**Assignment No:-5**

**Problem Statement:-**

Visualize the data using R/Python by plotting the graphs for assignment no. 1 and 2. Consider a suitable data set. a) Use Scatter plot, bar plot, Box plot and Histogram

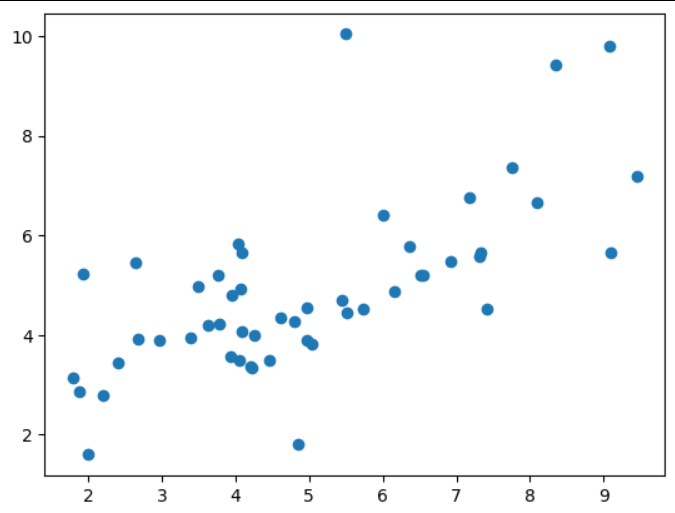
**Theory**:- **Methodology:**

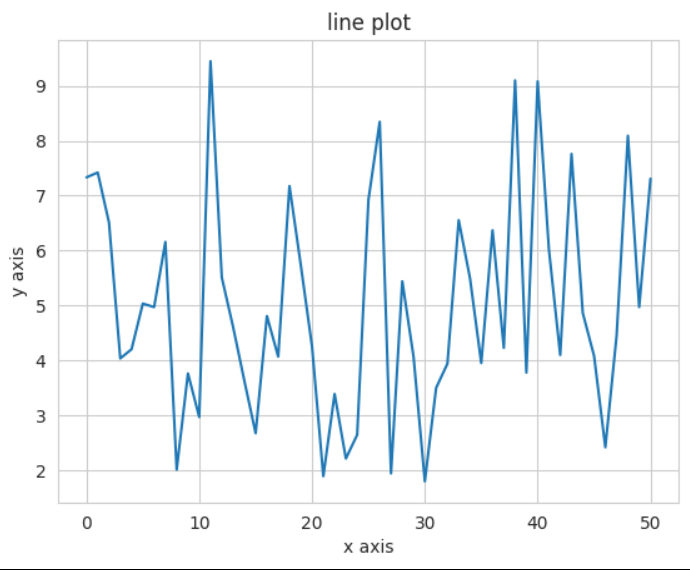
1. Selecting a Suitable Dataset:
   * Choose a dataset that contains numerical and categorical variables suitable for visualization with scatter plot, bar plot, box plot, and histogram.
2. Data Preprocessing:
   * Clean the data by handling missing values, outliers, and formatting issues.
3. Visualization:
   * Scatter Plot: Use numerical variables to plot points on a graph, providing insights into the relationship between variables.
   * Bar Plot: Utilize categorical variables to represent data with rectangular bars, illustrating comparisons between categories.
   * Box Plot: Display the distribution of numerical data through quartiles, identifying outliers and the range of the data.
   * Histogram: Illustrate the frequency distribution of numerical data by dividing it into bins and plotting the number of occurrences in each bin.
4. Interpretation:
   * Analyze the visualizations to identify patterns, trends, and outliers in the data.
   * Use visualization techniques to communicate insights effectively to stakeholders.

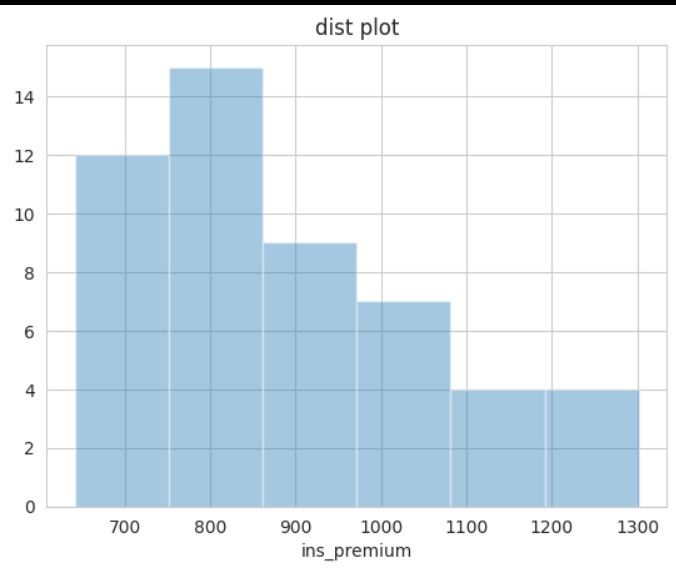
Advantages and Disadvantages & Limitation/Example:

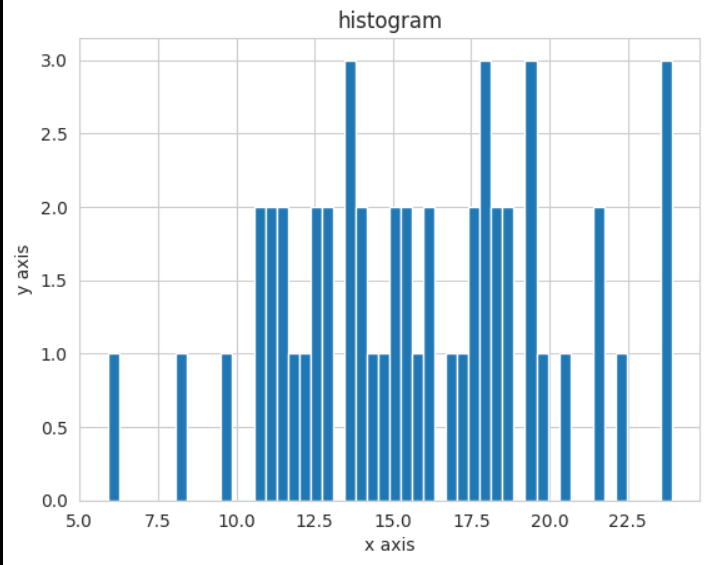
1. Advantages:
   * Scalability: Visualizations can be scaled to accommodate large datasets.
   * Interpretability: Visual representations make it easier to understand complex data patterns.
   * Communication: Visualizations facilitate communication of findings to stakeholders and decision-makers.
   * Insight Generation: Visualizations help in discovering trends, relationships, and outliers in the data.
2. Disadvantages & Limitations/Example:
   * Subjectivity: Interpretation of visualizations can be subjective and influenced by the viewer's biases.
   * Complexity: Creating effective visualizations requires knowledge of visualization techniques and tools.
   * Misleading Interpretation: Poorly designed visualizations can lead to misinterpretation of data.
   * Limited by Data Quality: Visualizations are only as good as the quality of the underlying data.

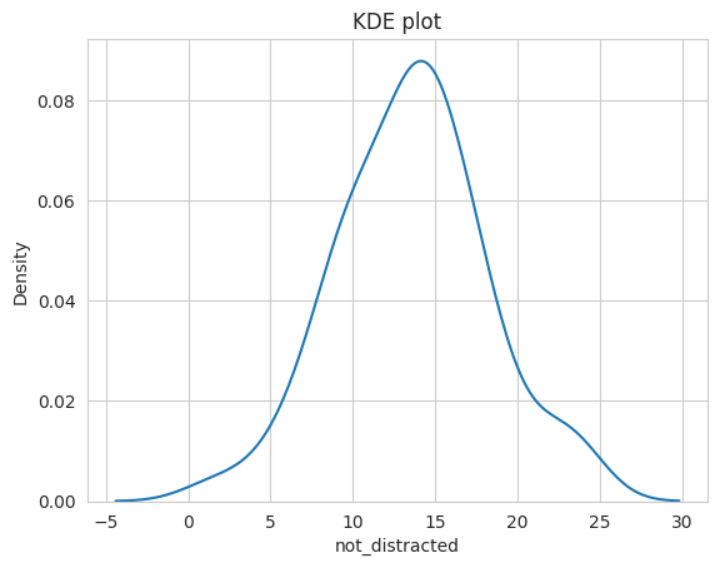
**Diagram**

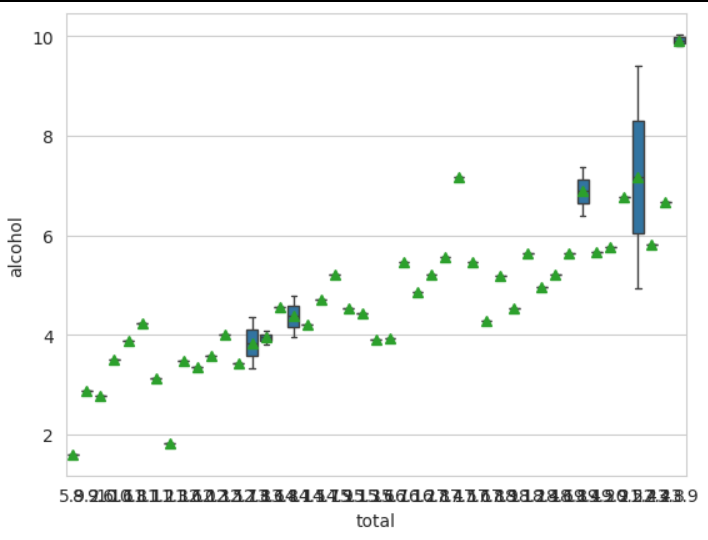
****

****

****

****

****

**Conclusion**

In conclusion, utilizing scatter plots, bar plots, box plots, and histograms allows for comprehensive data visualization. These visualizations offer insights into relationships between variables, comparisons among categories, distributions of data, and identification of outliers. By leveraging these graphical representations, stakeholders gain a deeper understanding of the dataset's characteristics and can make informed decisions based on the observed patterns and trends.