**INTRODUCTION**

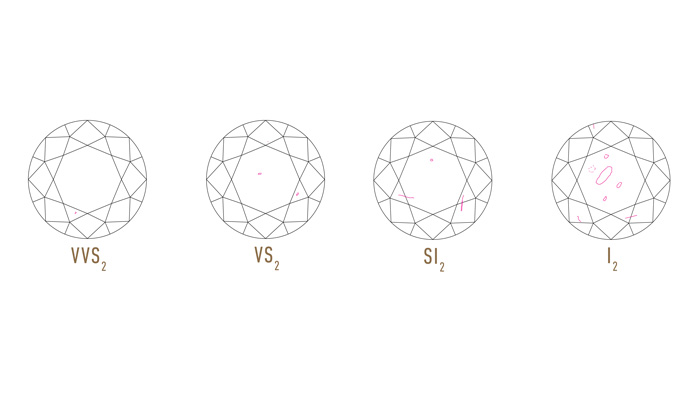
As the result of marketing of diamond industry, many of us believed diamond is the hardest material in the world and it is very very rare, as a good wish people also would like their relationship between their partner is also strongest in the world.

In the early 1940s, the diamond mining output increase is unprecedented, people want to figure out a way to assess a diamond’s value so they can have a universal method to assess diamond from different regions and retailers. The founder of Gemological Institute of America (GIA) Robert M. Shipley has introduced the term 4Cs to help his students to remember the four factors that can describe a cut diamond: color, clarity, cut and carat weight. Slowly this 4Cs became a standard way to measure the value of the diamond.

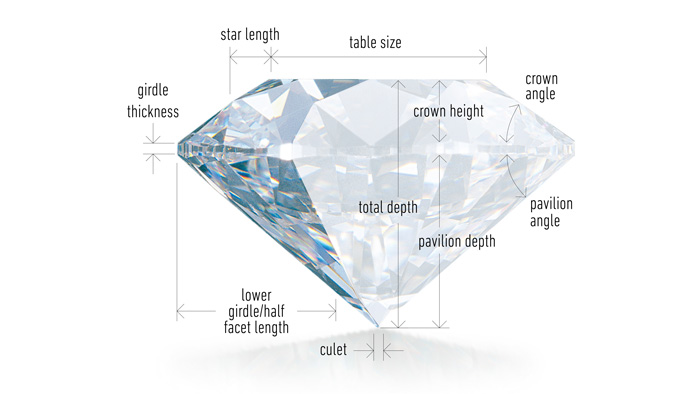
The first “c” s the color, the color used for describing a diamond is begin with grade D which is the first letter for diamond. From grade D to Z it became a color-grading system to measures the degree of colorlessness of a diamond. D is most colorless and Z is the color trend to yellow. When the other characters are the same a diamond with color D will have more value than color Z. If the diamond has a pure special color such as pink, yellow or green it will contain significantly higher value.



The second “c” refers to clarity, this indicates how much inclusions in the internal of a diamond and how much blemishes on the external of a diamond. The number, size, relief, nature and position of those inclusions and blemishes will help determine the value of a diamond. There are total of 11 grades for clarity, they are: Flawless (FL), Internally Flawless (IF), Very Very Slightly Included (VVS1 and VVS2), Very Slightly Included (VS1 and VS2), Slightly Included (SI1 and SI2) and Included (I1, I2 and I3). The higher grading the higher value the diamond has.



The third “c” means cut, which is the most important process in diamond manufactory. A well cut diamond will maximum the reflection of the light and make it looks brilliant. The cut is grading in the following grade: Excellent Cut, Very Good Cut, Good Cut, Fair Cut and Poor Cut. The better cutting a diamond has the higher value.



The last “c” is the carat weight of the diamond, carat is the unit to describe the weight of a diamond, a carat is equal to 200 milligrams. The heavier the more expensive.

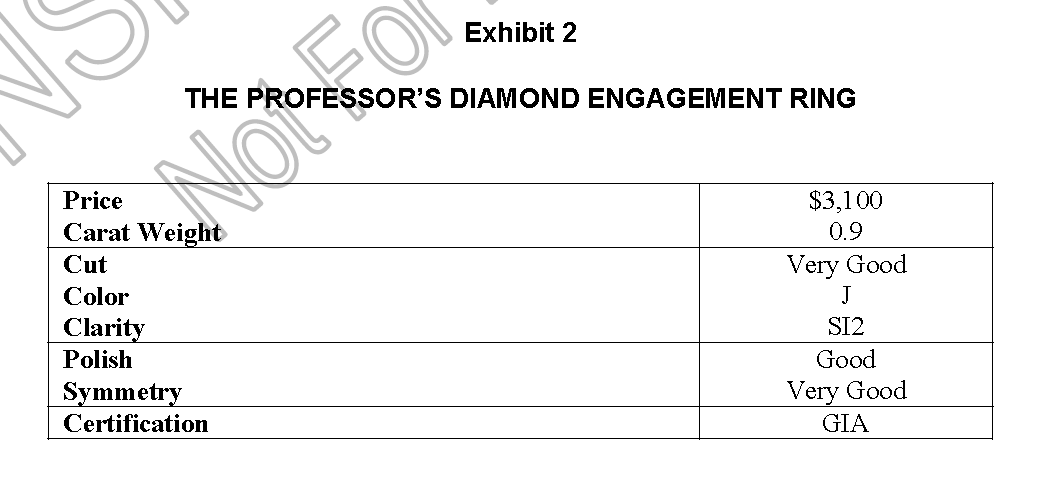


There are also other characters may affect the value of a diamond, they are Polish, Symmetry and Fluorescence.

**PROBLEM STATEMENT**

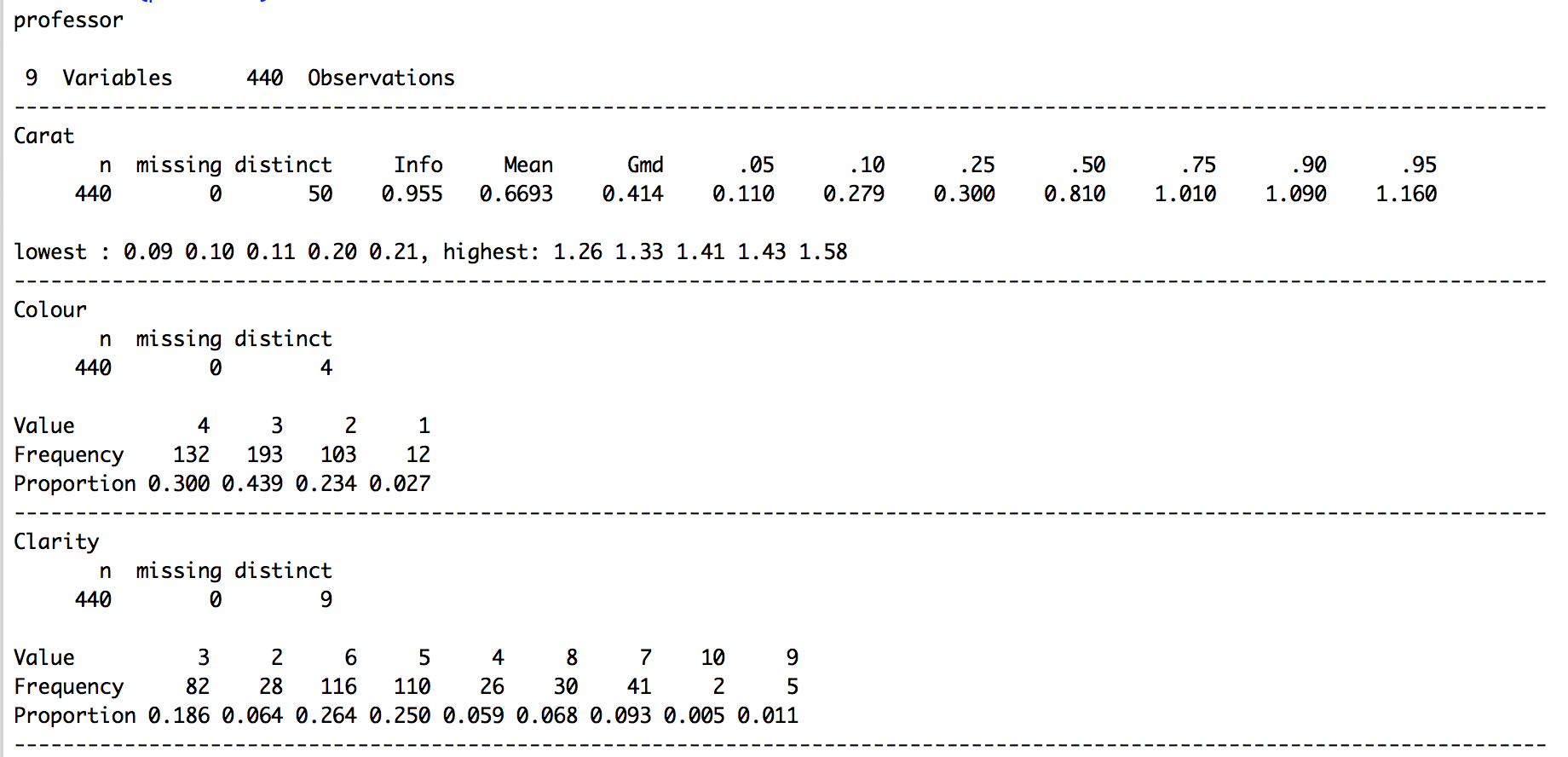
The professor’s girlfriend had already hinted third times about marriage, under this pressure the professor finally decided for the proposal. Find a nice diamond engagement ring is the first step but is not as easy as his initially thought.

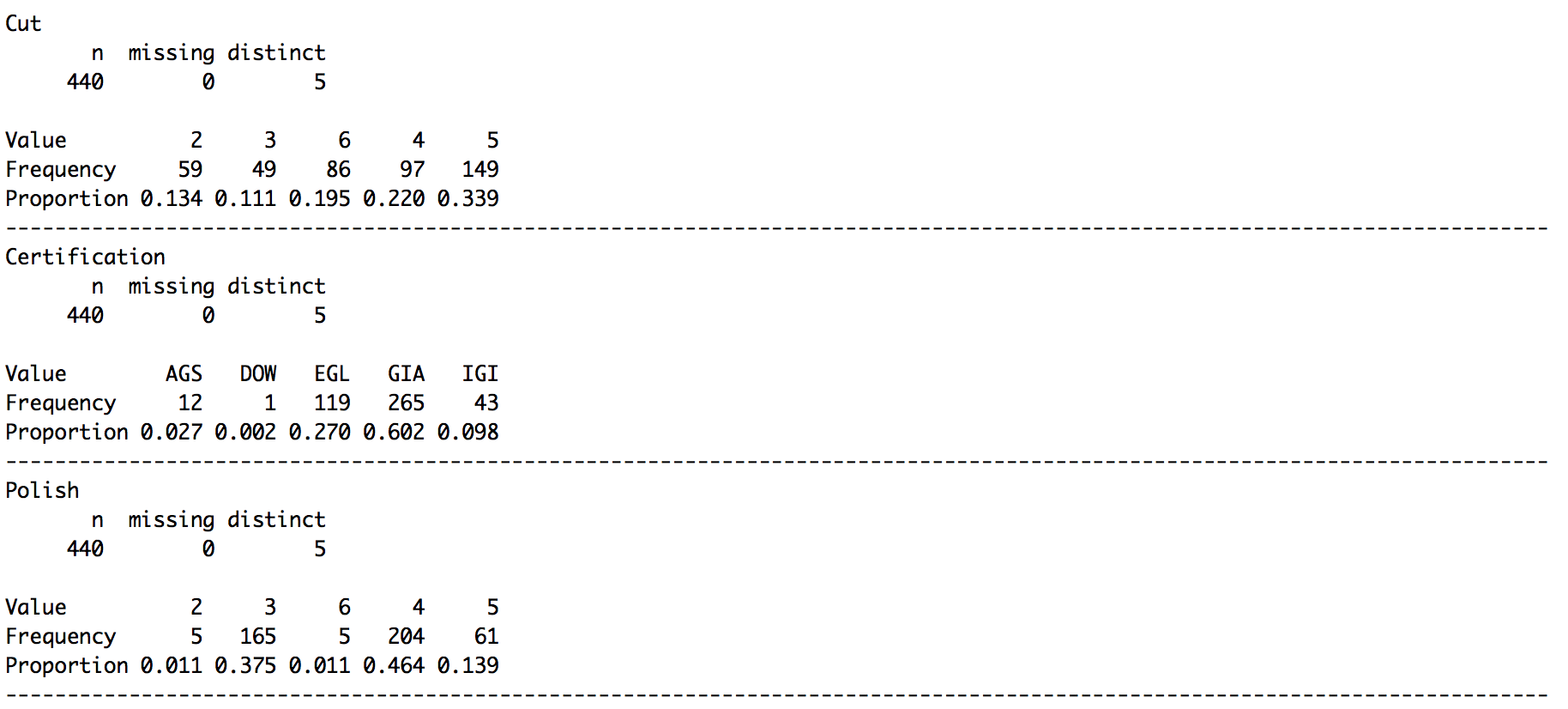
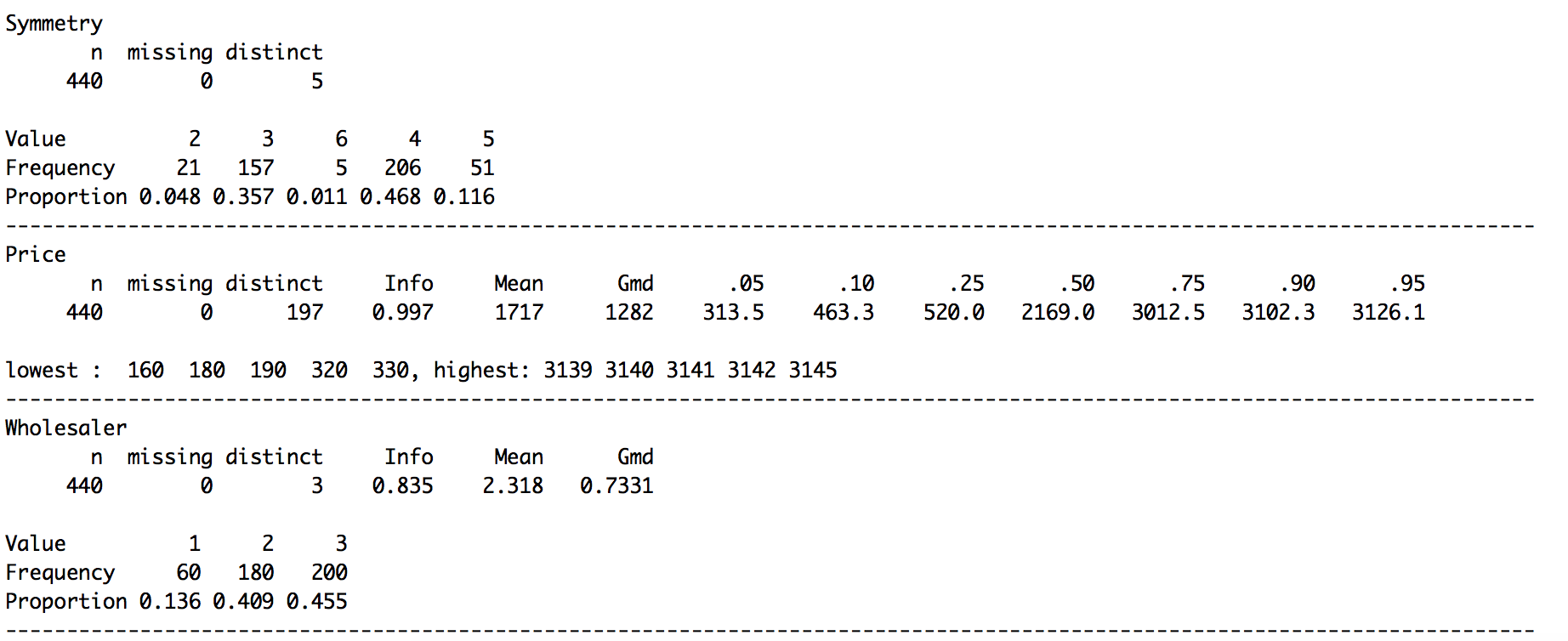
Is it fair to pay $3,100 for a diamond which has 0.9 carats in weight, J color, SI2 clarity, Very good cut, Good polish, Very Good Symmetry and come with GIA certificate in store become the biggest question in the professor’s mind.



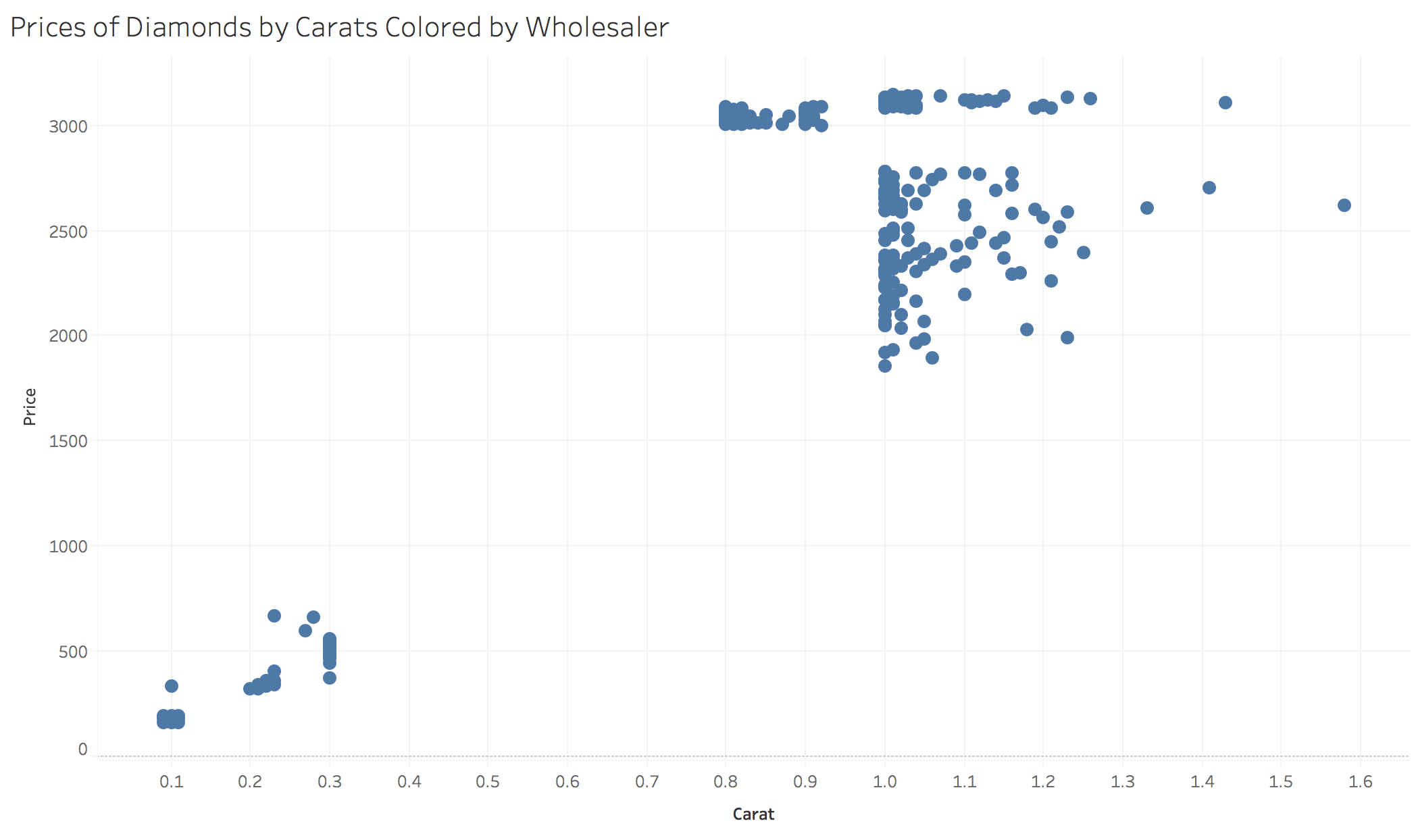
Now the professor decided to learn the 4Cs assessment standard and collecting available information from different diamond online retailer to build a module to assess a diamond's value based on the characters in order to answer his question.

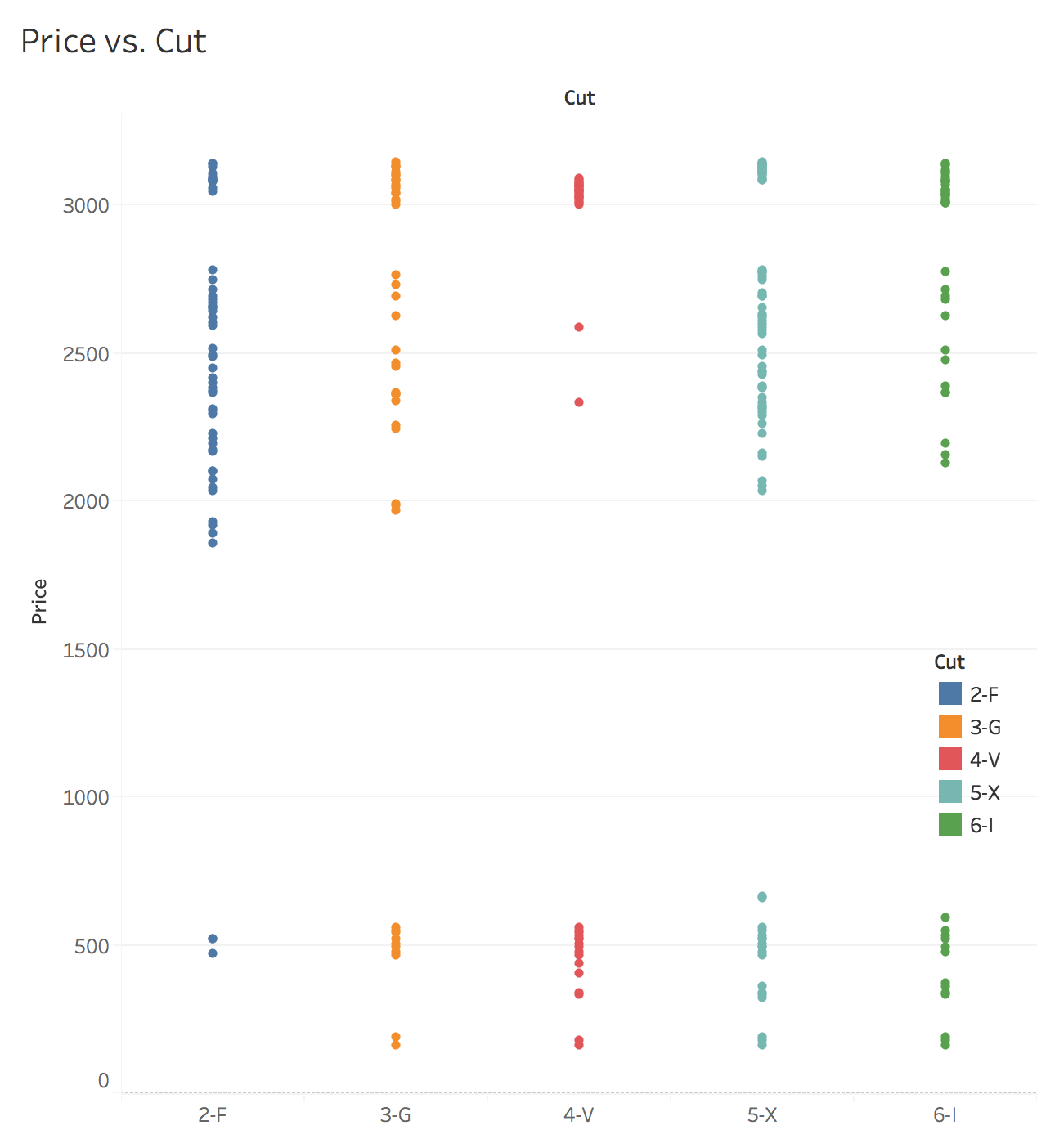
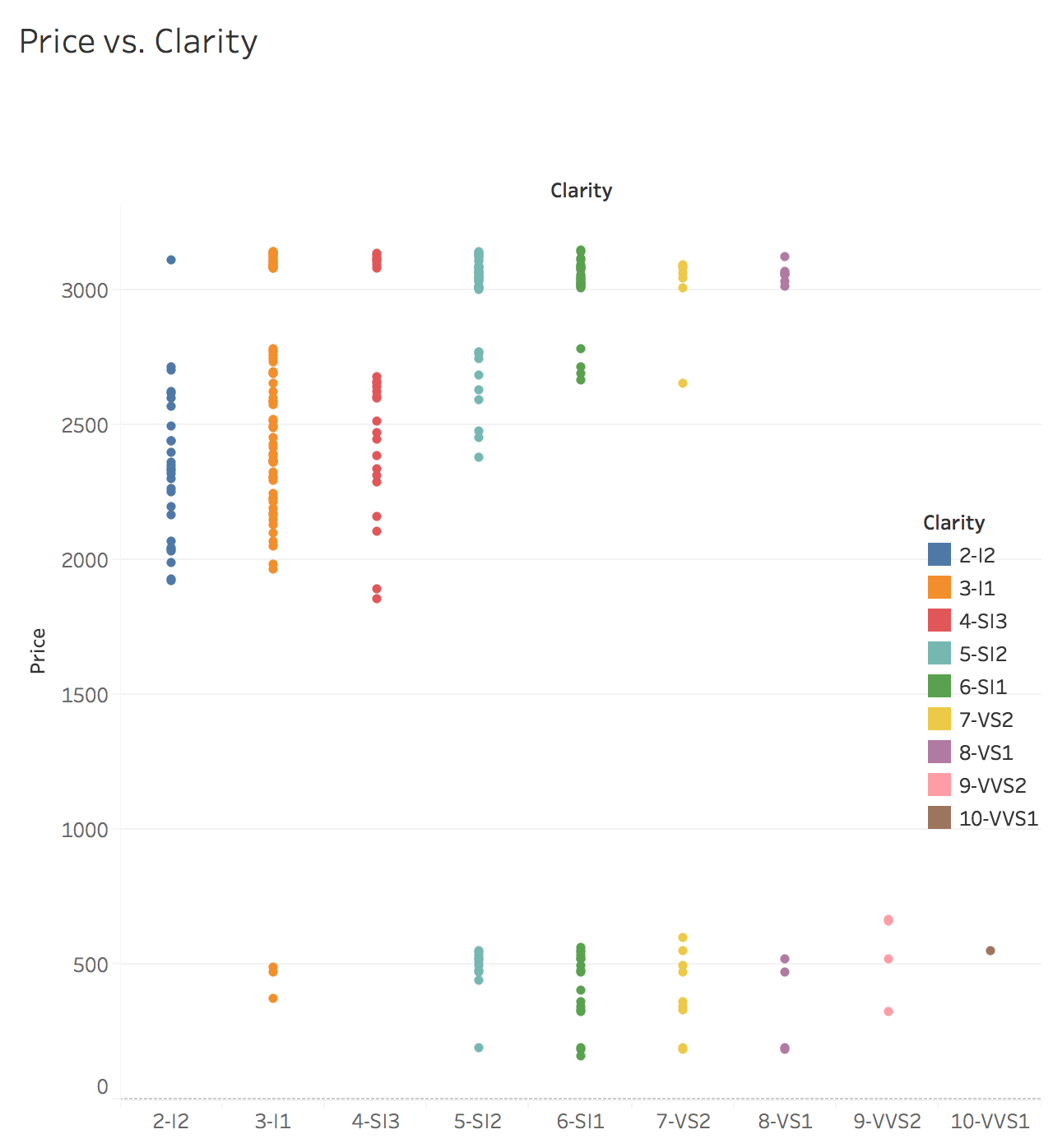
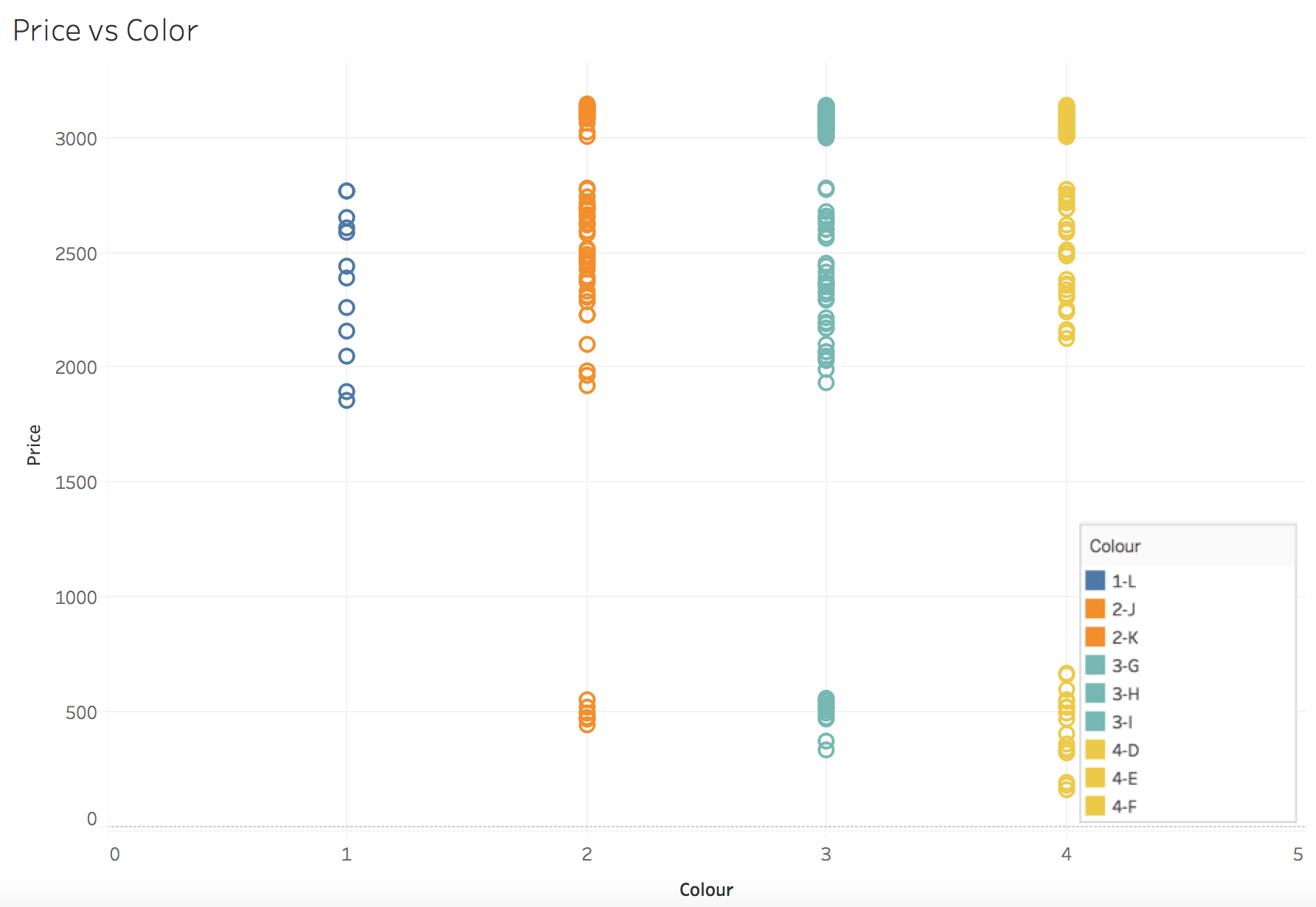
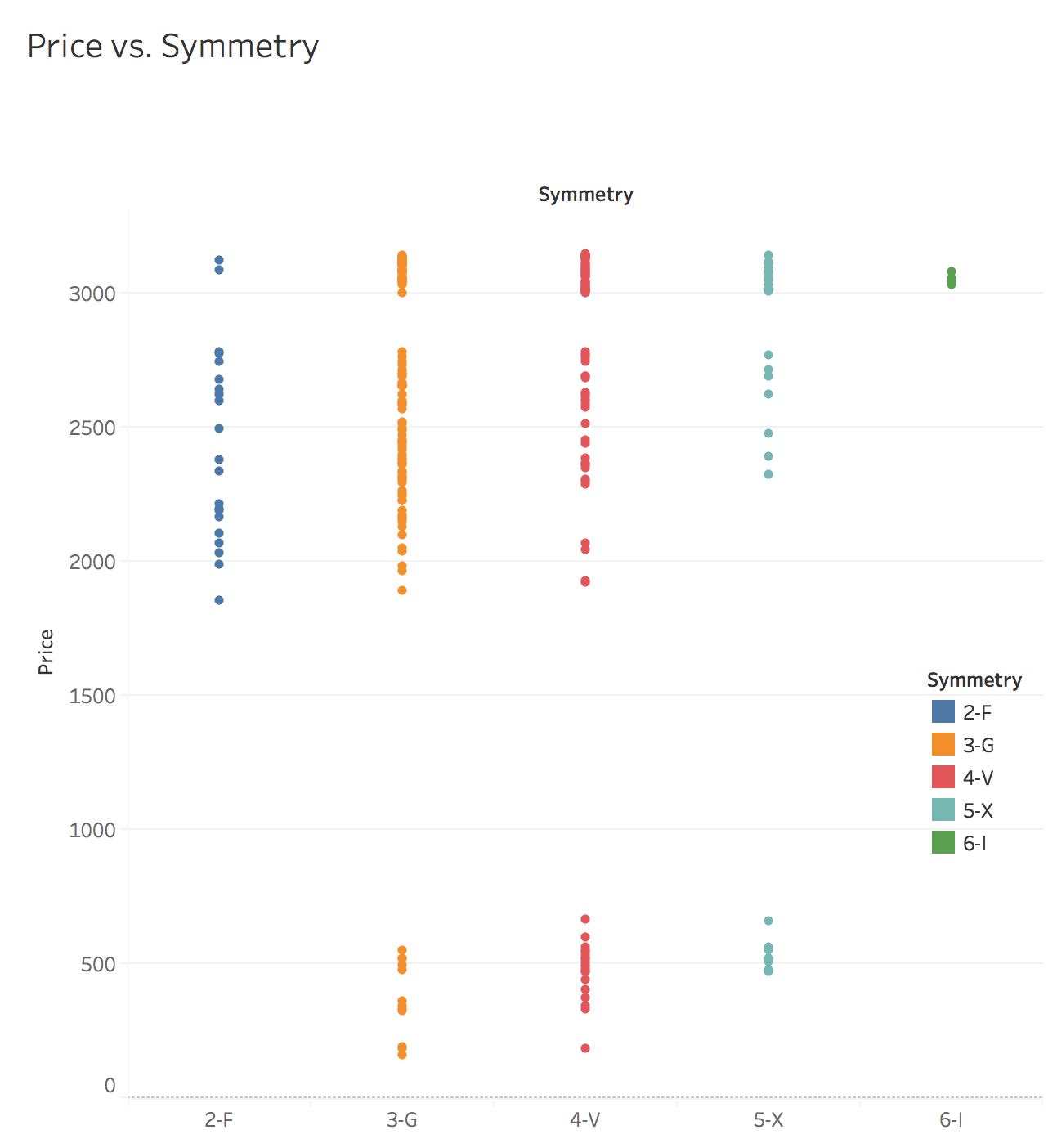
**DESCRIPTIVE ANALYSIS**

****

****

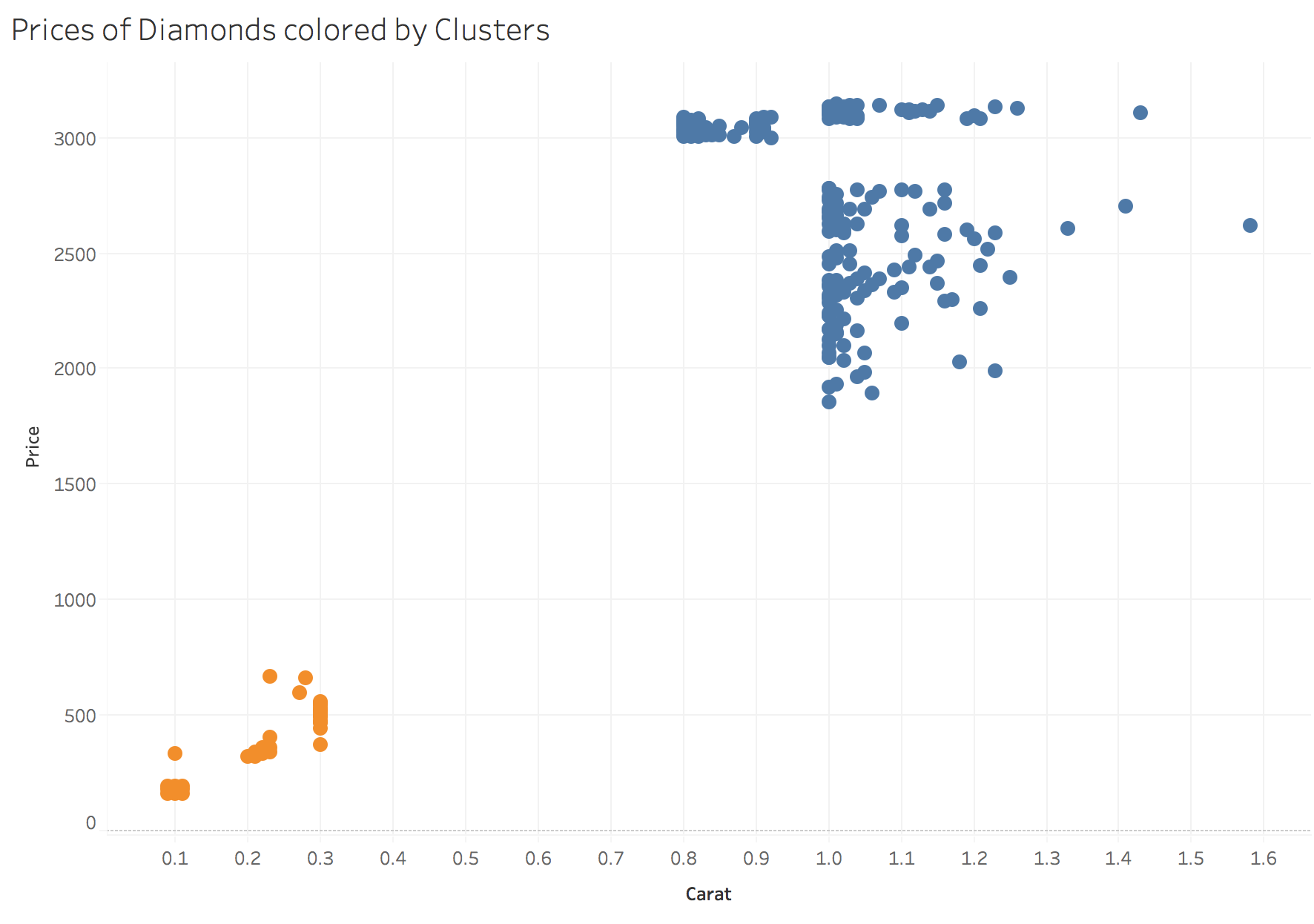
**BIVARIATE ANALYSIS**

****

** **

**MULTICOLLINEARITY**

**Rcorrplot.pdf**



DATA SELECTION FOR REGRESSION

When we plot carats vs prices we clearly see that the dataset is divided between two clusters. Then, two questions arise. Which cluster to analyze? or should the analysis be done for the overall dataset arises.

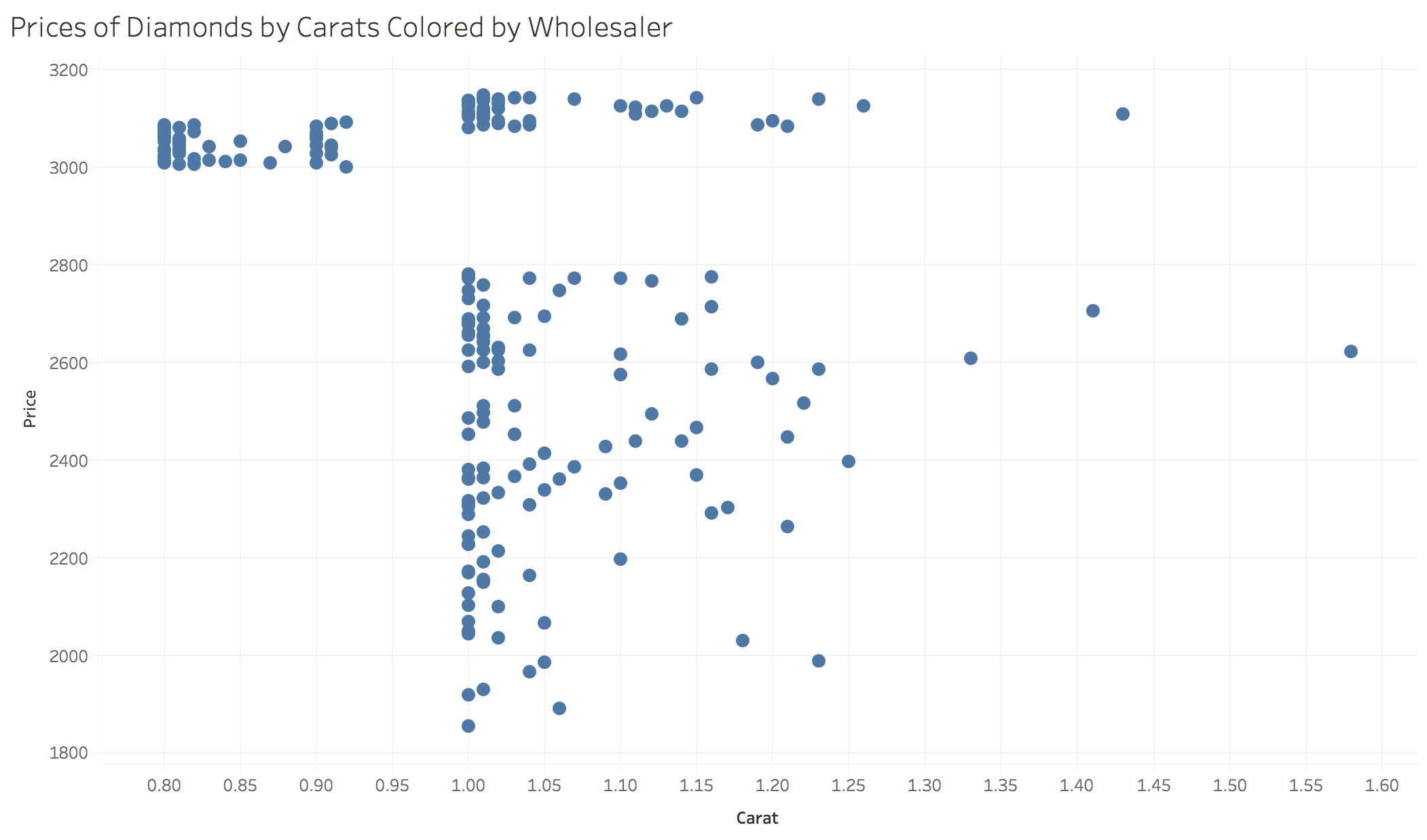
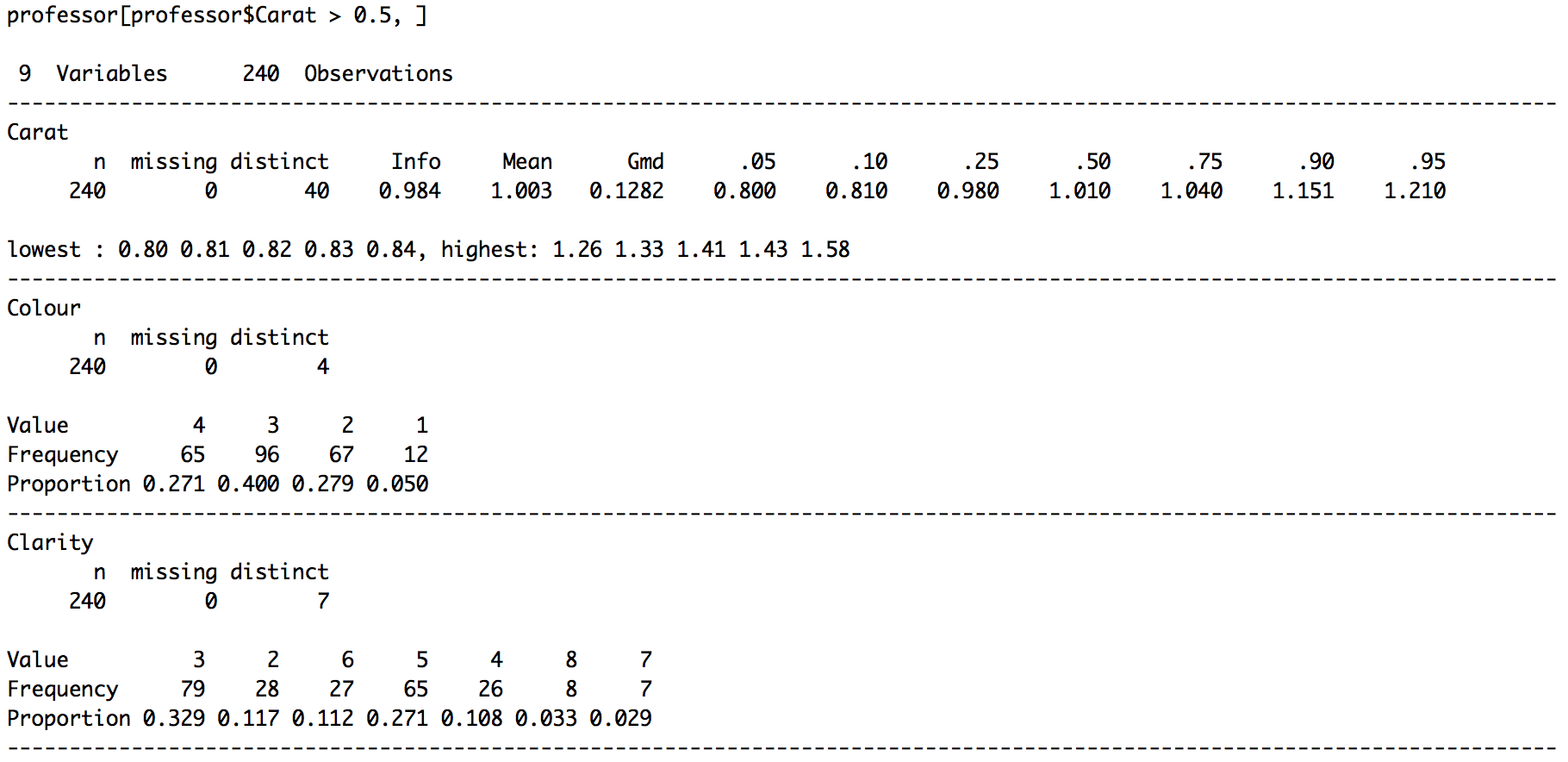
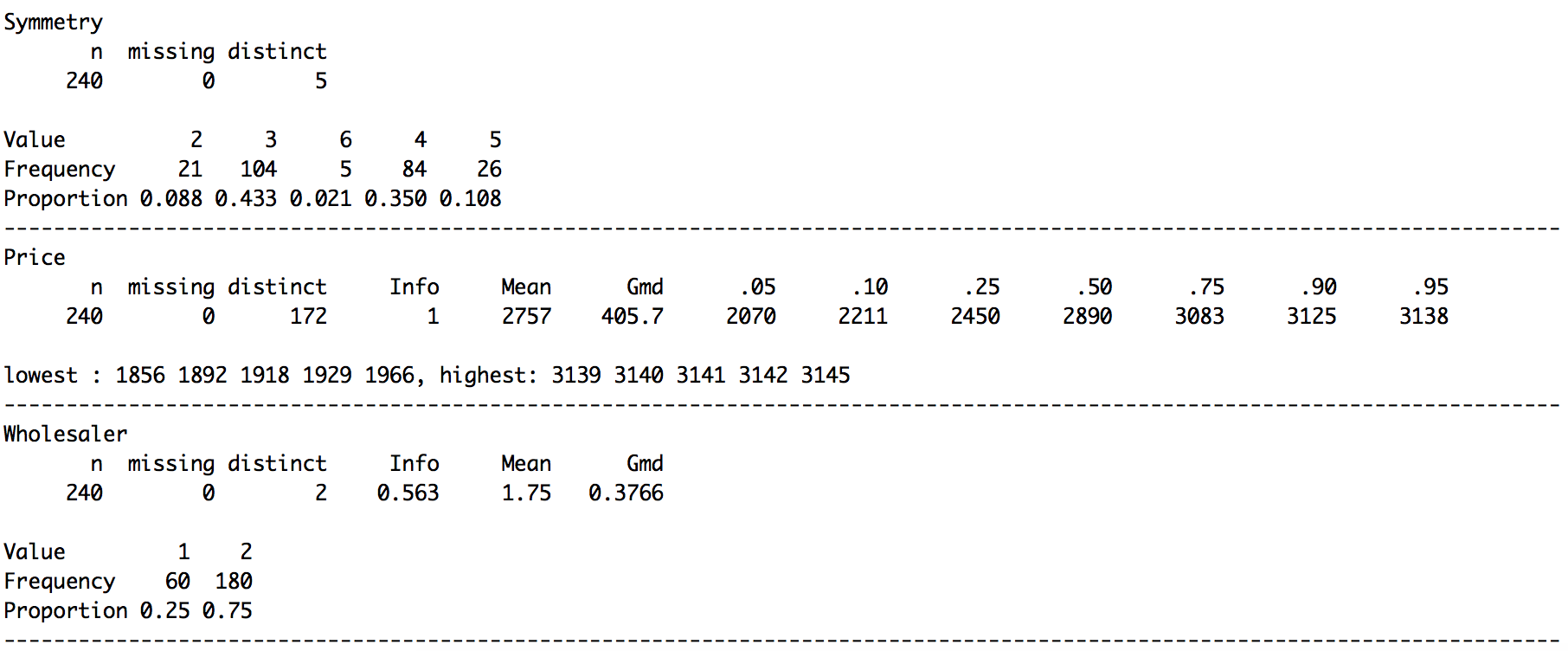
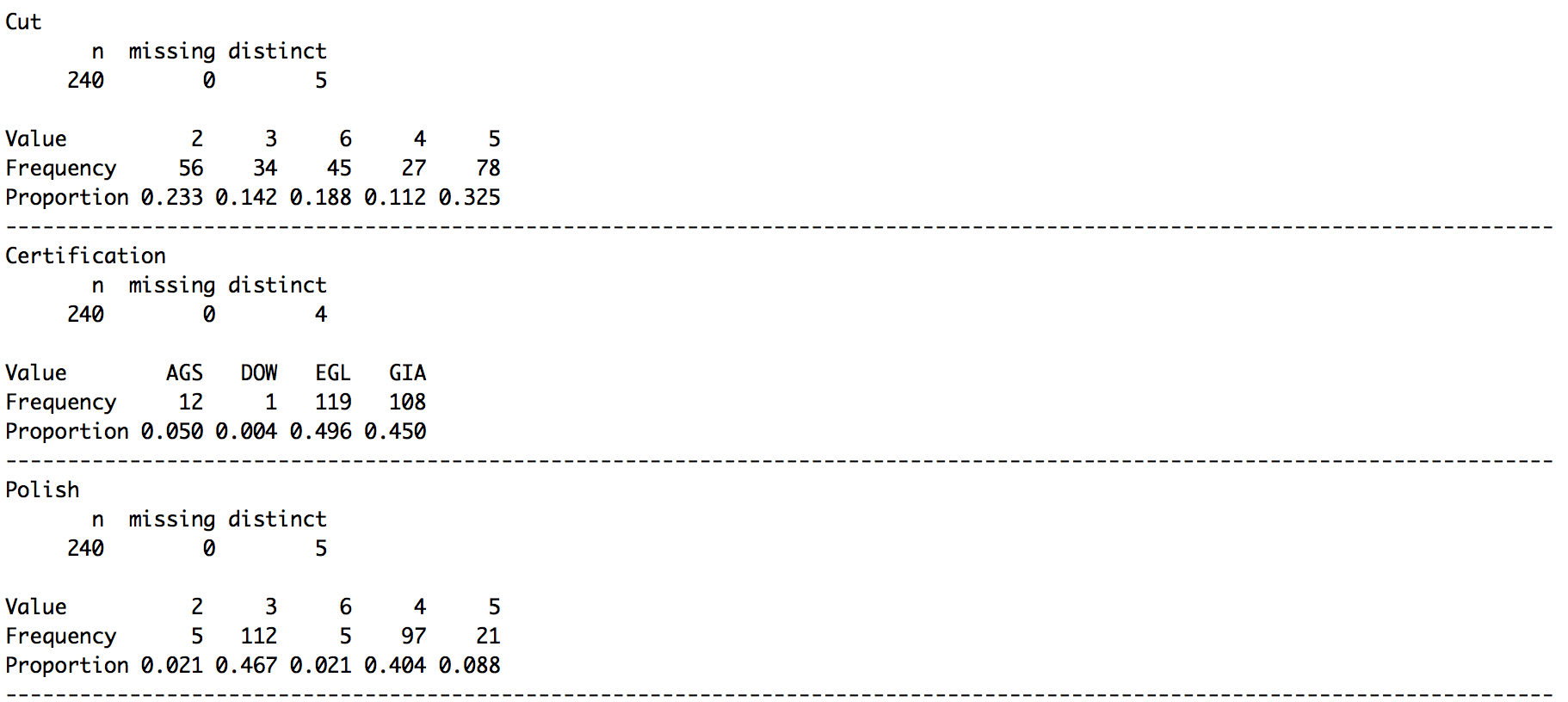
****

Figure: Cluster where professors ring belongs

**DESCRIPTIVE FOR THE CLUSTER**

**ANOVA**

**REGRESSION**

**Carat**

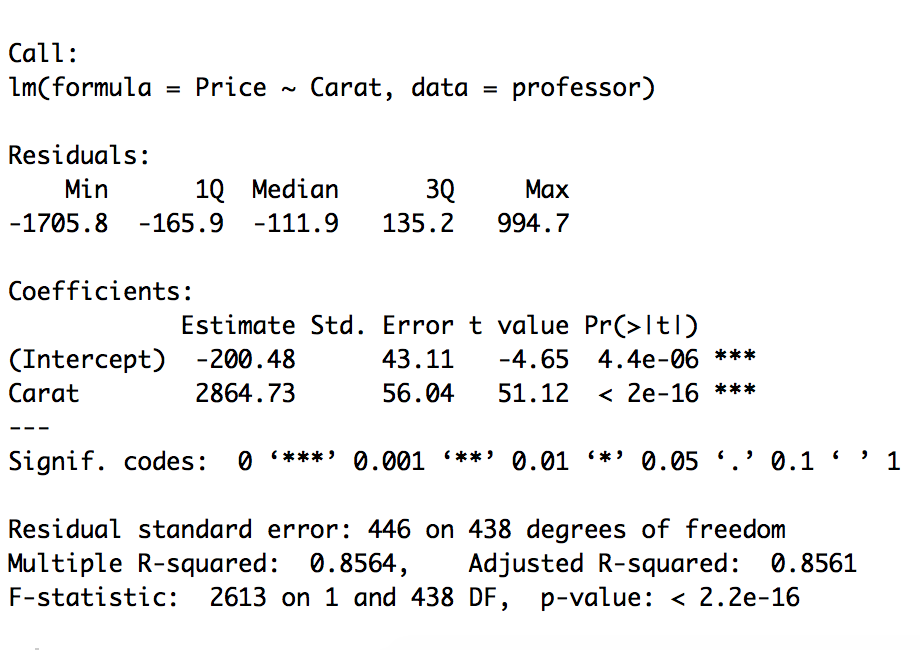
****

Figure: Carat vs Price 4 categories

**Colour**

Figure: Color vs Price 4 categories

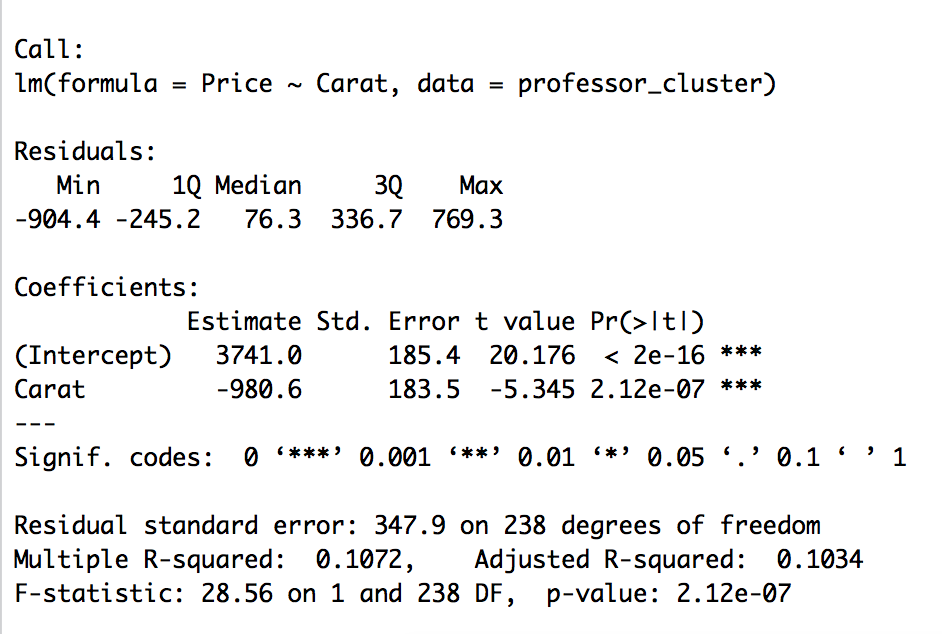


Figure: Colour vs Price 2 categories

**Clarity**

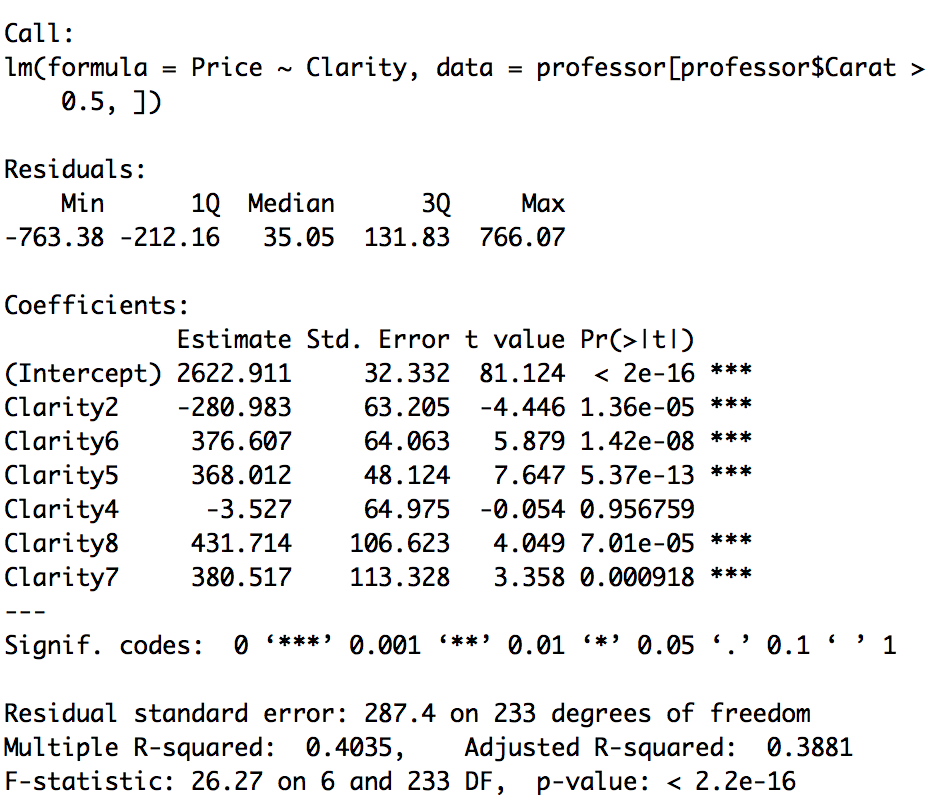
****

Figure: Price vs Clarity all levels

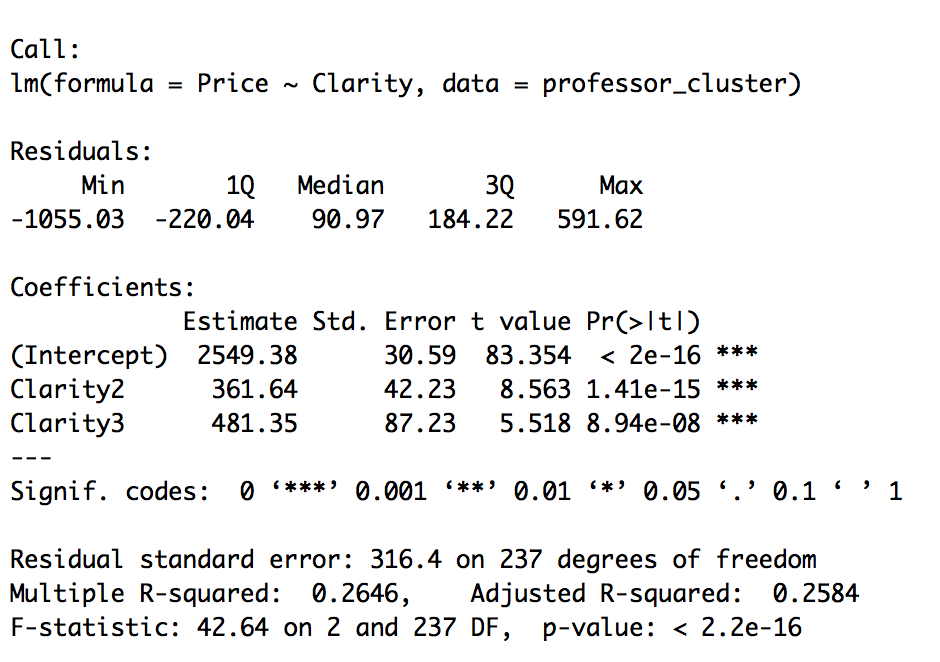
****

Figure: Price vs Clarity 3 levels

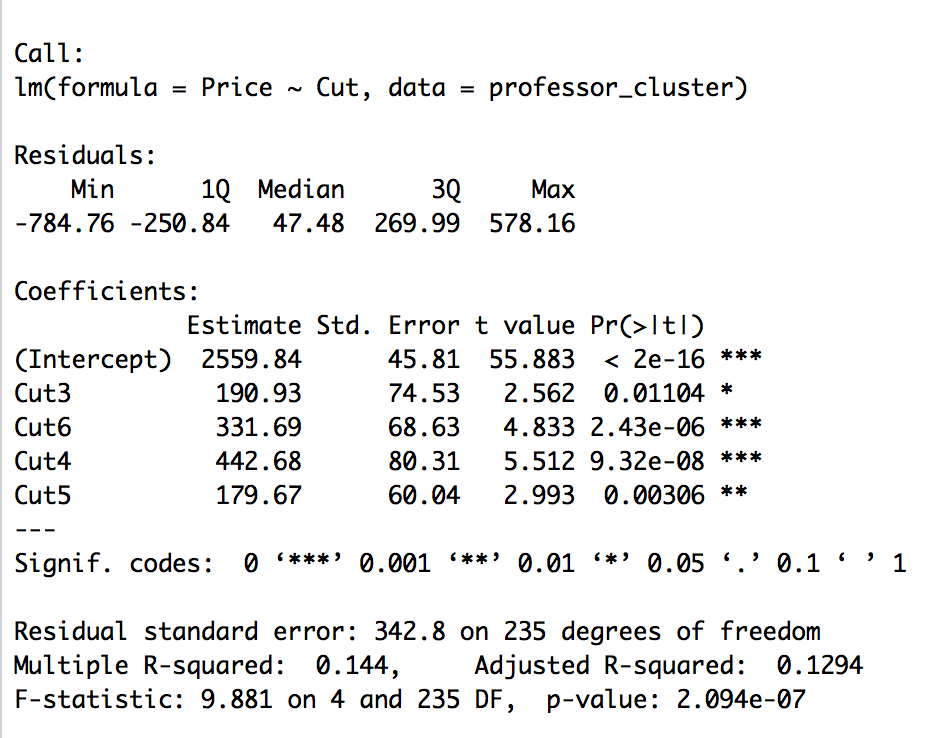
****

Figure: Price vs Cut

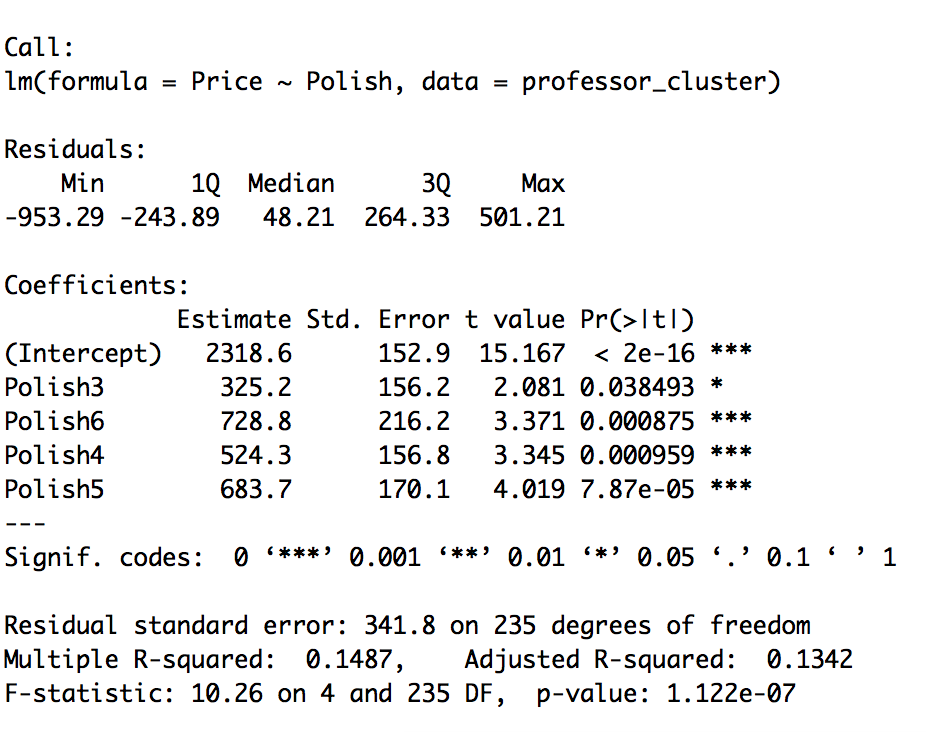
****

Figure: Price vs Polish

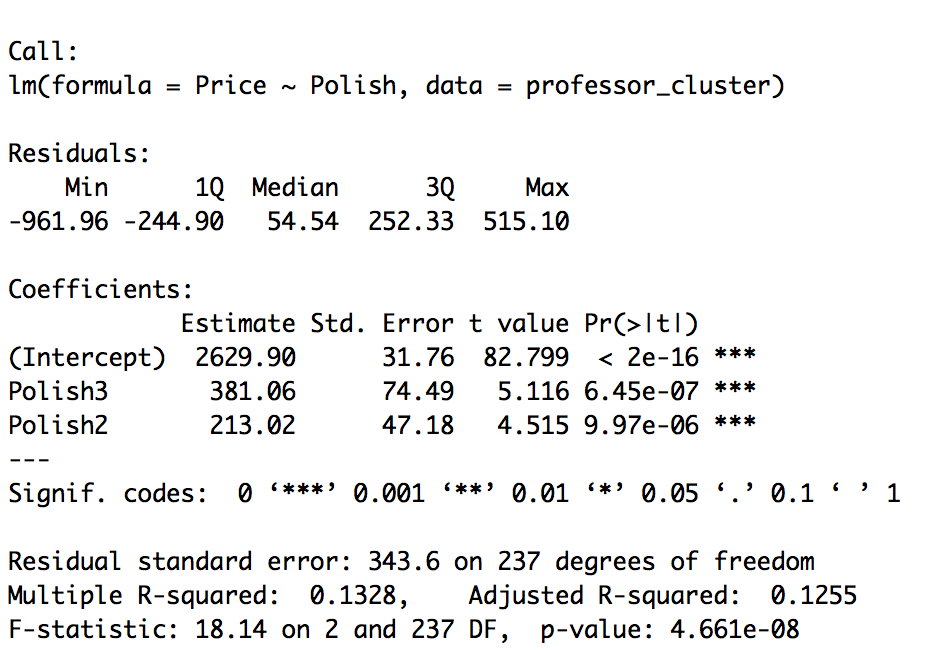


Figure: Price vs Polish 3 categories 2+3,4,5+6

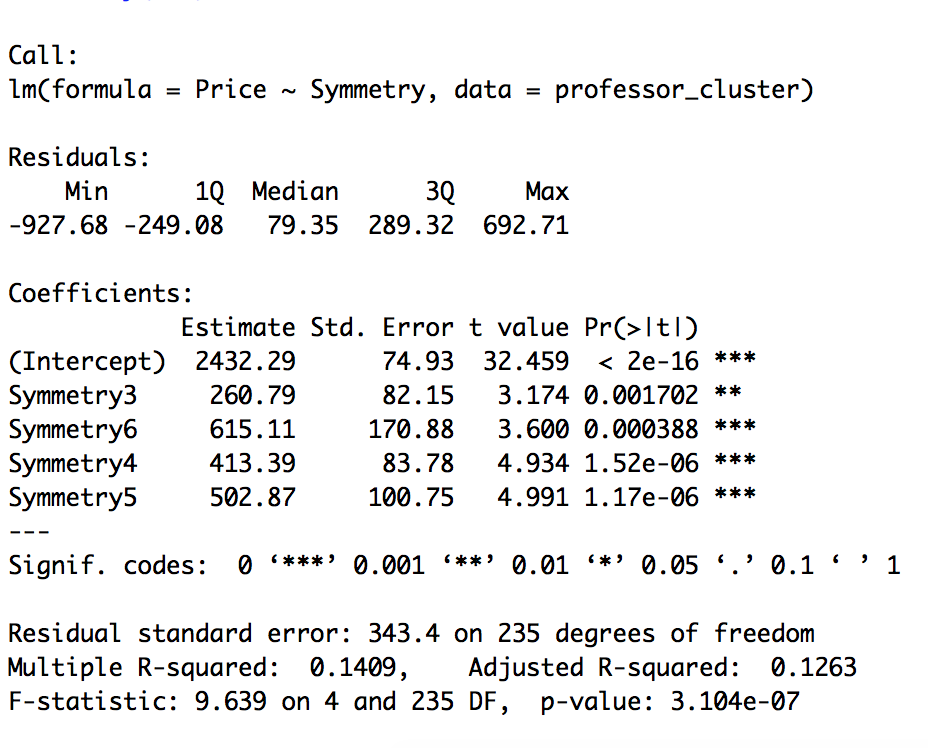


Figure: Price vs Symmetry

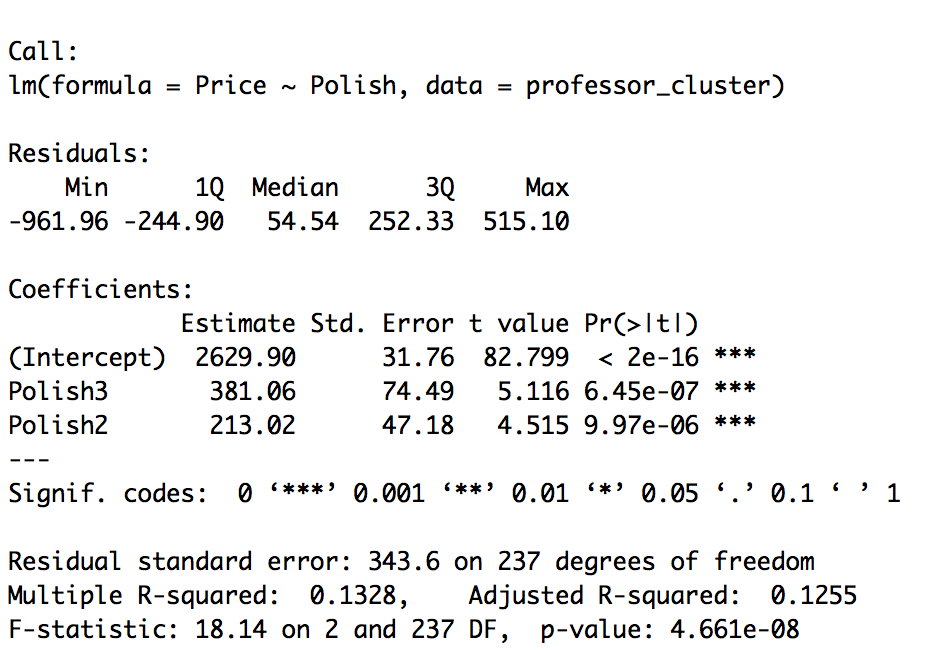
****

Figure: Price vs Symmetry 3 categories 2,3, 4+5+6