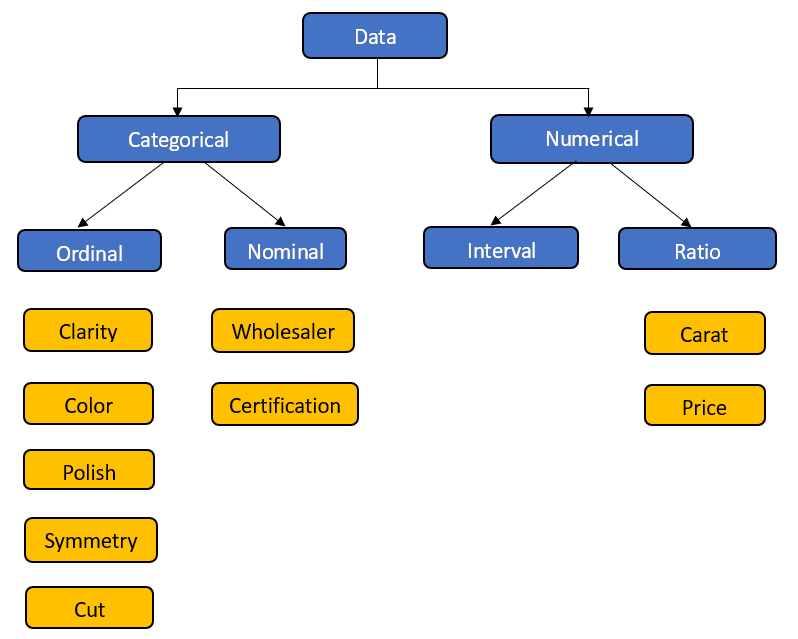
# **Descriptive Analysis**

# **Data Types of the Variables**

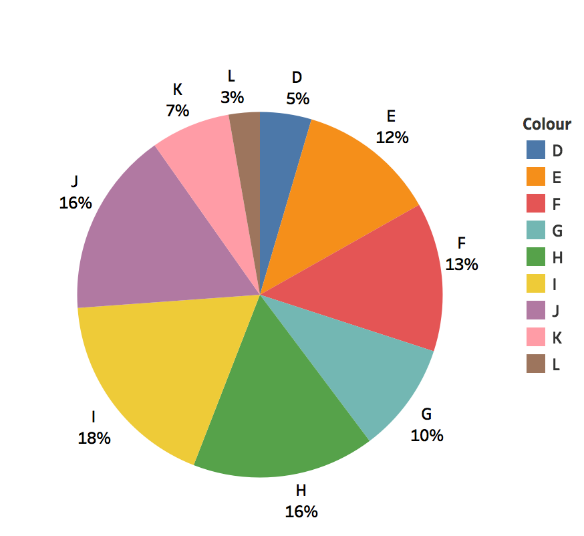
In the given case. there are nine variables. Seven of them are categorical and two are numerical. Within, the categorical variables 5 of them are ordinal variables. To gain an intuitive understanding of the hierarchy of these levels, each ordinal variable will be renamed from worst the best by following an increasing integer scale. Also, numerical variables carat and price are ratio variables within the numerical group.



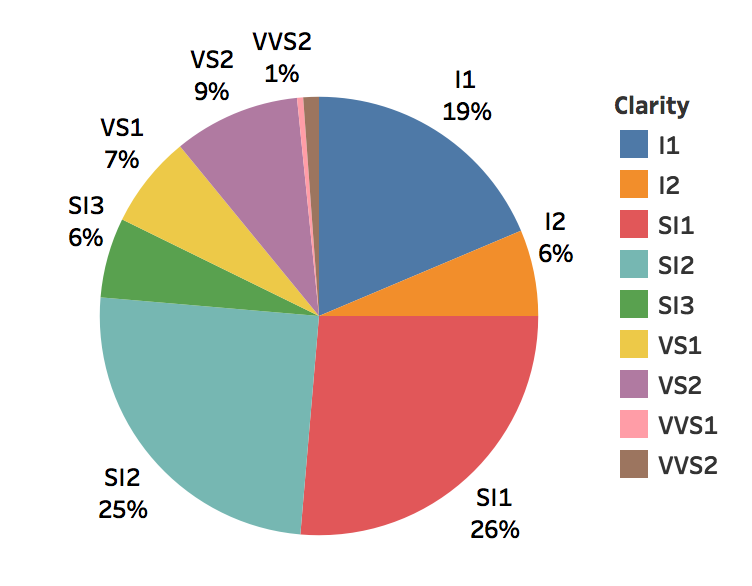
Dataset has 414 records with no missing values. The categories with low frequency on each independent variable are taken into consider for feature engineering.

## Distribution of Independent Variables

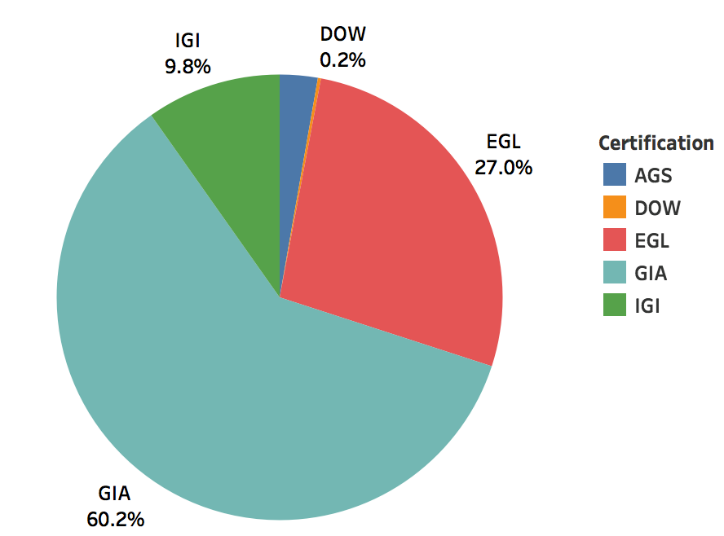
**Distribution of Color:**



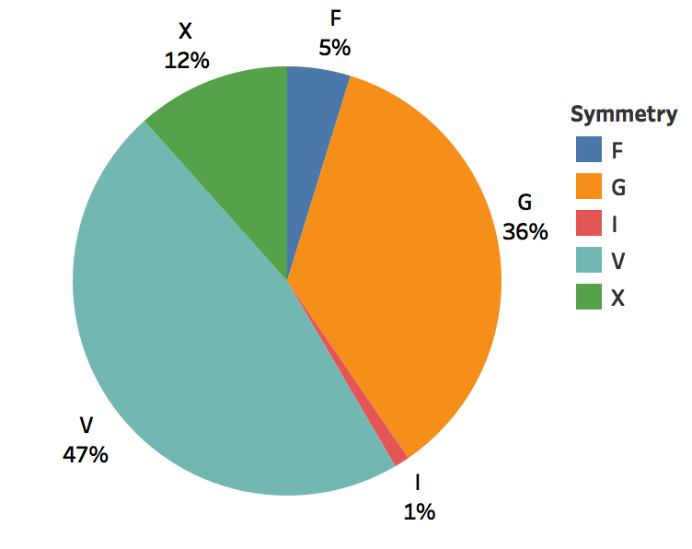
**Distribution of Clarity:**



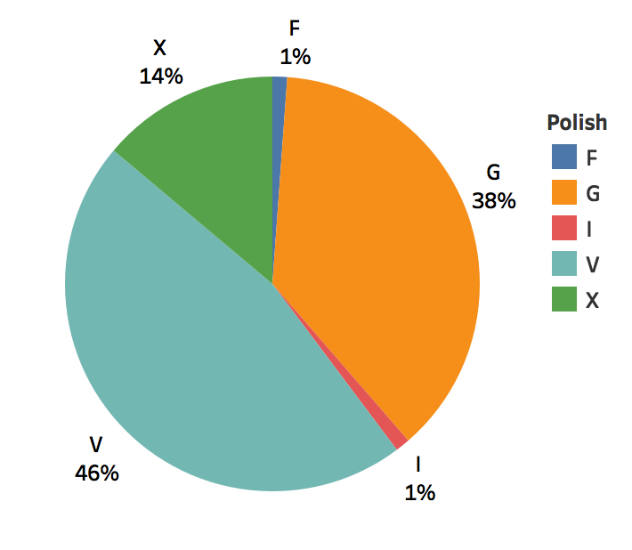
**Distribution of Certification:**



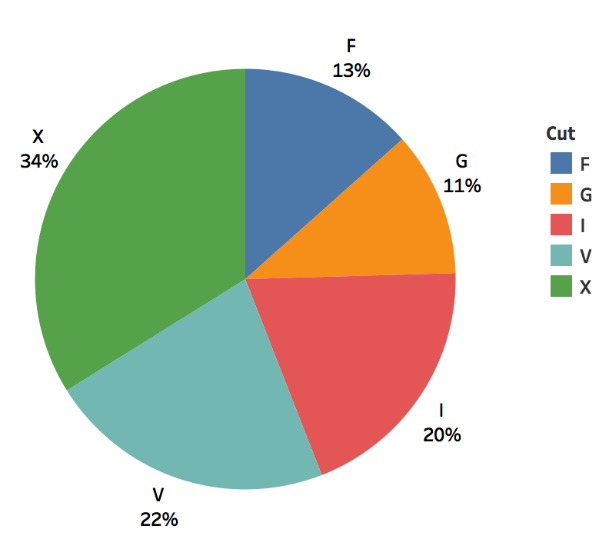
**Distribution of Symmetry:**



**Distribution of Polish:**



**Distribution of Cut:**



## Density Distribution

**Density Distribution of Price:**

**ADD EXPLANATION HERE. WHY WE HAVE IT FOR PRICE AND CARAT. HOW IS IT HELPING US. (SIGNALLING DIFFERENCE). WHAT ARE THE BLUE AND RED LINES**

****

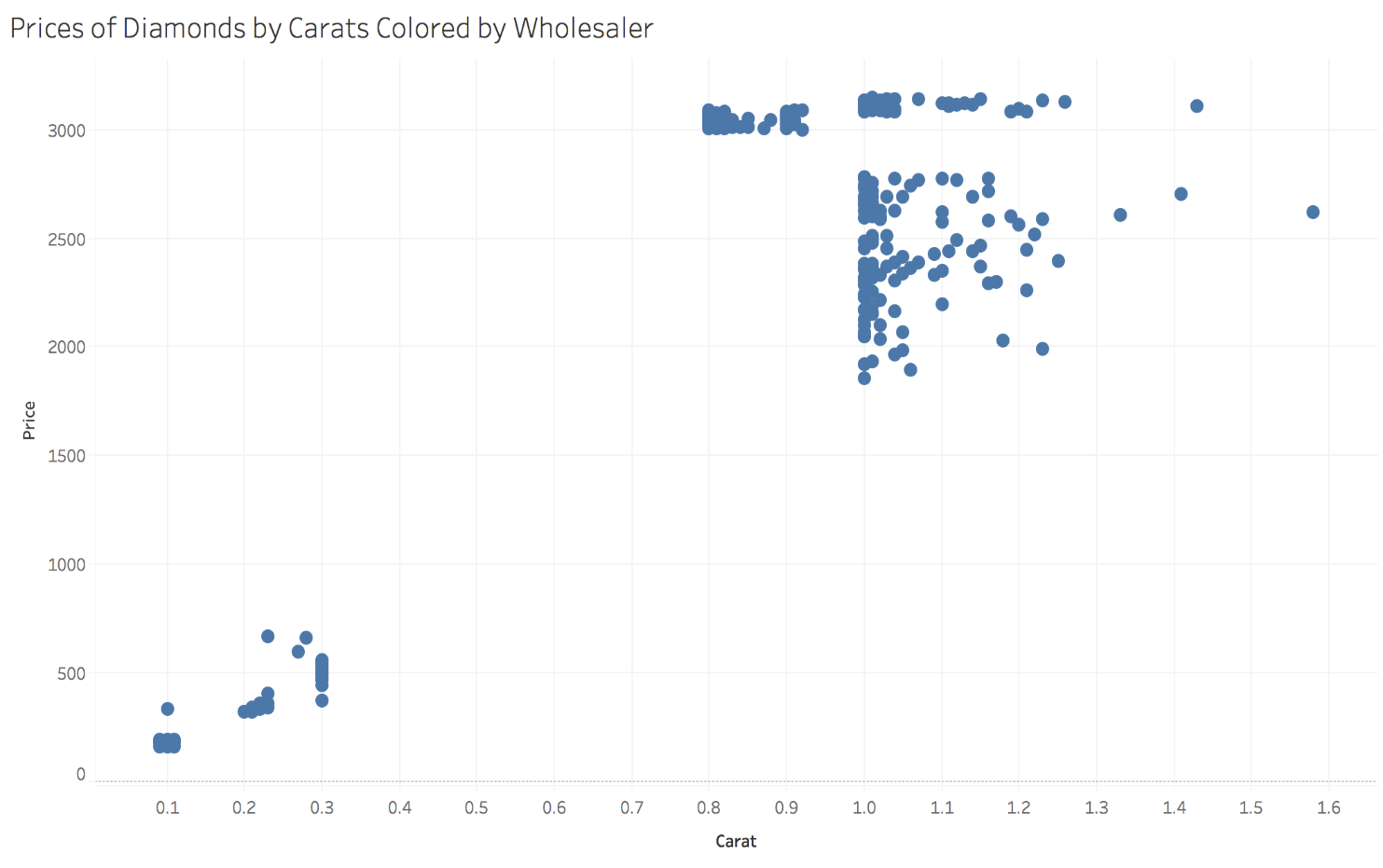
**Density Distribution of Carat:**

****

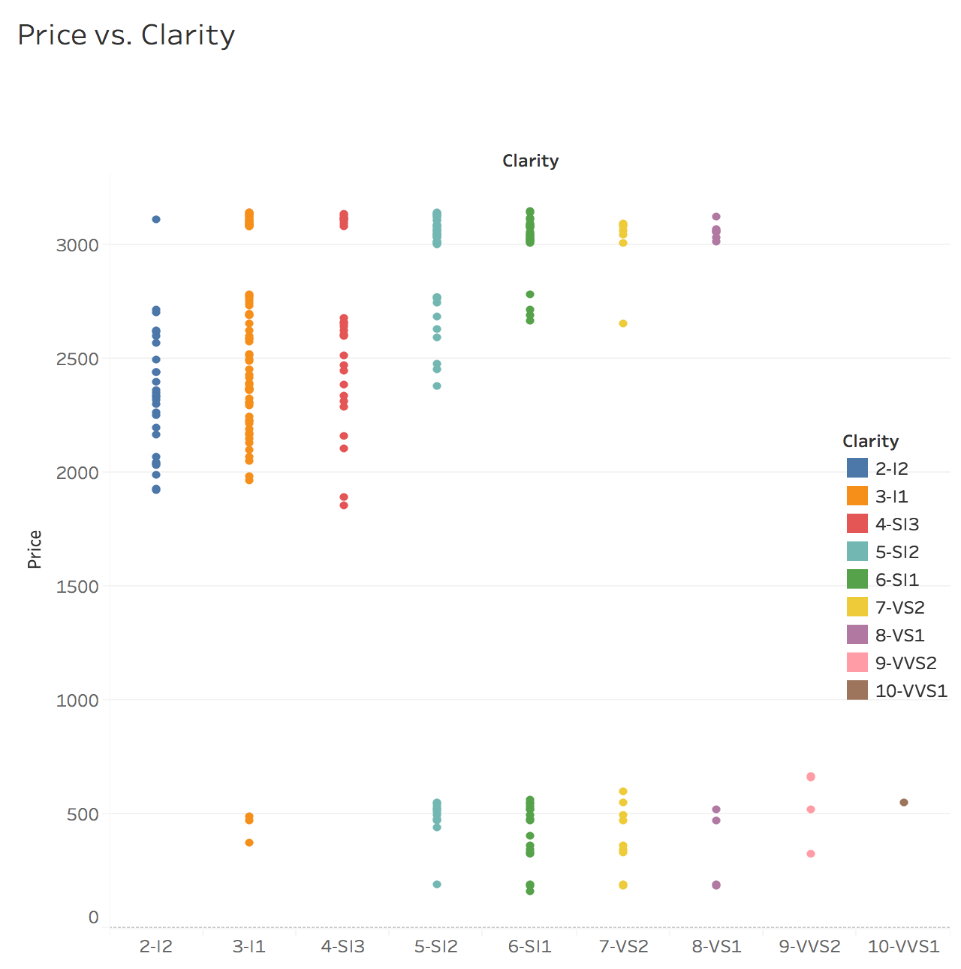
## Bivariate Distribution

ADD EXPLANATION HERE. PLOTTING BI-VARIATES HELP US UNDERSTAND RELATIONSHIP OF INDEPENDENTS VS. DEPENDENT VARIABLE PRICE. TRY TO EXPAND ON THIS LITTLE MORE.

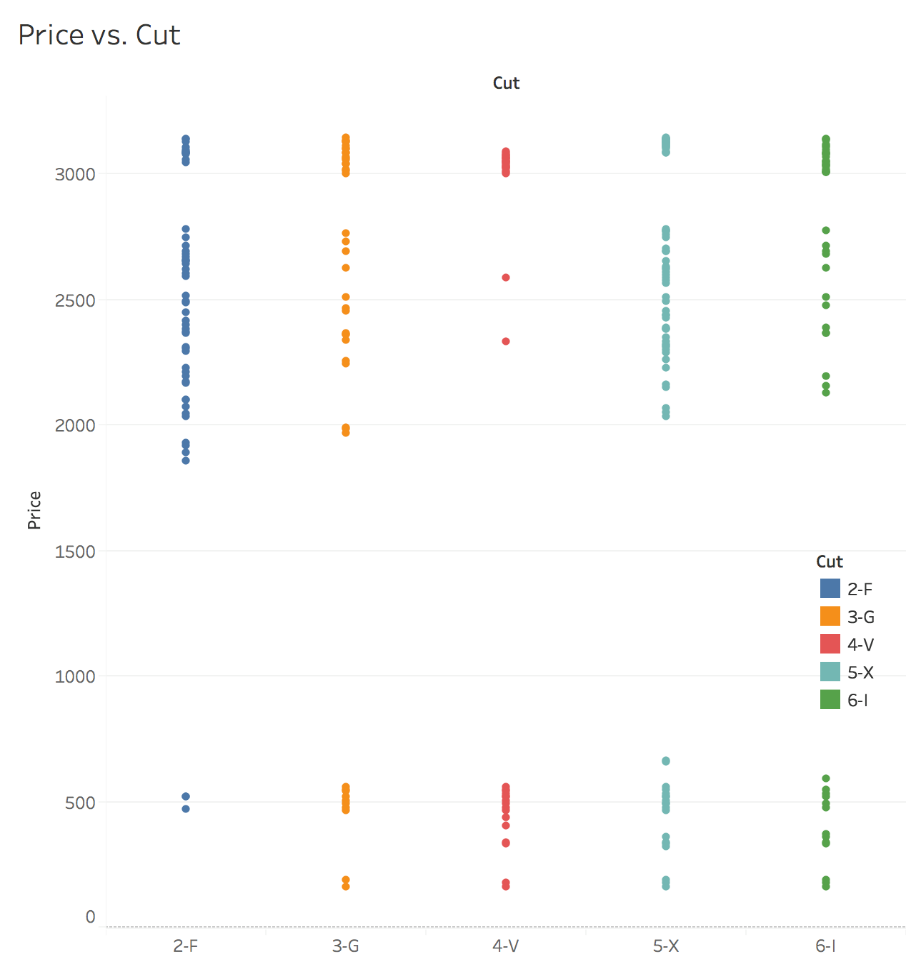
**Price vs Carat:**



**Price vs Clarity:**

****

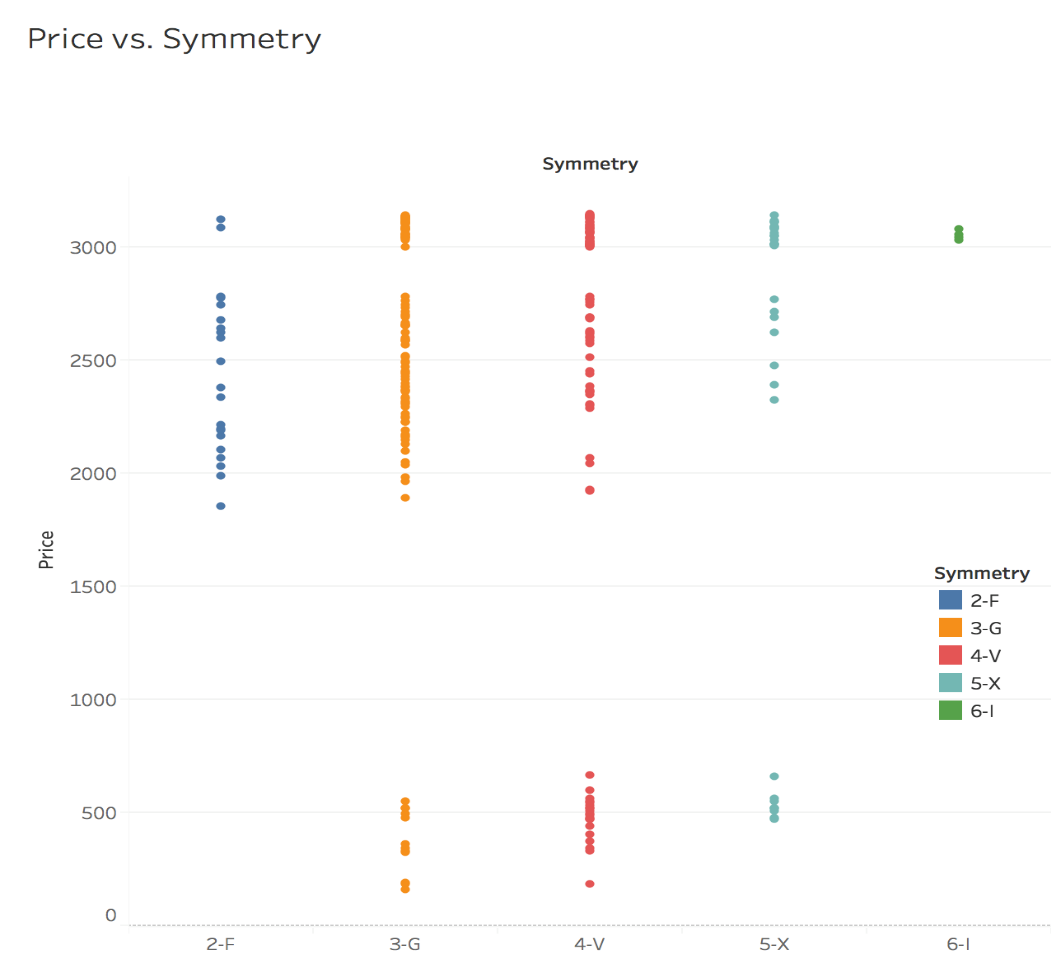
**Price vs Cut:**



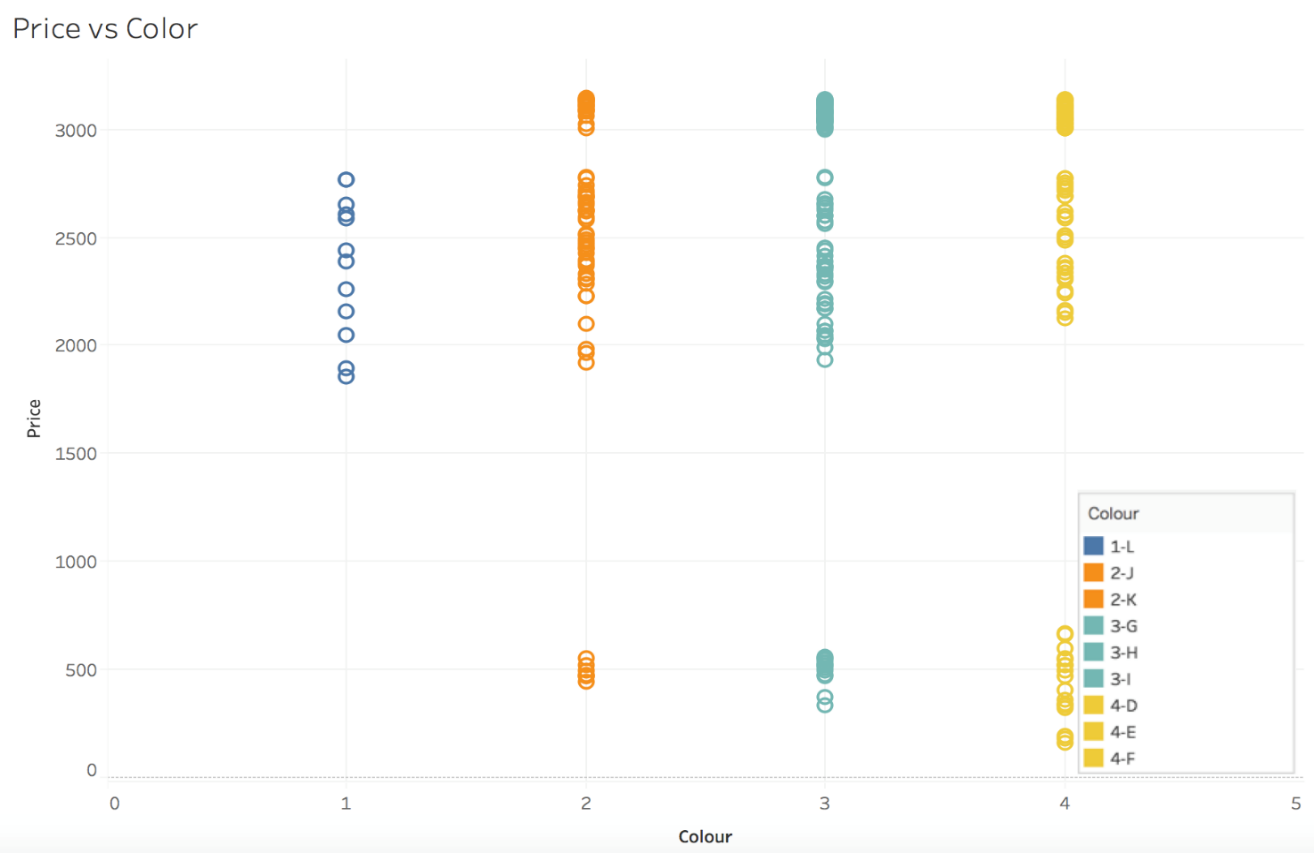
**Price vs Polish:**

****

**Price vs Symmetry:**

****

**Price vs Color:**

****