**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**:-

1. **Scatter Plot**:
   * Scatter plots are used to visualize the relationship between two continuous variables. Each point on the plot represents a single observation, with one variable on the x-axis and the other on the y-axis.
   * Scatter plots are helpful in identifying patterns, trends, and relationships between variables. For example, they can show if there's a correlation between two variables, if there's a linear or non-linear relationship, or if there are any outliers.
2. **Bar Plot**:
   * Bar plots are useful for visualizing and comparing the values of categorical variables or the summary statistics of continuous variables.
   * In a bar plot, each bar represents a category, and the height of the bar corresponds to the value of that category.
   * Bar plots are commonly used to show frequencies, counts, averages, or other summary statistics for different categories. They are particularly effective in comparing values across categories.
3. **Box Plot**:
   * Box plots (or box-and-whisker plots) are graphical representations of the distribution of a dataset.
   * They display the median, quartiles, and outliers of the data distribution. The box represents the interquartile range (IQR), which contains the middle 50% of the data. The whiskers extend to the minimum and maximum values within a certain range (often 1.5 times the IQR), and points beyond the whiskers are considered outliers.
   * Box plots are useful for detecting outliers, comparing the spread and skewness of different datasets, and identifying central tendencies.
4. **Histogram**:
   * Histograms are graphical representations of the distribution of a single continuous variable.
   * They display the frequency or count of observations within predefined intervals or bins along the x-axis, with the height of each bar indicating the number of observations falling within that bin.
   * Histograms are useful for understanding the shape, central tendency, spread, and skewness of a dataset's distribution. They can reveal patterns such as normal distributions, skewness, multimodality, or outliers.

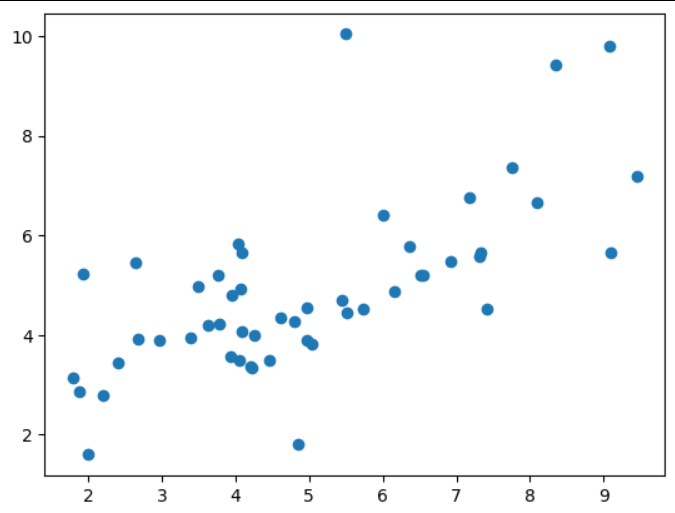
Advantages and Disadvantages & Limitation/Example:

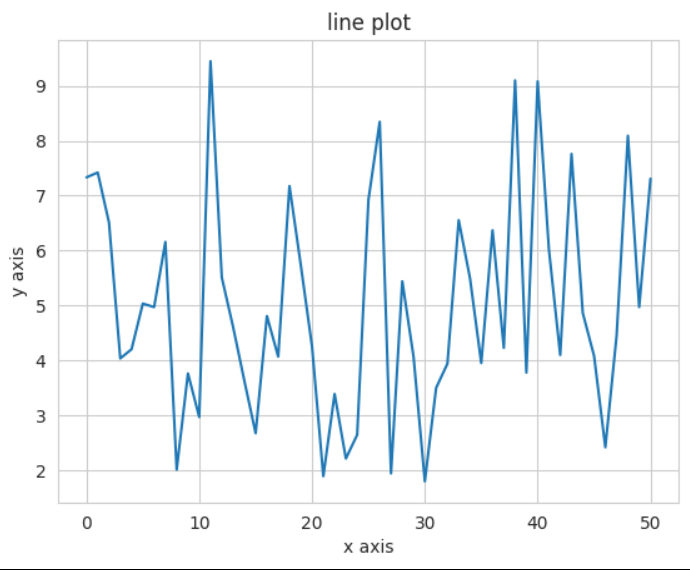
1. Advantages:
   * Scalability: Visualizations can be scaled to accommodate large datasets.
   * Interpretability: Visual representations make it easier to understand complex data patterns.
   * Communication: Visualizations facilitate communication of findings to stakeholders and decision-makers.
   * 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.

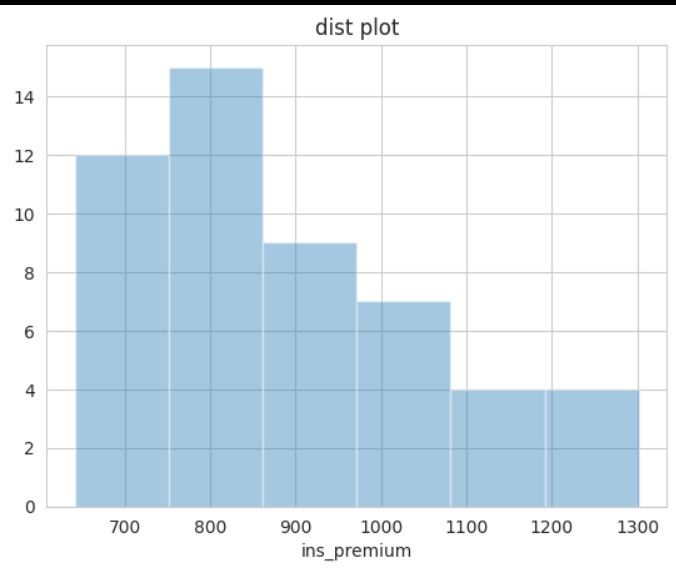
**Methodology:**

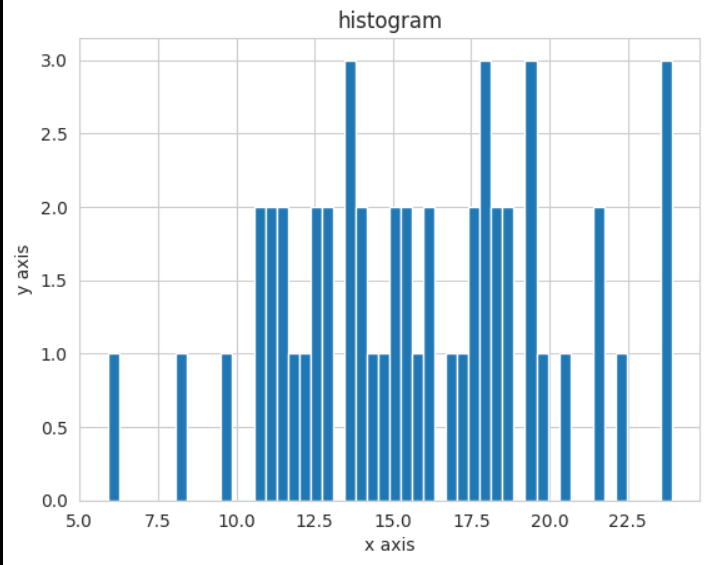
1. Selecting a Suitable Dataset:
   * Choose a dataset that contains numerical and categorical variables suitable for visualization with scatter plot, bar plot, box plot, and histogram.
2. Data Preprocessing:
   * Clean the data by handling missing values, outliers, and formatting issues.
3. Visualization:
   * 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.

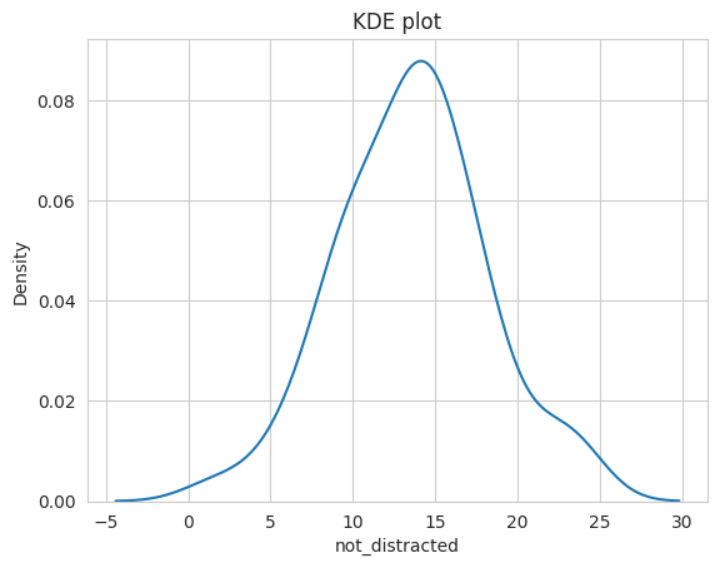
**Diagram**

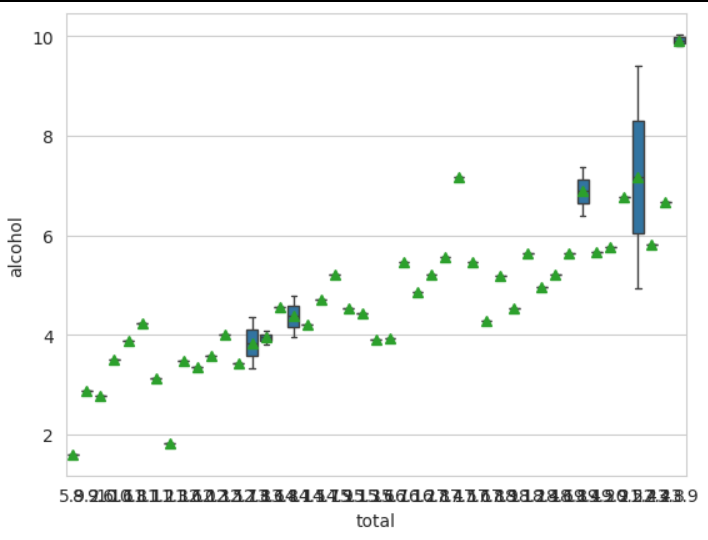
****

****

****

****

****

**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.