Color Writeup

Terrance Luangrath

2025-03-09

Data Description

The data set, `diamonds4.csv`, contains 5 varaibles that describes more than 1000 different diamonds for sale.

- `Carat`: *fill in description*
- `Clarity`: *fill in description*
- `Color`: how colorless a diamond is, the more colorless the diamonds, the rarer it tends to be
- `Cut`: *fill in description*
- `Price`: the price value of the diamond in USD

Variable Analysis: Color

The diamond color refers to how colorless a diamond is. From the Blue Nile website, color is the second most important of the 4Cs of diamond. The more colorless the diamond is the rarer it is. Diamond colors are classify in three main categories,

- Colorless Diamonds
- Near-Colorless Diamonds
- Faint Diamonds

from these three catgories, diamond color are grades from the ranges of D (colorless) to K (faintly colored). However, in this dataset, there exists only diamond graded from D to J.

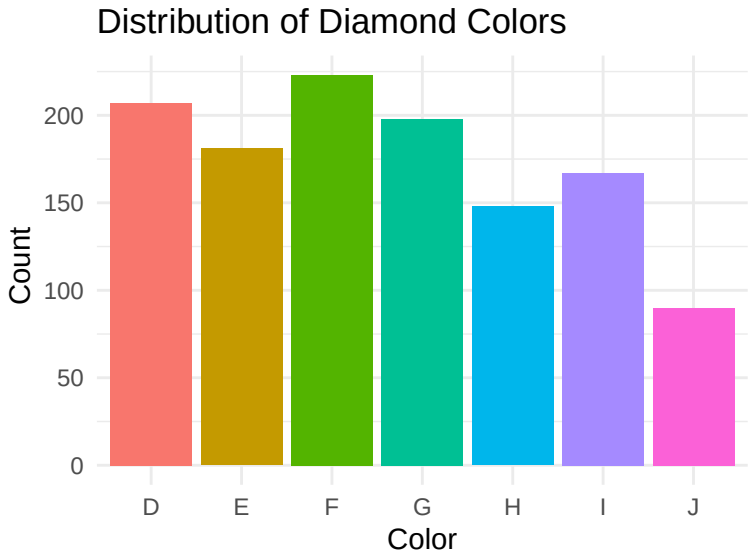


Figure 1: Diamond Color Distribution

Looking at Figure **??**, the distribution of diamond color appears fairly consistent, with counts ranging between 150-200 for most color grades. However, color `J` semms to have around 100 diamonds in the data set.

Each diamond color falls into one of the three categories from Colorless to Faintly Colored.

- Colorless diamonds: `D`, `E`, `F` Color Diamonds
- Near-colorless diamonds: `G`, `H`, `I`, `J` Color Diamonds
- Faint color diamonds: `K` Color Diamonds

Since there is no `K` in the data set, it would be assumed there is no faint color diamonds in the data set. To facilitate the analysis, we will create a new column called `color_cateogry` to group each `color` with their respected category.

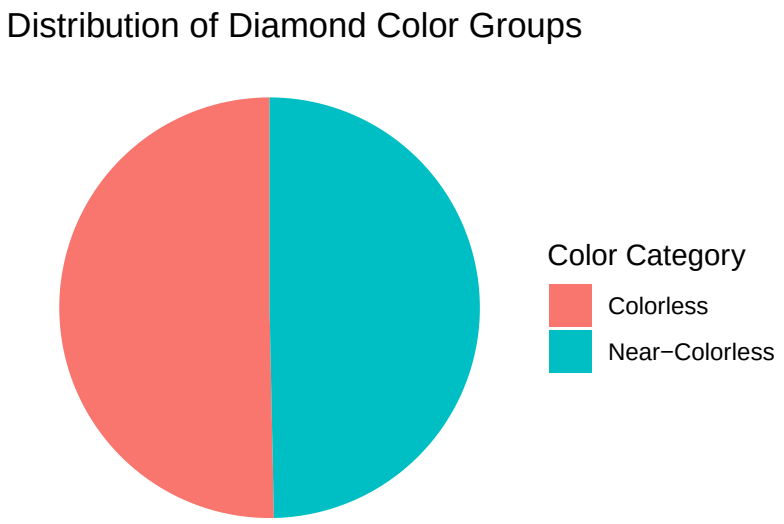


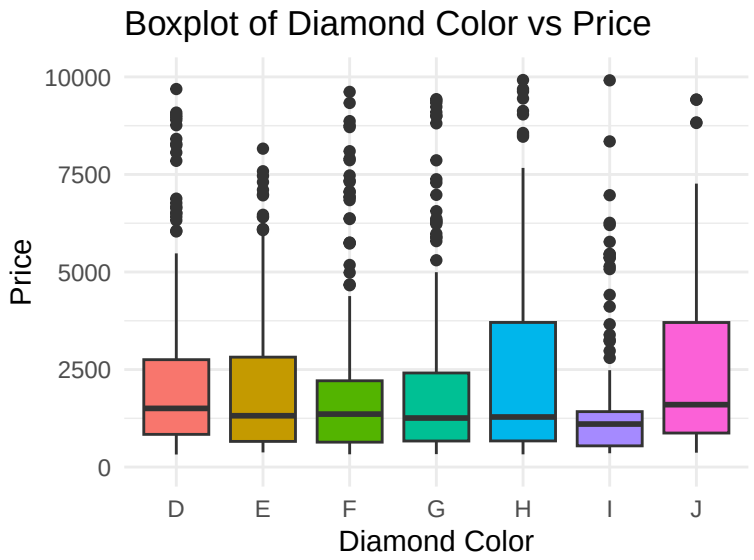
Figure 2: Diamond Color Category visualize through a Pie Chart

Looking at **??**, the distribution of diamond colors seems to be nearly even, with `Colorless` containing slightly higher amount of diamonds. With a well spread of diamond colors, further analysis is needed to understand the claim and how diamond color would be correlates with the other factors in the data set.

Bivariate Analysis

Bivariate Analysis on Prices

```
## # A tibble: 7 x 6
##   color   min    q1 median    q3    max
##   <chr> <int> <dbl>  <dbl> <dbl>  <int>
## 1 D       322  882    1781 6048. 355403
## 2 E       376  705    1602 5138  345397
## 3 F       328  704    1485 4372  227960
## 4 G       332  689.   1374 3872. 165766
## 5 H       326  728.   1575 6135  134856
## 6 I       354  574.   1212 1680. 123311
## 7 J       369  958.   1803 4504   40184
```

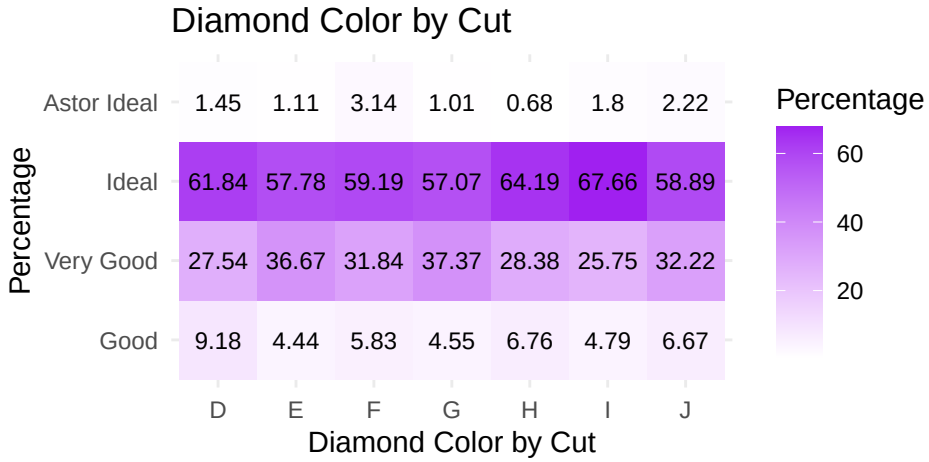


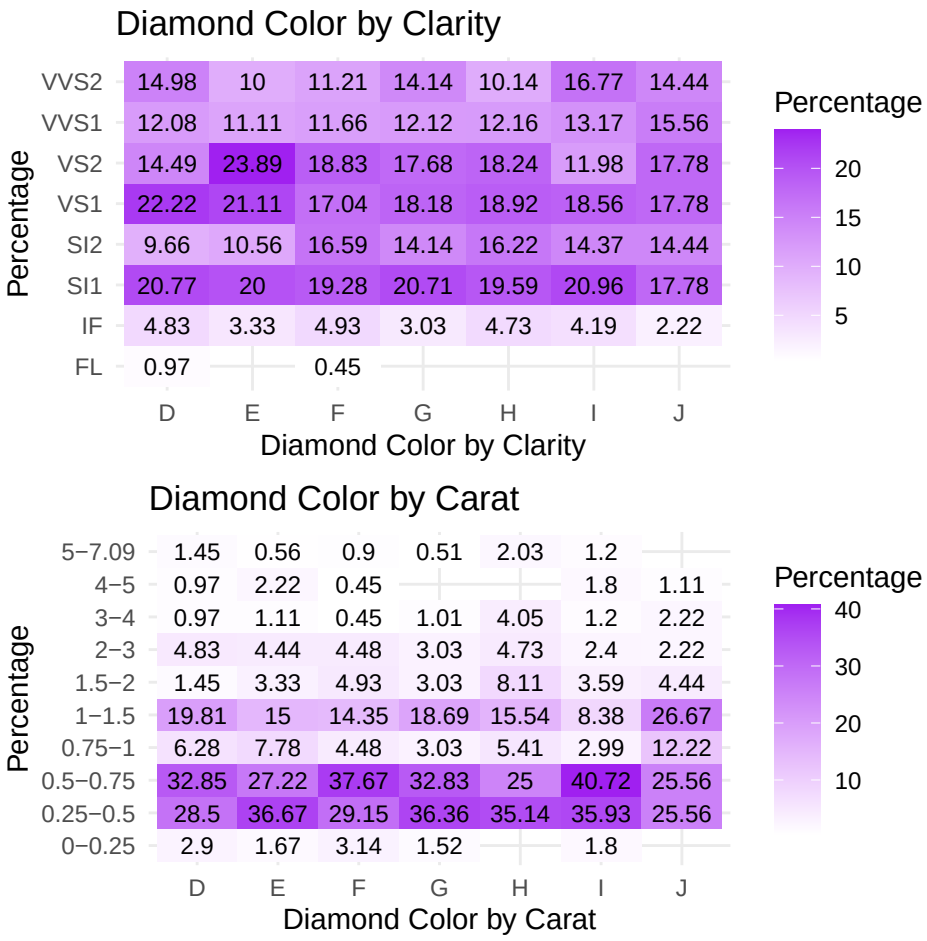
On the Blue Nile site, it mentions that color is the second most important among the 4C's for the price of the diamond. Reviewing the box plot, the colorless diamonds (`D`, `E`, and `F`) tends to have higher prices than the near-colorless diamond (`G`, `H`, `I`, and `J`). However, some colorless diamond have higher median values than the near-colorless diamond:

- `H` have higher median values than the colorless diamond `F`
- `J` (the lowest color grade) has a higher median value than all the colorless diamond.

This suggests that color does play a significant factors in the diamond price, but it's not the most important factors to the diamond prices. Other characteristics, such as `cut`, `clarity`, and `carat`, needs to be considered to help prove the claim.

Color vs Other Diamond Characteristcs




Diamond Color by Clarity
VVS2 14.98 10 11.21 14.14 10.14 16.77 14.44
VVS1 12.08 11.11 11.66 12.12 12.16 13.17 15.56
VS2 14.49 23.89 18.83 17.68 18.24 11.98 17.78
VS1 22.22 21.11 17.04 18.18 18.92 18.56 17.78
SI2 9.66 10.56 16.59 14.14 16.22 14.37 14.44
SI1 20.77 20 19.28 20.71 19.59 20.96 17.78
IF 4.83 3.33 4.93 3.03 4.73 4.19 2.22
FL 0.97 0.45
D E F G H I J
Percentage
Diamond Color by Clarity
Percentage
20
15
10
5
Diamond Color by Carat
5−7.09 1.45 0.56 0.9 0.51 2.03 1.2
4−5 0.97 2.22 0.45 1.8 1.11
3−4 0.97 1.11 0.45 1.01 4.05 1.2 2.22
2−3 4.83 4.44 4.48 3.03 4.73 2.4 2.22
1.5−2 1.45 3.33 4.93 3.03 8.11 3.59 4.44
1−1.5 19.81 15 14.35 18.69 15.54 8.38 26.67
0.75−1 6.28 7.78 4.48 3.03 5.41 2.99 12.22
0.5−0.75 32.85 27.22 37.67 32.83 25 40.72 25.56
0.25−0.5 28.5 36.67 29.15 36.36 35.14 35.93 25.56
0−0.25 2.9 1.67 3.14 1.52 1.8
D E F G H I J
Percentage
Diamond Color by Carat
Percentage
40
30
20
10