Git setup - Adv R - Homework2

Jesinthan Joseph Durairaj - 24213676

Introduction

The `diamonds` dataset from the `ggplot2` package contains the prices and attributes of over 53,000 diamonds, including carat weight, cut, color, clarity, and price. This dataset is widely used for practicing data visualization and modeling in R.

In this brief report, we explore how **diamond price** varies with **carat size**, grouped by **cut quality**.

Data Summary

Below are some descriptive statistics for the two variables used in our plot:

- *Carat*

  - Minimum: 0.2
  - 1st Quartile (Q1): 0.4
  - Median: 0.7
  - Mean: 0.798
  - 3rd Quartile (Q3): 1.04
  - Maximum: 5.01
  - Standard Deviation: 0.474

- *Price (USD)*

  - Minimum: $326
  - 1st Quartile (Q1): $950
  - Median: $2401
  - Mean: $3933
  - 3rd Quartile (Q3): $5324
  - Maximum: $18,823
  - Standard Deviation: $3989

Scatter Plot of Diamond Price vs Carat

We visualize the relationship between `carat` and `price`, grouped by the diamond's **cut quality**.

Plot carat vs. price by cut

Diamond Price vs Carat by Cut