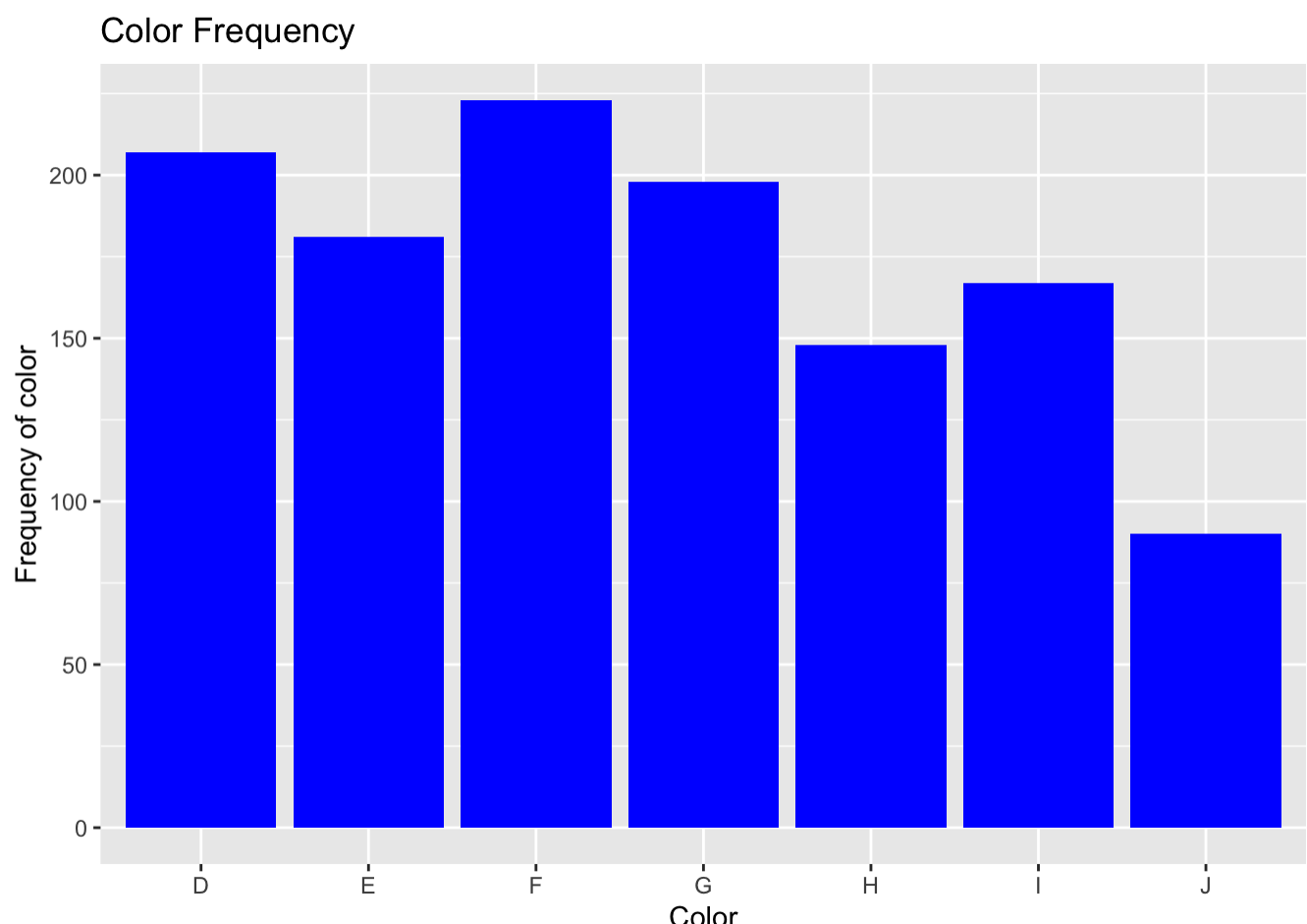


Color

Out of the 4 C attributes, color is the second most important one. The GIA diamond color grades, the Industry standard for a diamonds' color scale, range from D (colorless) to Z (light yellow or brown). Bluenile only sells colorless(D-F) and near colorless(G-H). In general, the more colorless the diamond the more expensive it is supposed to be.

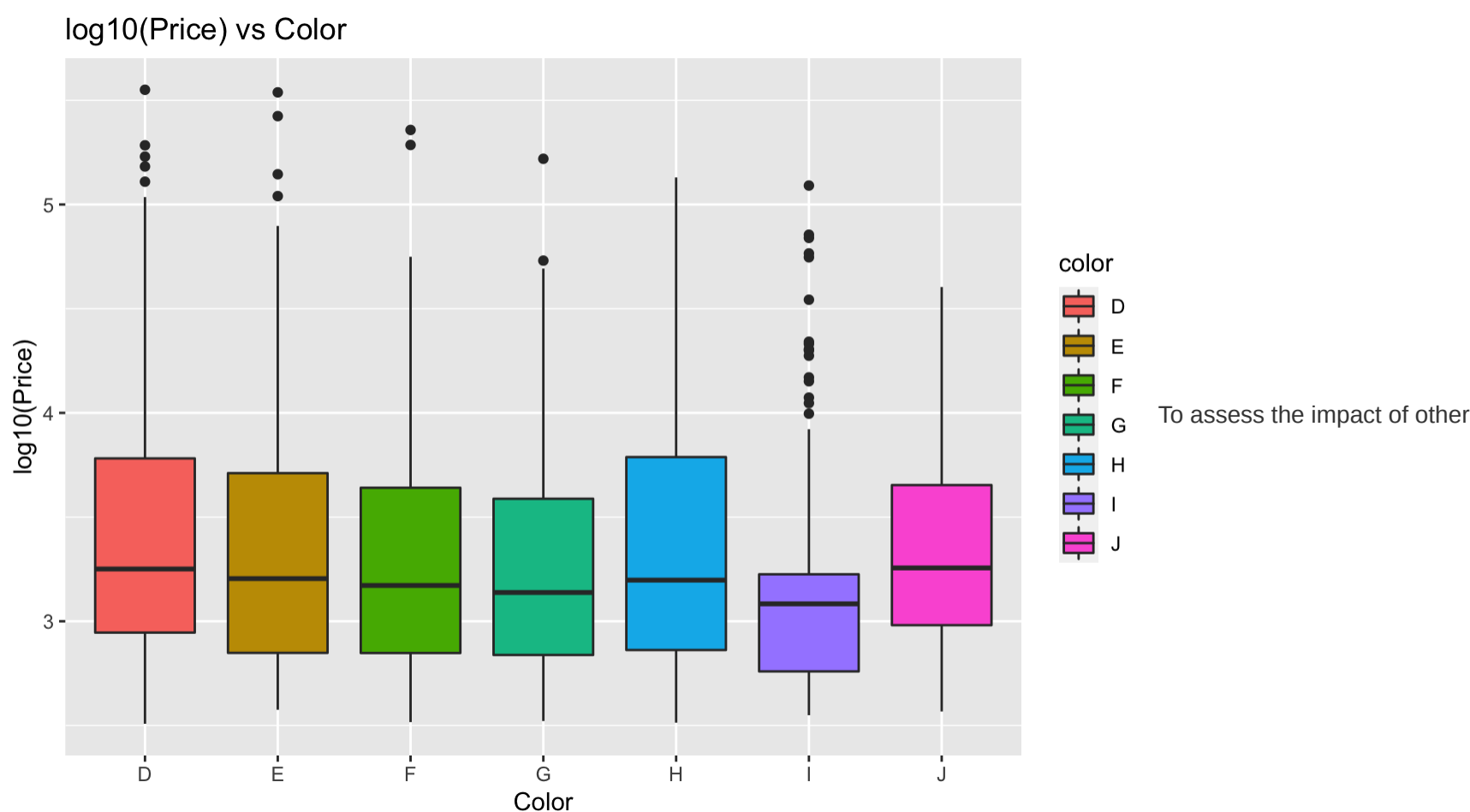
Color Frequency

We started off by checking the frequency of occurrences of various colors of diamonds in our data set. It looks like all the colors are represented well enough in our data set.



Price vs Color

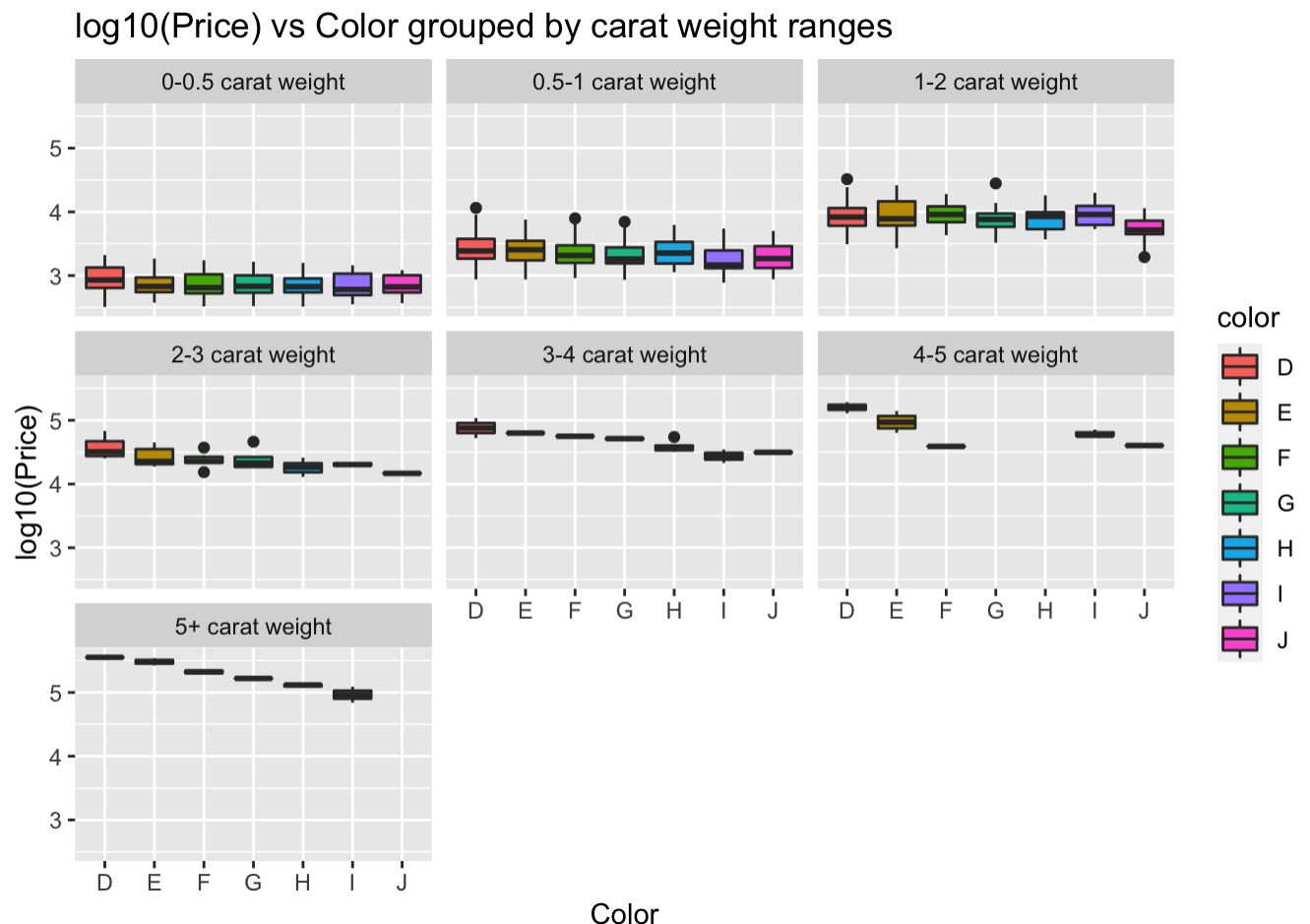
Then we created a box plot to compare price ranges for different colored diamonds. We transformed the Price variable using the log10 function to reduce the variation between the data points. The below box plot fails to support the claim that the more colorless a diamond the more it should cost. Surprisingly, H and J grade diamonds show a higher mean price than the colorless diamonds(D, E, F). We suspected some variables other than color might be impacting the price of the diamond.



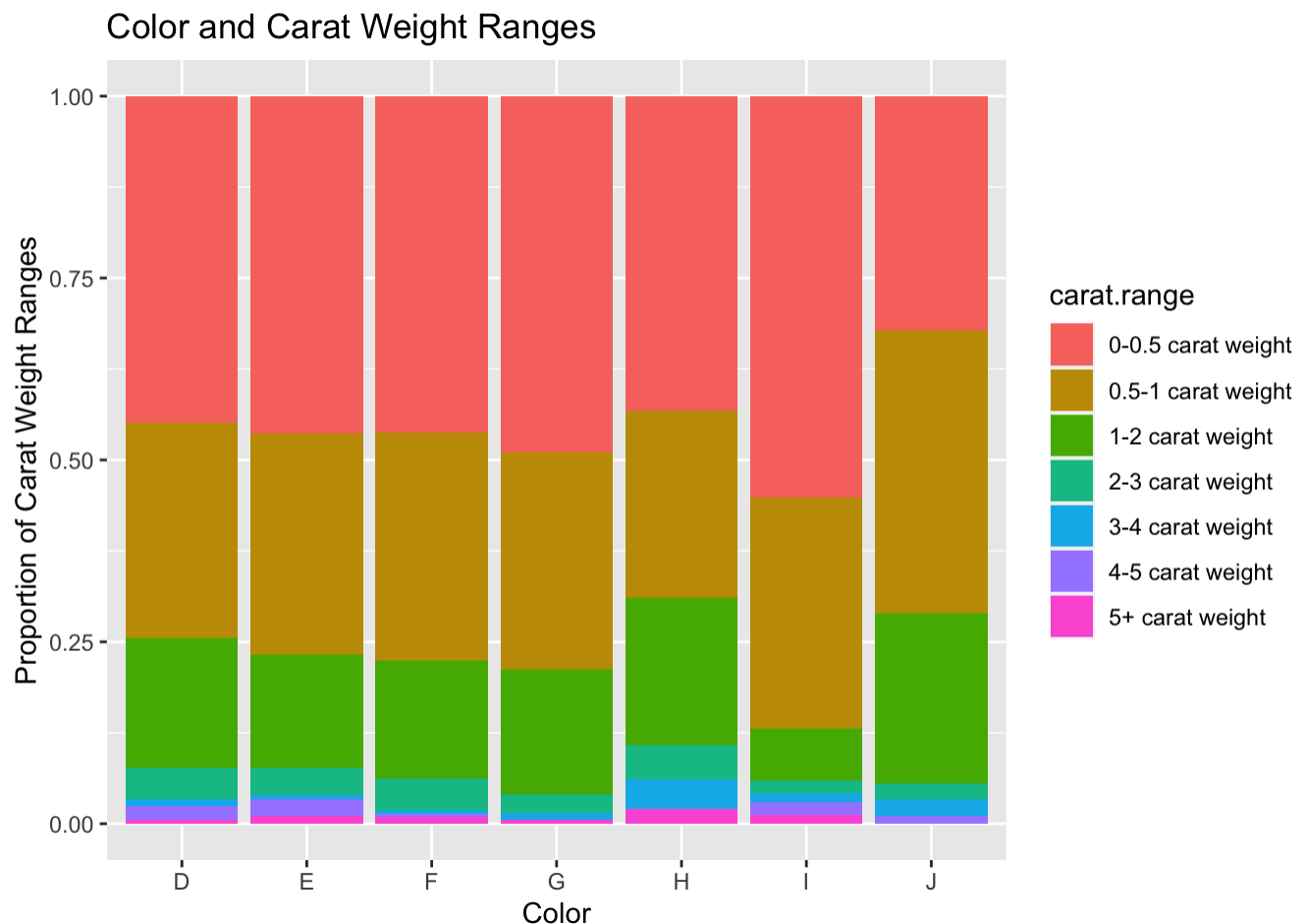
variables along with color on the price, we created box plots by grouping the other variables into their respective categories and then assessed the impact of color on price.

Color and Carat Weight Ranges

Here we have split our data set into different carat weight ranges, and then tried to assess the impact of color on each of these ranges separately. In general, the more colorless the diamond the more expensive it is supposed to be. However, this trend becomes much more apparent for diamonds with >1 carat weight, which can be seen from the boxplots below. As the carat size increases, the colorless diamonds become much more expensive than the more colored diamond. This directly supports the claim made by Bluenile that color is more visible in large diamonds.

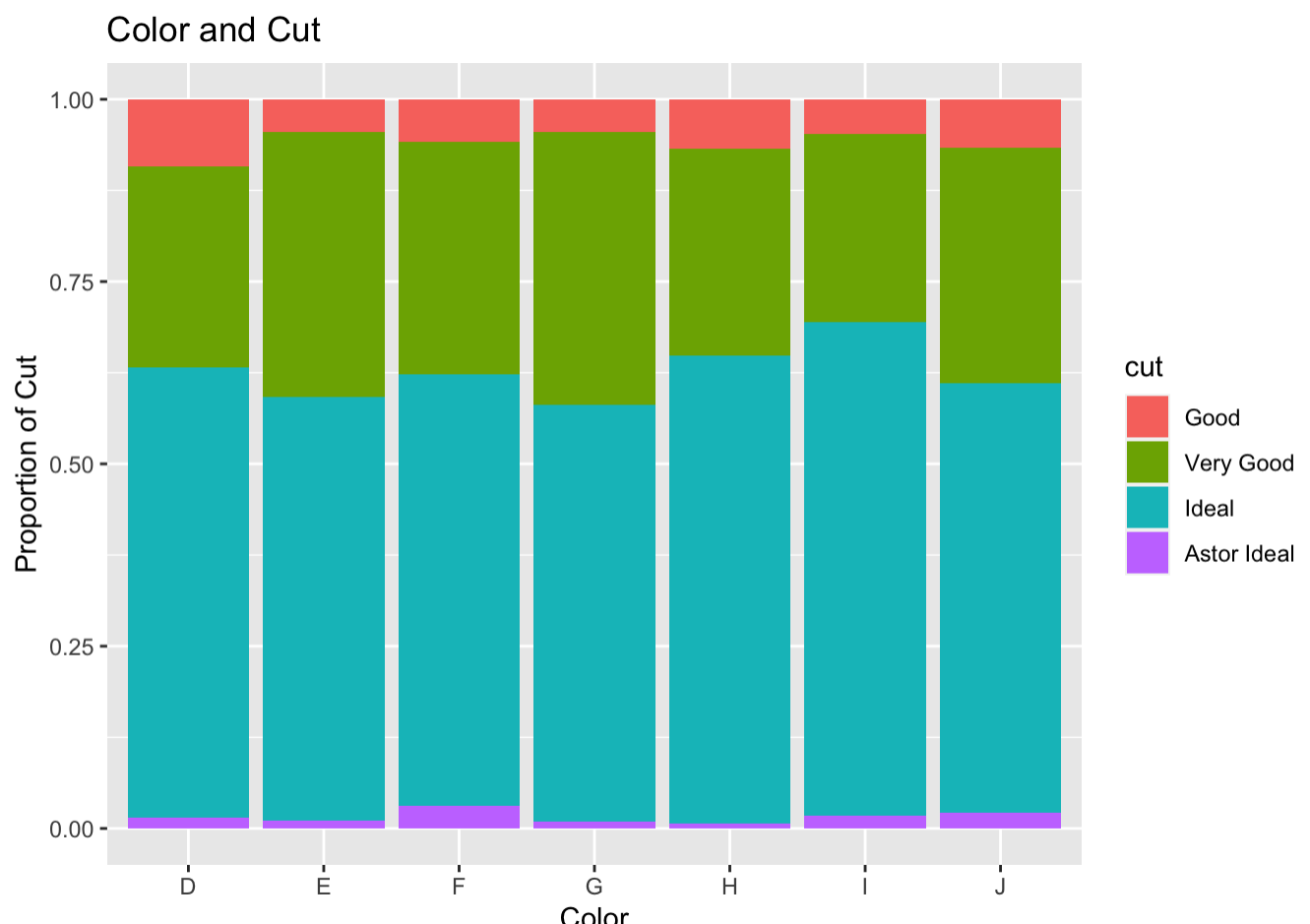


Then to find a correlation between color and carat weight ranges we also created proportion bar charts. This chart will help us answer questions such as does a particular carat weight range occur more often for a particular color in our data set? The most interesting find in the below color and carat weight ranges chart is that the proportion of diamonds with more than 1 carat weight is much higher in both H and J color-grade diamonds. This actually explains why H and J color-grade diamonds overall had a higher than expected mean price, almost comparable to D/E/F diamonds, in the log10(Price) vs Color box plot.



Color and Cut

To find a correlation between color and cut we also created proportion bar chart. This chart will help us answer questions such as does a particular cut type occur more often for a particular color in our data set? In the below color and cut plot, we can see that the proportion of Astor ideal cut is higher for F color-grade diamond compared to others. But no other major bias was found between color and cut.



Color and Clarity

To find a correlation between color and clarity we also created proportion bar chart. This chart will help us answer questions such as does a particular clarity occur more often for a particular color in our data set?. In the below color and clarity plot, we can interestingly see that the best clarity, which is FL, only occurs for D and F grade diamonds in our data set.

