The Maharaja Sayajirao University of Baroda
Department of Statistics, Faculty of Science
M. Sc. Previous(Sem-1)
Data Visualization
Assignment-2

Date: 06-08-2025

Note : Ensure every graph includes a proper title and axis label.

Q.1 Consider the diamonds dataset in ggplot package in R. This dataset contains contain the prices and other attributes of almost 54,000 of diamonds.

1. Design a visualization to explore the relationship between diamond carat and price. Your plot should minimize overlapping of data points to enhance clarity. Apply a clean, minimal visual style, and include an appropriate title along with axis label.
2. Create a plot that display the relationship between diamond carat and depth based on their clarity using colour and include an appropriate title along with axis label.
3. The plot divide in subplot based on the various levels of clarity, so each subplot represents one clarity level.
4. Create a plot that shows the relationship between diamond carat and price. Use colour fill to differentiate between diamond cut types. Also, include a smooth curve to represent the overall trend.
5. Create a visualization that compares diamond prices across distinct colour grades. Show the outlier with the distinct colour. Use a vibrant colour palette to distinguish between the colour categories.
6. Create a bar chart to compare the frequency of each cut across different clarity levels. Group the bars by clarity using position adjustments that shows multiple bar graph and use colour for different clarity levels.
7. Create a horizontal boxplot to visualize the distribution of diamond prices across different colour grades. Reverse the price axis and customize the axis label and break the axis.
8. Construct a scatter plot where the transparency of each point reflects the carat value and enhance the visual clarity by colour scale for the price variable.
9. Create a circular bar plot to visualize the total diamond price for each colour grade.

Q.2 Simulate data from different probability distributions and use ggplot2 to create visual representations. Interpret the shapes and properties of each distribution through graphical analysis.

1. Generate 500 random variables from a Uniform(0, 1) distribution. Visualize the distribution using a histogram with an overlaid density curve. Customize the bins and add appropriate titles and axis labels.
2. Compare bar plots of Binomial($n = 10$, $p = 0.5$) and Binomial($n = 20$, $p = 0.5$) in one panel. Use facets to display differences in spread.
3. Simulate 1000 samples each from Uniform(0,1), Normal(0,1), and Exponential(1) distributions. Plot all three densities in a single graph using colour coding and legends.
4. Create side-by-side boxplots for Normal($\mu = 0$, $\sigma = 1$), Normal($\mu = 5$, $\sigma = 2$), and Exponential($\lambda = 1$) samples. What differences do you observe in spread and shape?
5. Simulate 1000 values from a Gamma distribution with shape = 2, rate = 1. Plot a histogram and overlay a smoothed density curve.
6. Create a simulation to compare the Poisson distribution for $\lambda = 2$, 5, and 10. Plot them side by side using facets or line plots.