# Module 1 - Refresher Assignment

## Sox, Morgan

#install.packages("tidyverse")  
library(tidyverse)

## ── Attaching packages ───────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.1.0 ✔ purrr 0.2.5  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.7  
## ✔ tidyr 0.8.2 ✔ stringr 1.3.1  
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## ── Conflicts ──────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

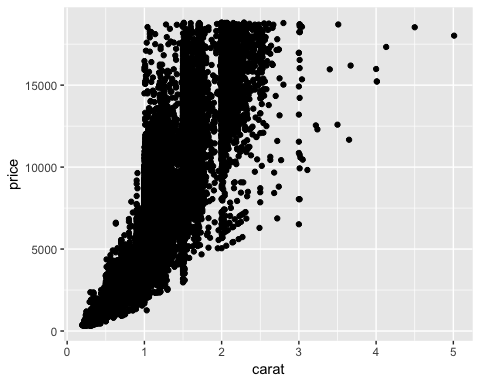
diamonddata = diamonds  
nrow(diamonddata)

## [1] 53940

ncol(diamonddata)

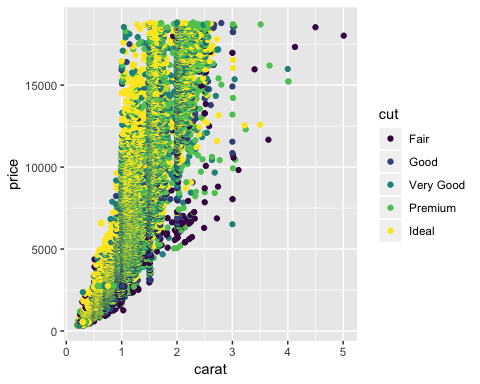
## [1] 10

ggplot(diamonds, aes(carat,price))+  
 geom\_point()



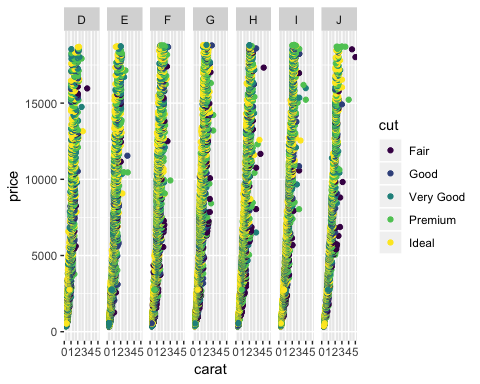
We notice as the size of the carat increases the price does increase in some cases, but we notice that the price is not completely dependent on the carat size. We should probably explore more variables to see if there is a better relationship to price than using carat.

ggplot(diamonds, aes(carat,price,color=cut))+  
 geom\_point()



It seems like cut isn’t very signifigant in price, but it is also very hard to see in the highly concentrated carats. We do notice that it appears that ideal cut and are concentrated at or below the 2 carat mark, but they have a wide range of prices.

Dia <- ggplot(diamonds, aes(carat,price,color=cut))+  
 geom\_point()  
  
Dia + facet\_grid(. ~ color)



Based on the graph above it is very hard to see a strong relationship between the 4 variables. There are many diamonds at the same price point from all cut catagories, color catagories, and many different sized carats. To look into the relationship more deeply taking a same of the population would be helpful.

inventory <- read\_csv("InventoryData.csv")

## Parsed with column specification:  
## cols(  
## `Item SKU` = col\_character(),  
## Store = col\_character(),  
## Supplier = col\_character(),  
## `Cost per Unit ($)` = col\_double(),  
## `On Hand` = col\_integer(),  
## `Annual Demand` = col\_integer()  
## )

inventoryA <-filter(inventory, Supplier == "A")  
inventoryA = mutate(inventoryA, OnHandRatio = `On Hand` / `Annual Demand`)

In the new inventoryA dataframe there are 3695 observations and 6 variables. The mutate line of code created a new variable which is the On Hand variable divided by the Annual Demand variable as a ratio.

avg\_cost <- inventory %>% group\_by(`Item SKU`) %>% summarise(SKUAvgCost = mean(`Cost per Unit ($)`))

Based on my previous experience in R the most challenging topics would be extracting or replacing subsets in dataframes. I also had a challenge in using the different aes functions on certain charts.