Sucoen Yeater

Assignment 1

Friday, September 11, 2020 9:37 PM

Requirements:

For the assignment, you are to conduct a deeper exploration of the diamonds data addressing the question "What factors affect the price of diamonds, and what is the nature of the relationship?"

Once you have done this exploration, select two different graphical displays that you believe provide insight into the relationship between price and one or more of the other variables in the data set.

For each graph:

1. Provide a copy of the graph.
2. Provide a detailed description of the relevant features of the graph.
3. Summarize your findings in a response to the question "What factors affect the price of diamonds, and what is the nature of the relationship?"

I wanted to start off with the graph, shown in figure 1, we ended the 9/2 class with since there was a very short discussion. Basically, the graphs that aren't clarity of I1 look messy. The I1 graph looks like it gives useful information. One thing to notice is that the Ideal cuts, for the most part, are higher price than other cuts that have the same carat. For the next graph, SI2, it looks like a majority of the Ideal cuts are above the other cuts, but there are some cases where an ideal cut is below the price of an 'inferior' cut of the same carat. This trend continues. As the clarity gets better, the cuts look to have a less overall effect on the price and some other variable(s) need to be accounted for. Let's add another variable to this by using facet_grid instead of facet_wrap, as shown on the next page.

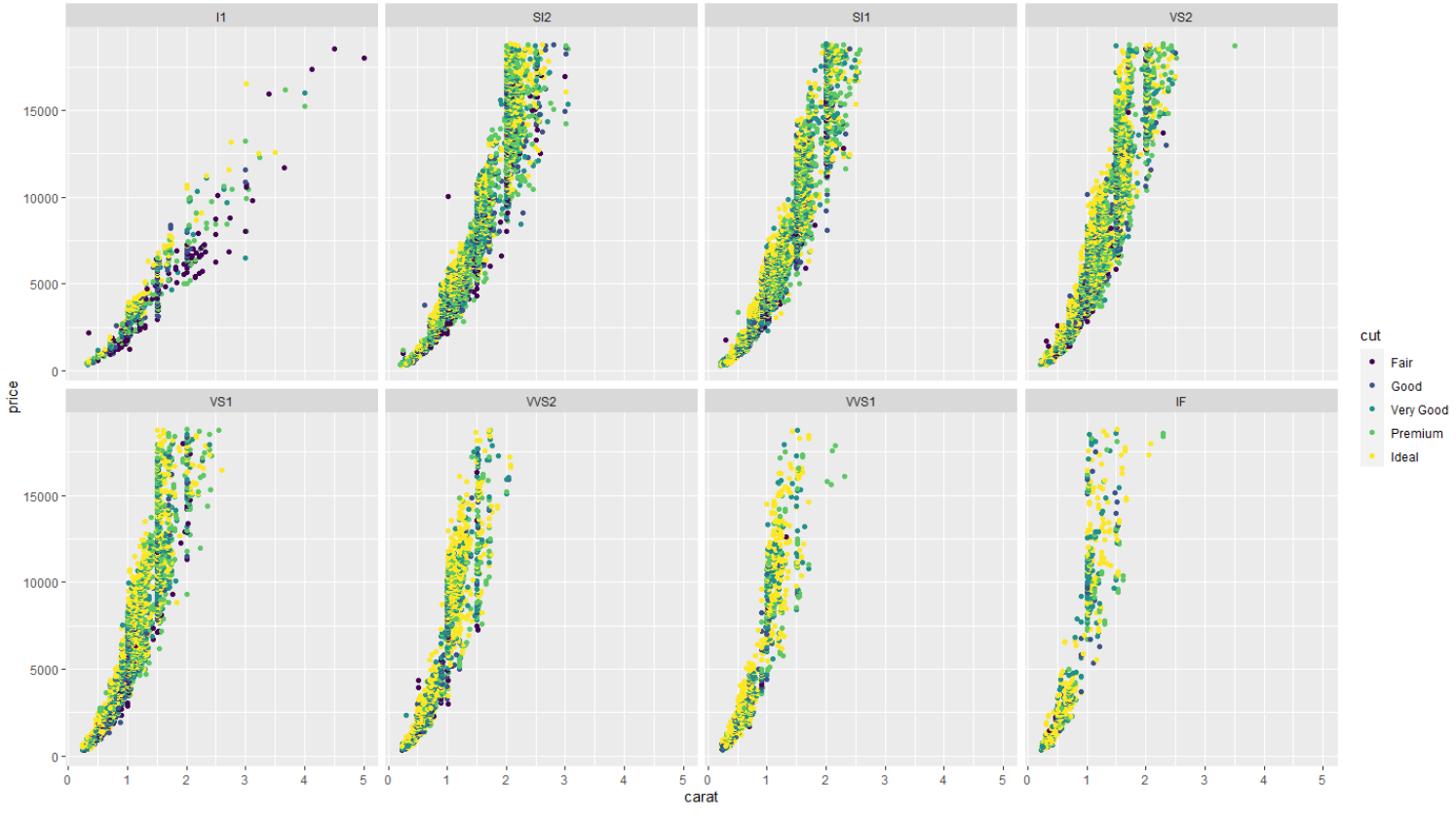


Figure 1: Four variables plotted: Carat, Price, Cut, Clarity

Sucoon Yeater

Assignment 1

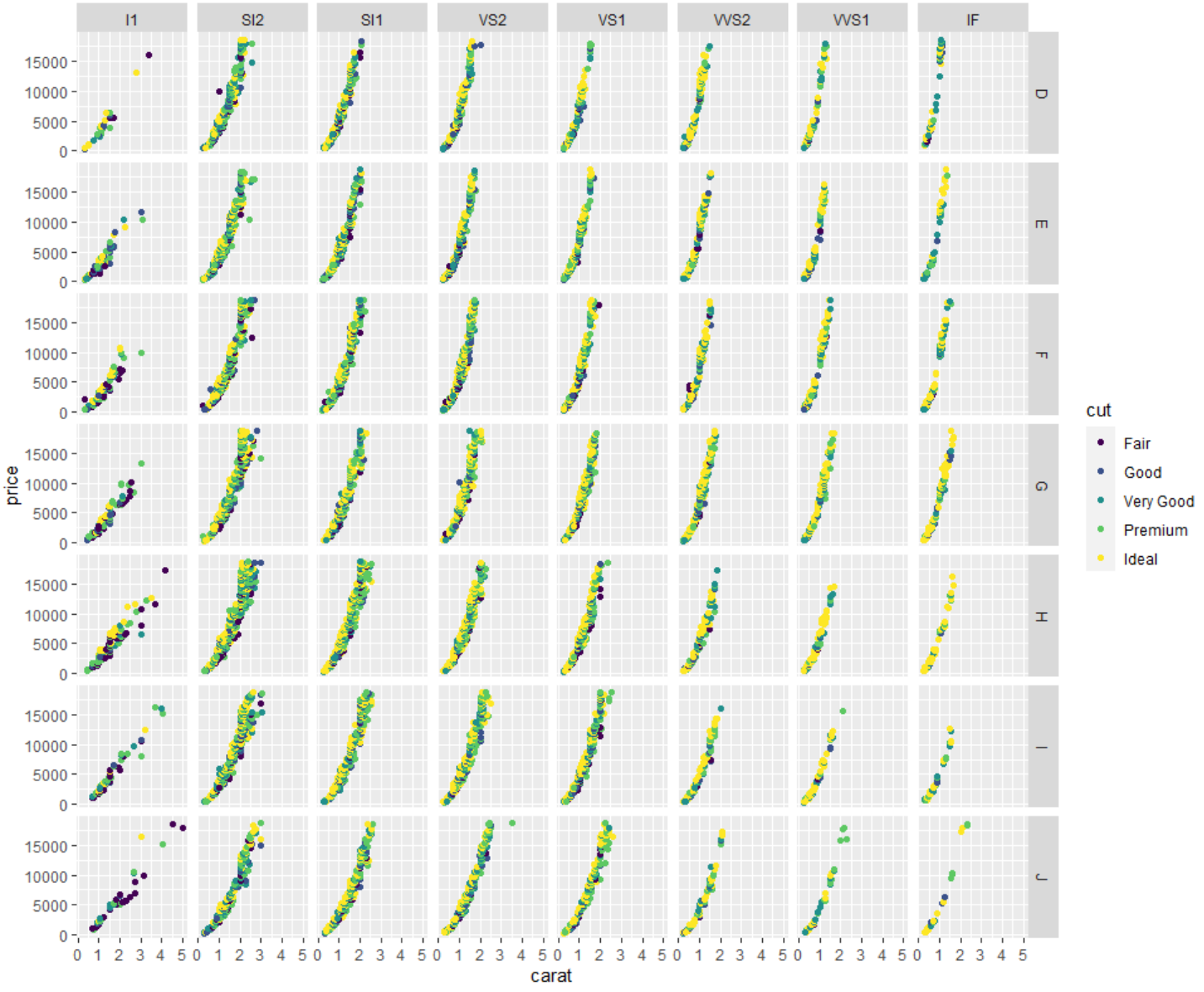
Friday, September 11, 2020 9:57 PM

Using the following code

```
d<-ggplot(data=diamonds)+
  geom_point(mapping=aes(x=carat, y=price, color=cut)) +
  facet_grid(color~clarity)
```

Results in the graph shown below in figure 2.

Each column represents the clarity rating where I1 is the worst and each row represents the color rating where D is the best. The data is a little more informative now that the color (D, E, F, G, H, I, J) has been included in the graph. This will only plot I1 and D on one graph and so on. Just looking at the first graph (I1, D) there are probably even more factors that will affect the price of a diamond. I say this because a fair cut 3.5ish diamond is more expensive than an ideal cut 2.8ish diamond. Also note that highest priced diamond with the best clarity, IF, the best color, D, does not have an ideal cut like one might first guess. So there has to be at least another variable that is effecting the price of the diamond. Looking into the help of the diamonds dataset shows that there are other factors that could affect the price such as length, width, depth, and total depth. For fun, let's see which variable most effects the price as shown on the next page.

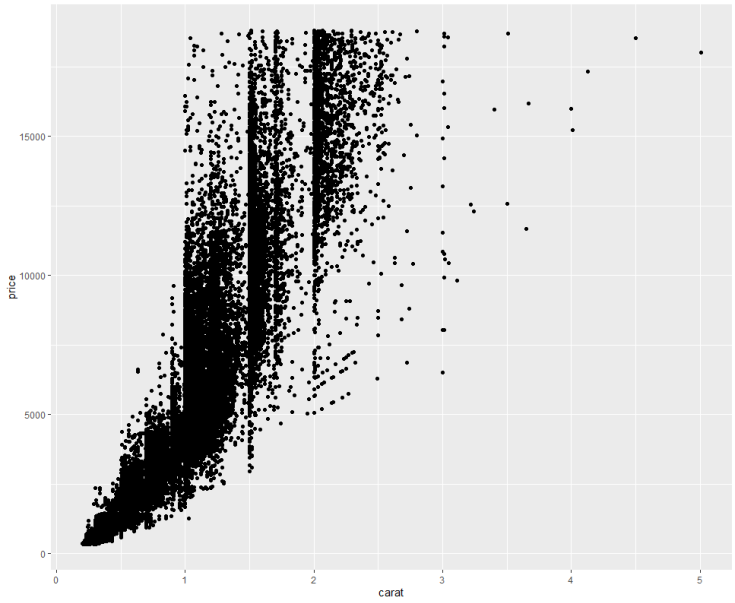


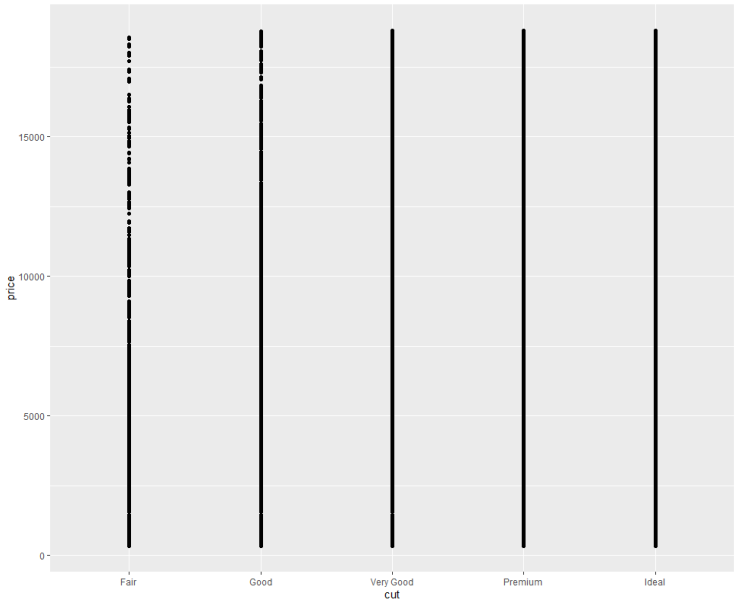
Sucoen Yeater

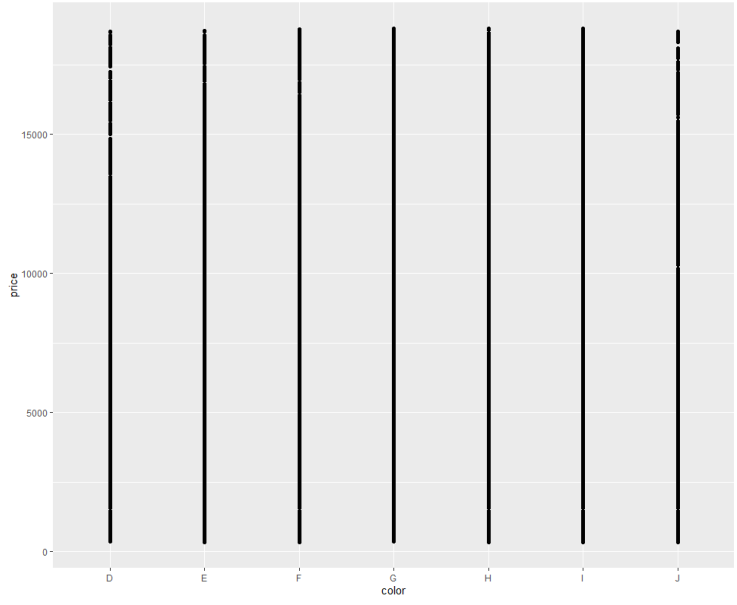
Assignment 1

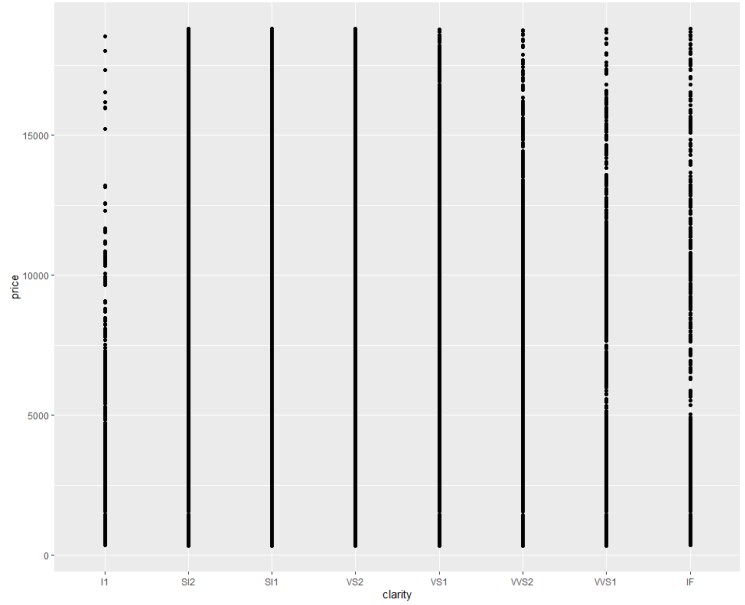
Friday, September 11, 2020 9:57 PM

As previously mentioned, there are multiple variables that can affect the price of a diamond. Let's look at each variable individually plotted vs. price to find out which variable most affects the price of a diamond.







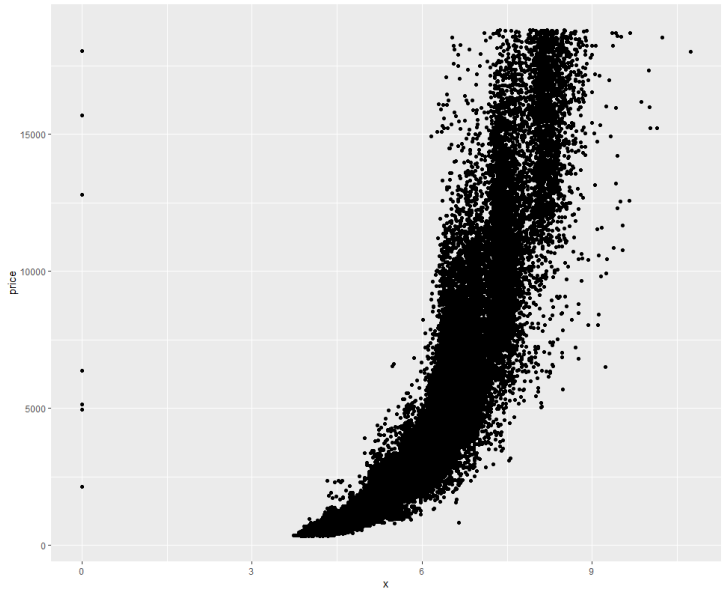


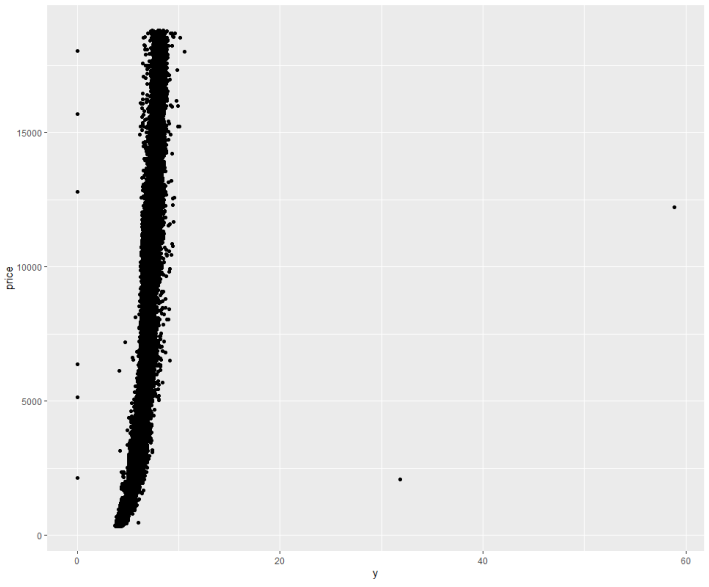
Three additional graphs as well as the final discussion on next page

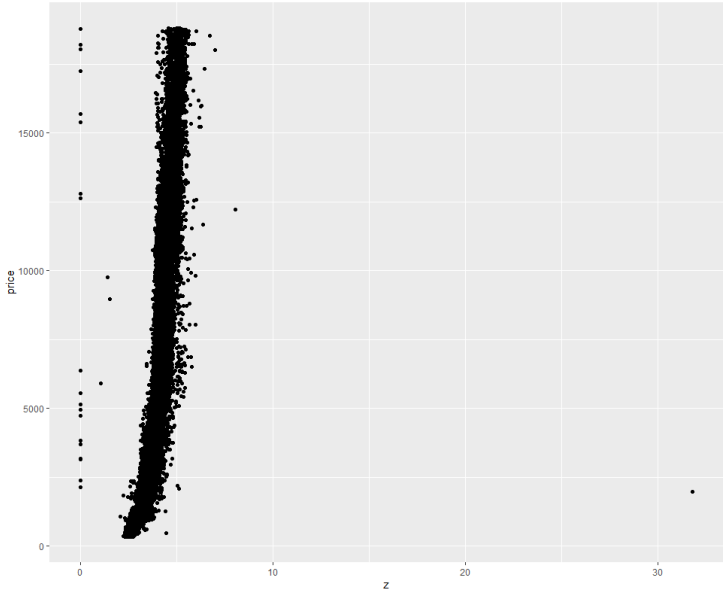
Sucoca Yeater

Assignment 1

Friday, September 11, 2020 9:57 PM







Starting with carat it can be seen that as the carat increases the price exponentially increases. There are points on the upper right portion of the graph that don't necessarily follow this trend. I believe these would be outliers and just more about the diamond needs to be known.

The cut, color, and clarity don't seem to have a direct impact on the price. The combination of them with the carat as well as the length, width, and depth of the diamond more so affects the price. There looks to be each type of cut, color, and clarity for most prices where the customer would then further customize with carat, length, width, and depth to get the final price of the diamond.

The three graphs shown above on this page are the length, width, and depth respectively. Each has a clear exponential trend. If any of those increase then there is an exponential increase in price. Also to note is the width and depth seem to have the exact same effect on price, which does make sense because they are similar variables in that the volume of the diamond will be the same whether there was an increase in depth or an equal increase in width.