdspacelearning.com/gcse-maths/statistics/box-plot/



Data Visualisation

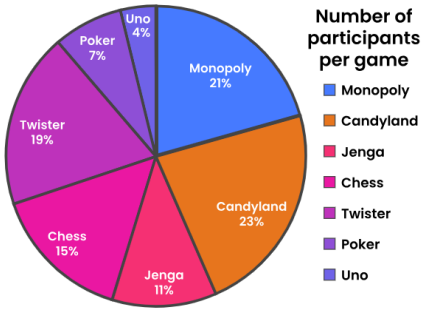
Data Visualization - Recap:

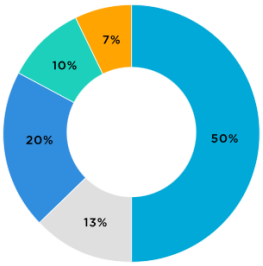
Recapitulating data visualization involves summarizing the key concepts, techniques, and tools covered in the module. It reinforces learning and helps learners consolidate their understanding of data visualization principles and practices.

Pie Chart:

A pie chart is a circular statistical graphic divided into slices to illustrate numerical proportions. The arc length of each slice is proportional to the quantity it represents. Pie charts are useful for displaying the relative proportions of categorical data.

https://www.houseofmath.com/encyclopedia/statistics-and-probability/statistics/charts/what-does-pie-chart-mean





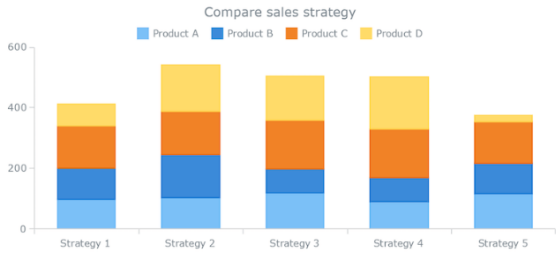
Donut Chart:

A donut chart is similar to a pie chart but with a hole in the center. It is created by removing the center part of a pie chart. Donut charts are visually appealing and can be useful for displaying the same information as a pie chart while allowing additional space for annotations or labels.

https://www.spotfire.com/glossary/what-is-a-donut-chart

Stacked Bar Plot:

A stacked bar plot is used to represent categorical data with multiple variables. Each bar in the plot represents a category, and the length of the bar is divided into segments corresponding to different subcategories. Stacked bar plots are effective for comparing the total values across categories while showing the contribution of each subcategory.



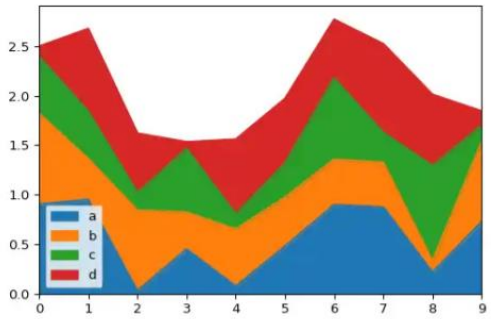
Relative Stacked Bar Plot:

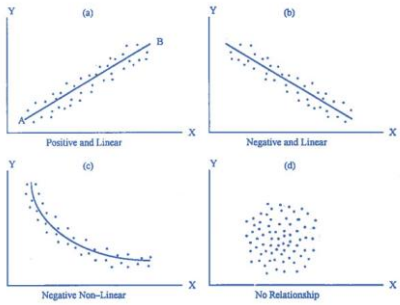
A relative stacked bar plot, also known as a normalized stacked bar plot, represents the proportion of each subcategory within a category. Instead of absolute values, the height of each segment in the stacked bar plot represents the relative frequency or proportion of the corresponding subcategory within the category.

Data Visualisation

Stacked Area Plot:

A stacked area plot is similar to a stacked bar plot but with the data represented as areas instead of bars. It is useful for visualizing changes in the composition of data over time or across different categories. Stacked area plots highlight the cumulative contribution of different variables to the total value.



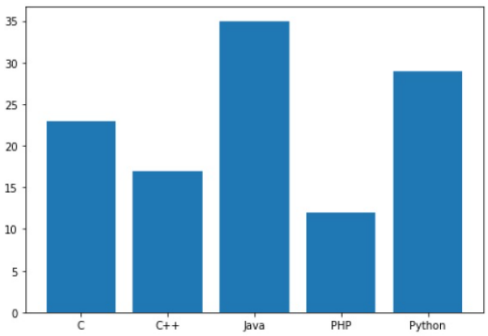


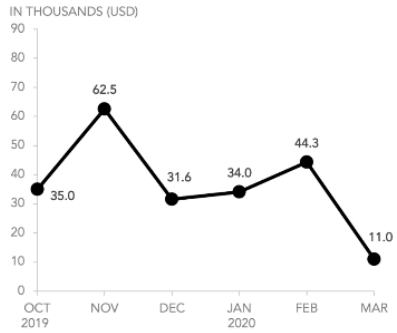
Scatter Plots:

A scatter plot is a two-dimensional plot that displays the values of two variables as points on a Cartesian plane. Each point represents an observation in the dataset, with the x-coordinate corresponding to one variable and the y-coordinate corresponding to the other variable. Scatter plots are useful for visualizing relationships and patterns between variables.

Bar Plot:

A bar plot is a graphical representation of categorical data with rectangular bars. The length or height of each bar corresponds to the frequency, count, or proportion of the corresponding category. Bar plots are effective for comparing the values of different categories.





Continuous vs Continuous Plot:

Continuous vs continuous plots, such as scatter plots or line plots, are used to visualize the relationship between two continuous variables. These plots help identify patterns, trends, correlations, and outliers in the data.

Line Plot:

A line plot, also known as a line graph, is used to display data points connected by straight lines. Line plots are commonly used to visualize trends or patterns in time-series data, where the x-axis represents time and the y-axis represents the variable of interest.

Data Visualisation

Line Plot Covid Data:

A line plot of COVID-19 data typically represents the trend of cases, deaths, or other relevant metrics over time. Each point on the plot represents the value of the metric at a specific date, and the line connects these points to show the overall trend. Line plots of COVID-19 data are essential for tracking the progression of the pandemic and informing public health decisions.

Category and Continuous: Bar Plot, Box Plot, Violin Plot, Swarm Plot, Point Plot

Continuous and Continuous: Scatter Plot, Line Plot, Heatmap, Hexbin Plot, Contour Plot

Category and Category: Stacked Bar Plot, Grouped Bar Plot, Clustered Column Chart, Waterfall Chart, Diverging Stacked Bar Chart