# Module 3 - Assignment 2

# Subramanian, Subbu

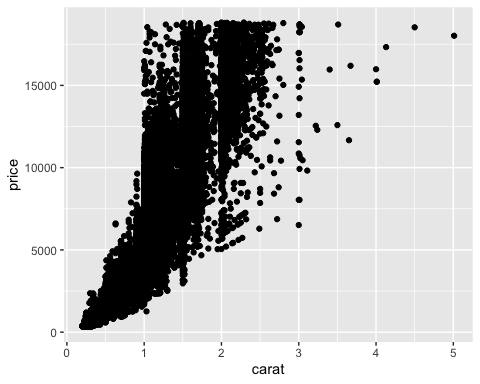
# Exploratory Data Analysis

Diamond Color and Price

library(tidyverse)

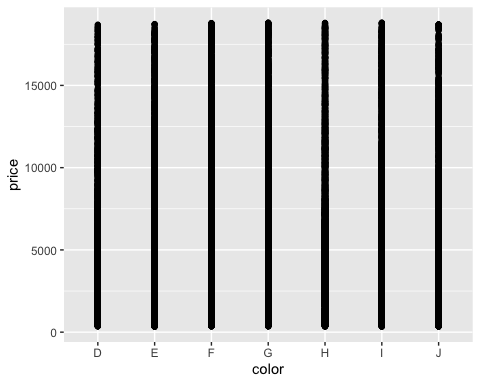
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

ggplot(data=diamonds,aes(x=carat,y=price))+  
 geom\_point()



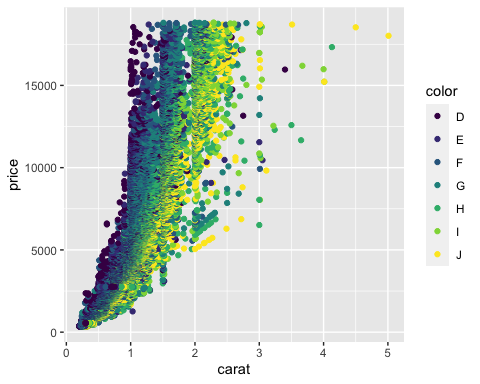
1.) What do you notice from the scatterplot as the carat size increases? Price Increases 2.) From the scatterplot, what carats are most represented within the diamonds dataset? 1 to 2

library(tidyverse)  
ggplot(data=diamonds,aes(x=color,y=price))+  
 geom\_point()



This plot isnt that useful as the price doesnt seem to vary much based on the color of the diamond.

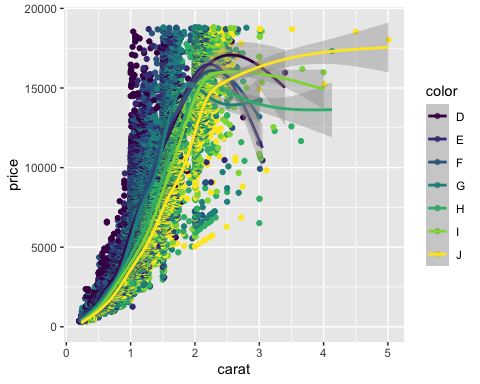
library(tidyverse)  
ggplot(data=diamonds,aes(x=carat,y=price,color=color))+  
 geom\_point()



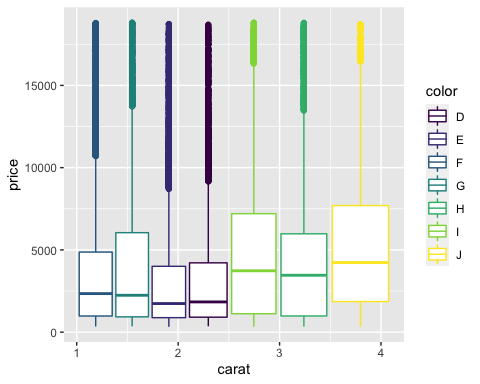
1.) Does color impact the price? Again not conclusive as the same set of colors appears to be on bothe low price and high price ends. 2.) Are certain colors associated with carat size? Provide an example. Yes, the larger the carat the color worsens to J

library(tidyverse)  
dsample<-diamonds[sample(nrow(diamonds), 100),]  
ggplot(data=diamonds,aes(x=carat,y=price,color=color))+  
 geom\_point()+  
 geom\_smooth()

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'



library(tidyverse)  
dsample<-diamonds[sample(nrow(diamonds), 100),]  
ggplot(data=diamonds,aes(x=carat,y=price,color=color))+  
 geom\_boxplot()



I think, when the carat is less the color of the diamond is very good, whereas when the carat size increases the color of the diamond detoriates.