Exploratory Data Analysis

Tabitha Hagen

2022-07-22

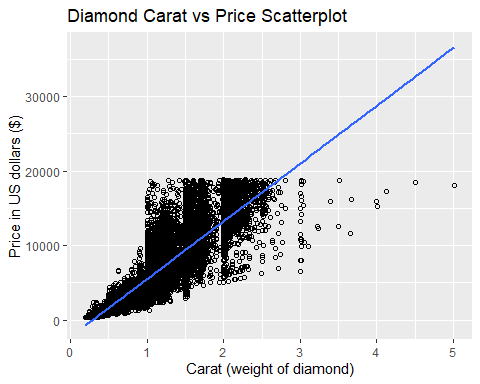
# Module 3 - Assignment 2

## Tabitha Hagen

### Exporatory Data Analysis

#### Diamond Color and Price

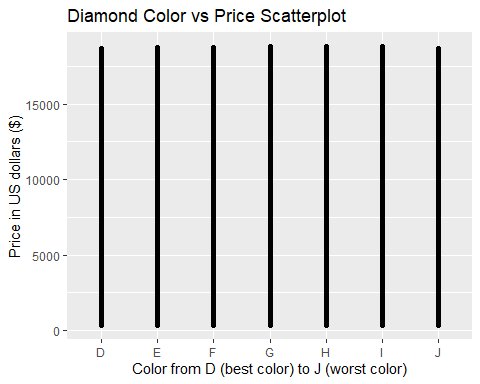
ggplot(data=diamonds, aes(x=carat, y=price)) +   
 geom\_point(shape=1) + # Creates a Scatterplot on the graph  
 geom\_smooth(method = "lm", formula = "y~x") + # Adds a fitted line to the graph  
 labs(title = "Diamond Carat vs Price Scatterplot", # Plot Title  
 x = "Carat (weight of diamond)", # x axis label  
 y = "Price in US dollars ($)" # y axis label  
 )



1.) What do you notice from the scatterplot as the carat size increases? There is a direct + correlation between the carat size & price as shown by the line. The more the diamond weighs, the higher the price is.

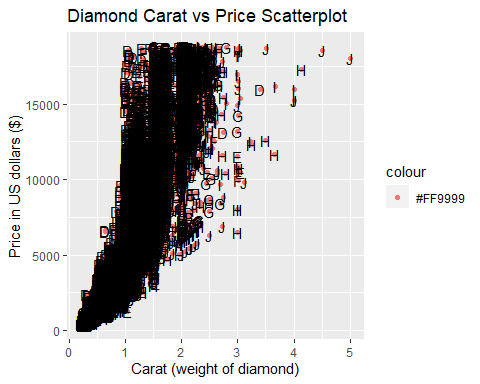
2.) From the scatterplot, what carats are most represented within the diamonds dataset? There appears to be more data with the Carats being between 0.5 and 2.0. The graph shows almost solid color between these values.

ggplot(data=diamonds, aes(x=color, y=price)) +   
 geom\_point() + # Creates a Scatterplot on the graph  
 labs(title = "Diamond Color vs Price Scatterplot", # Plot Title  
 x = "Color from D (best color) to J (worst color)", # x axis label  
 y = "Price in US dollars ($)" # y axis label  
 )



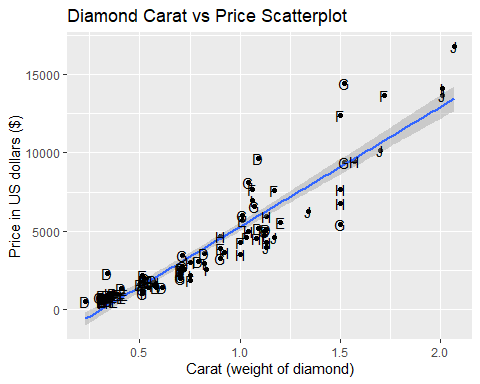
The scatterplot of Color vs. Price shows that Colors can have low or high price, spread evenly from the D (best color) to J (worst color). It is not particularly useful to compare these 2 characteristics.

ggplot(data=diamonds, aes(x=carat, y=price)) +   
 geom\_point(aes(colour="#FF9999")) + # Creates a Scatterplot on the graph  
 geom\_text(aes(label=color)) + # Labels points D to J on the color scale   
 labs(title = "Diamond Carat vs Price Scatterplot", # Plot Title  
 x = "Carat (weight of diamond)", # x axis label  
 y = "Price in US dollars ($)" # y axis label  
 )



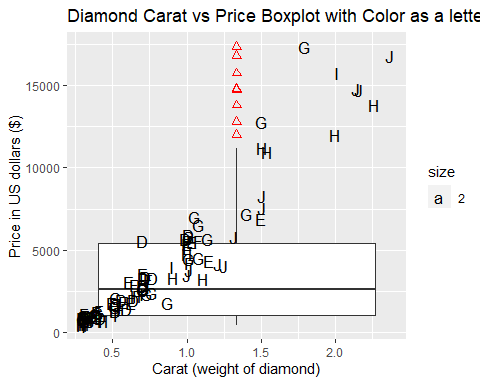
After adding the color codes to the Carat vs Price Scatterplot, it looks even messier than before. I do not see color affecting price. However, the larger diamonds are of color J or the worst color. Coincidentally, these too can be just as expensive as the Colors D-F.

dsample <- diamonds[sample(nrow(diamonds), 100), ] # take a sample of 100 items from Diamonds  
ggplot(data=dsample, aes(x=carat, y=price)) +   
 geom\_point() + # Creates a Scatterplot on the graph  
 geom\_smooth(method="lm", formula= "y~x" ) + # Adds a fitted line to the graph  
 geom\_text(aes(label=color)) + # Labels points D to J on the color scale   
 labs(title = "Diamond Carat vs Price Scatterplot", # Plot Title  
 x = "Carat (weight of diamond)", # x axis label  
 y = "Price in US dollars ($)" # y axis label  
 )

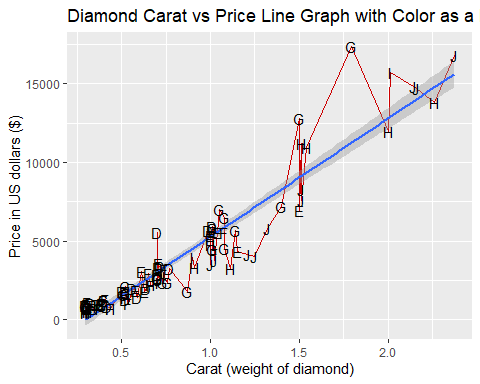


#### Using a Sample of the Diamonds Dataset

dsample <- diamonds[sample(nrow(diamonds), 100), ] # take a sample of 100 items from Diamonds  
  
ggplot(data=dsample, aes(x=carat, y=price, group=1)) +   
 #stat\_boxplot (geom = "errorbar", width = 0.25) + # add error bars to your whiskers  
 geom\_boxplot(outlier.colour="red", outlier.shape=2,  
 outlier.size=2, notch=FALSE) + # Creates a Boxplot on the graph w/ red outliers  
 geom\_text(aes(label=color, size=2)) + # Labels points D to J on the color scale   
 labs(title = "Diamond Carat vs Price Boxplot with Color as a letter D to J", # Plot Title  
 x = "Carat (weight of diamond)", # x axis label  
 y = "Price in US dollars ($)" # y axis label  
 )



ggplot(data=dsample, aes(x=carat, y=price)) +   
 geom\_line(aes(group=1), colour="#CC0000") + # Creates a line connecting data points  
 geom\_smooth(method="lm", formula= "y~x" ) + # Creates a fitted line   
 geom\_text(aes(label=color)) +  
 labs(title = "Diamond Carat vs Price Line Graph with Color as a letter D to J", # Plot Title  
 x = "Carat (weight of diamond)", # x axis label  
 y = "Price in US dollars ($)" # y axis label  
 )



#### Summary

This sample of the Diamonds dataset shows that the larger the carat, the higher the price. It also shows that the larger the carat, the less quality in color you get. Lastly, it shows that more smaller diamonds are bought than larger diamonds, probably because the color is not so good at the larger size or that smaller ones are simply more affordable.