Homework 1

Nur Balqis Binti Roslan Ghani

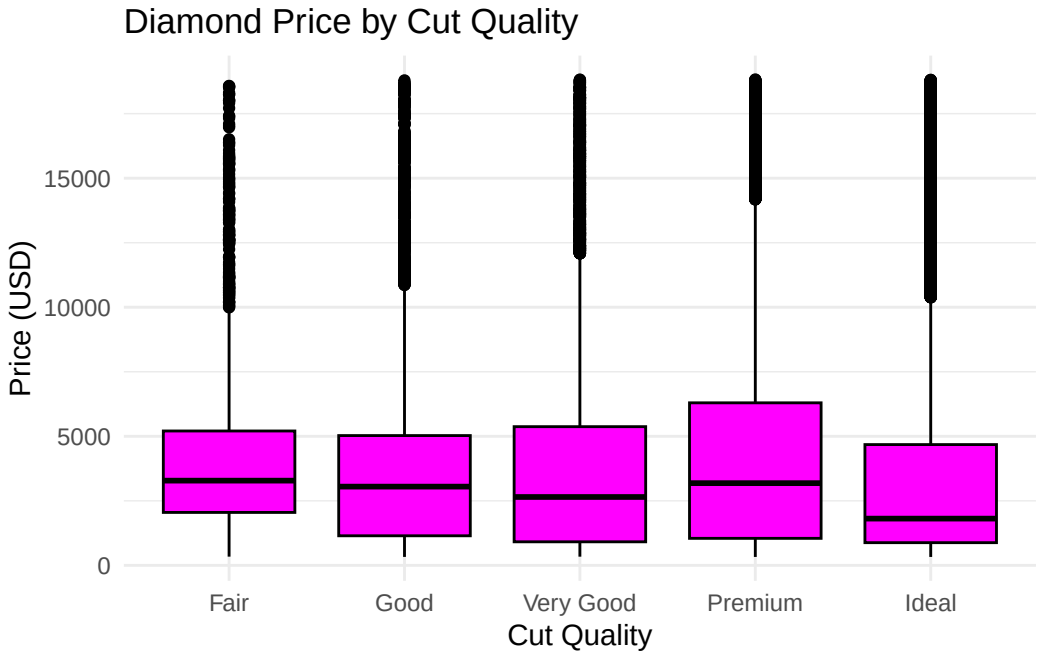
Introduction

The dataset used in this document is the diamonds dataset from the ggplot2 package. It contains over 50,000 observations on diamond characteristics such as carat, cut, colour, and price.

The relationship between a diamond's cut and price.

Graph

The distribution of diamond across various cut quality levels is displayed in the boxplot below. It enables us to comprehend how quality grading effects pricing.



This boxplot shows how diamond prices vary by cut quality using data from the ggplot2::diamonds dataset. The y-axis shows the price in USD, and the x-axis shows the five cut quality levels: Fair, Good, Very Good, Premium, and Ideal. The interquartile range is summarized by each box; outliers are shown by points beyond the whiskers, while the median is shown by a line. Ideal cuts, although of the finest quality, have somewhat lower median prices than Premium cuts, which have the highest median values. The reason for this could be that ideal cuts are usually less expensive because they have less carats. Many high-priced outliers and the largest price spread are seen in fair cuts. All things considered, it would appear that although cut influences pricing, other elements such as carat, weight and clarity are also important.