# Marci Copeland

# Module 1 - Assignment 3 - R and R studio Refresher

#install.packages("tidyverse")

library(tidyverse)

## -- Attaching packages ---------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

diamonddata = diamonds  
nrow(diamonddata)

## [1] 53940

ncol(diamonddata)

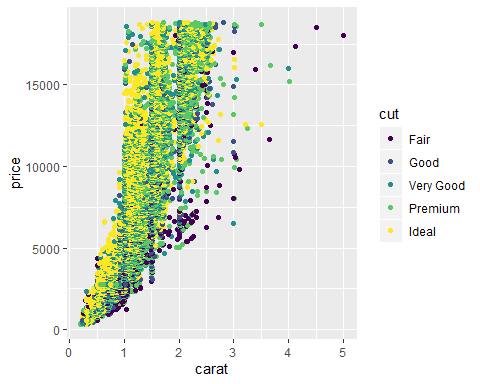
## [1] 10

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



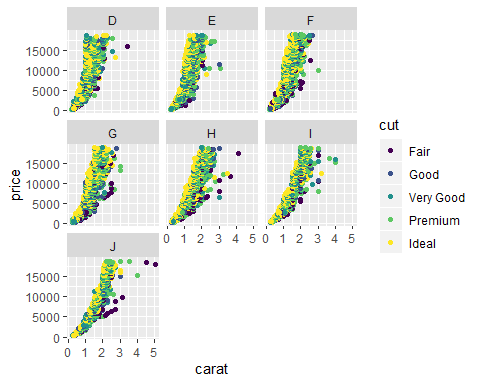
The relationship between carat and price in the diamonddata dataset shows as the carat size increases, the diamond price increases.

ggplot(diamonddata,aes(x=carat,y=price,color=cut)) +  
geom\_point()



The relationship between carat, price, and cut in the diamonddata dataset shows as the carat size increases, the diamond price increases, however it doesn’t appear cut has a direct relationship with price and or carat since the cut qualities are found at each price point and each carat size.

ggplot(diamonddata,aes(x=carat,y=price,color=cut)) +  
 facet\_wrap(~color) +  
 geom\_point()



Color and cut has very small impact on price, the plots show that carat has more impact on price than color or cut. Based on the scatterplot, certain colors are not associated with carat size except colors H, I, and J are the only colors that contain 4 and 5 carat diamonds and cut qualities are found at each price point and most carat sizes.

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

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

View(inventory)

inventoryA<-inventory %>%  
 filter(Supplier == "A")  
nrow(inventoryA)

## [1] 3695

inventoryA = mutate(inventoryA, OnHandRatio = `On Hand` / `Annual Demand`)

Mutate function in the code provided above is creating a new variable - OnHandRatio which is the calculation of On Hand divided by Annual demand.

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

All exercixes were a review for me and ponly required a review of textboook information and notes from the previous semester of taking MIS 503, however nothing was very challenging or difficult to complete utilizing my notes.