# Refresher Project - Module 1

## BAN 502 - Predictive Analytics

### Cooper, Sarah

### Task 1

# install.packages("tidyverse", dependencies = TRUE)  
# install.packages("zeallot")  
library(tidyverse)

library(zeallot)

### Task 2

data(diamonds)  
  
ncol(diamonds)

## [1] 10

nrow(diamonds)

## [1] 53940

## Task 3

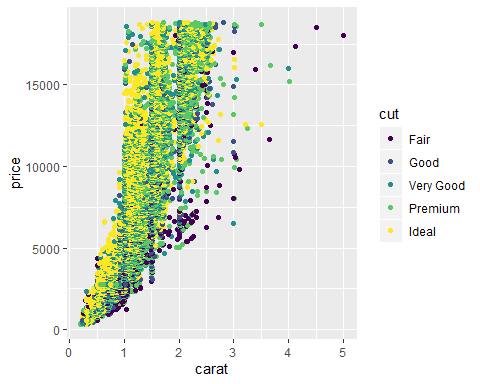
ggplot(diamonds, aes(x = carat, y = price)) +  
 geom\_point()



**The above graph is signaling that the larger the carat size, the higher the price.**

##Task 4

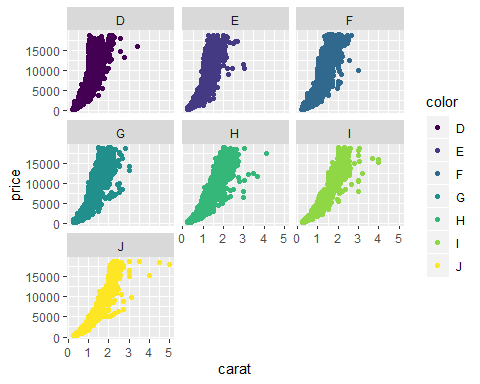
ggplot(diamonds, aes(x = carat, y = price, color = cut)) +  
 geom\_point()



**From the above graph we can see a correlation in an ideal cut diamond being smaller in size. With an increase in size, there is also an increase in price. There are several diamonds in a fair cut that are much larger in size but cheaper in price.**

## Task 5

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



**The above graphs are displaying a relationship between a more tinted diamond (scale H, I, and J) often being in larger size and less expensive, while the colorless diamonds (scale D, E, and F) are often smaller but more expensive.**

## Task 6

library(readr)  
InventoryData <- read\_csv("C:/Users/Sarah/Downloads/InventoryData.csv")

inventory <- InventoryData  
str(inventory)

summary(inventory)

## Item SKU Store Supplier Cost per Unit ($)  
## Length:13561 Length:13561 Length:13561 Min. : 0.0   
## Class :character Class :character Class :character 1st Qu.: 137.0   
## Mode :character Mode :character Mode :character Median : 377.5   
## Mean : 504.4   
## 3rd Qu.: 775.5   
## Max. :1982.3   
## On Hand Annual Demand   
## Min. : 0.0 Min. : 0.0   
## 1st Qu.: 50.0 1st Qu.: 483.0   
## Median :101.0 Median : 965.0   
## Mean :100.5 Mean : 966.2   
## 3rd Qu.:151.0 3rd Qu.:1448.0   
## Max. :200.0 Max. :2150.0

## Task 7

inventoryA = inventory%>% filter(Supplier=="A")  
ncol(inventoryA)

## [1] 6

nrow(inventoryA)

## [1] 3695

## Task 8

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

## Task 9

avg\_cost = inventoryA%>% group\_by(`Item SKU`)%>%  
 summarise(SKUAvgCost =mean(`Cost per Unit ($)`, na.rm = TRUE))  
head(avg\_cost)

## # A tibble: 6 x 2  
## `Item SKU` SKUAvgCost  
## <chr> <dbl>  
## 1 0100 125.   
## 2 011 12.3  
## 3 0113 197.   
## 4 0122 81.7  
## 5 013 14.3  
## 6 0133 203.

## Task 10

**The most common challenge I had in MIS 503 was being unclear on the format of formulas. What made the challenge harder was the specificity of the formula (one too many comma’s or parenthesis just won’t do). The error’s that R throws back at you during these times is difficult to interpret, only adding frustration and time to the assignment. There is also so many ways to accomplish the same task. Often times I found the R for Data Science book helpful in the understanding but not in the execution. In my last class I leaned heavily on Google and public forums for troubleshooting.**