Homework 7: Plotting univariate and bivariate data

Problem 1: Custom Histograms

In this problem, you will work with the `diamonds` dataset from the **ggplot2** package. First, subset the data to include only diamonds with an **Ideal** cut using base R logical subsetting. Then complete the following tasks:

1. **ggplot2 Histogram**:

   - Create a histogram of the variable `price` using a binwidth of 500.

   - Customize the plot by setting black borders and a darkorange fill.

   - Add appropriate x-axis and y-axis labels and a title.

```
# Load necessary library for the diamonds dataset and ggplot2 functions
library(ggplot2)

# Write code here
```

2. **Base R Histogram**:

   - Using the same subset of data, produce a histogram with 10 breaks.

   - Customize it by adding an x-axis label, y-axis label, and main title.

```
# Write code here
```

Problem 2: Density Plots and Normality Assessment

Using the same `diamonds_ideal` dataset, complete the following tasks:

1. **Density Plot with ggplot2**:

   - Create a density plot for `price` with an adjust value of 0.9.

   - Customize the plot with a darkorange fill at 50% opacity and limit the x-axis to the range 300 to 8000.

2. **Overlay Normal Density**:

   - On the same plot, overlay a Normal probability density function using `stat_function()`.

   - Use the sample mean and standard deviation for the Normal curve and set its color to royalblue.

3. **Q-Q Plot**:

   - Produce a Q-Q plot using ggplot2 (with `geom_qq()` and `geom_qq_line()`) to assess the normality of `price`.

4. **Normality Tests**:

   - Perform a Shapiro-Wilk test and a Kolmogorov-Smirnov test for normality on `price` and print the results.

Problem 3: Scatterplots and Smoothing Curves

Investigate the relationship between `carat` and `price` using the full `diamonds` dataset.

1. **ggplot2 Scatterplot**:

   - Create a scatterplot with `carat` on the x-axis and `price` on the y-axis.

   - Customize the points (set size and color) and add a smoothing curve using `geom_smooth()` with a span of 1.0.

2. **Base R Scatterplot**:

- Create a scatterplot of the same variables using Base R with a customized point shape and labels.

Problem 4: Overlaid Density and Bar Plots for Categorical Data

Using the full `diamonds` dataset, complete the following tasks:

1. **Overlaid Density Plot**:

- Create an overlaid density plot of `price` by `cut` using ggplot2.

- Use custom fill colors for each level of `cut` (e.g., darkorange, darkorchid, cyan4, gold, steelblue) and adjust the transparency so the curves can be compared.

2. **Bar Plots for Categorical Data**:

- Create a bar plot of the variable `color` to show the frequency of diamonds for each color using ggplot2.

- Then modify the bar plot to display the relative frequencies (proportions).

3. **Pareto Chart**:

- Produce a Pareto chart for `color` by sorting the bars in descending order of frequency.