# BAN 502

## Assignment 2 – R and RStudio Refresher

### Jessica Sisco

#install.packages(tidyverse)  
  
library(tidyverse)

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

## v ggplot2 3.1.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.8  
## v tidyr 0.8.2 v stringr 1.3.1  
## v readr 1.2.1 v forcats 0.3.0

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

## Diamond Data

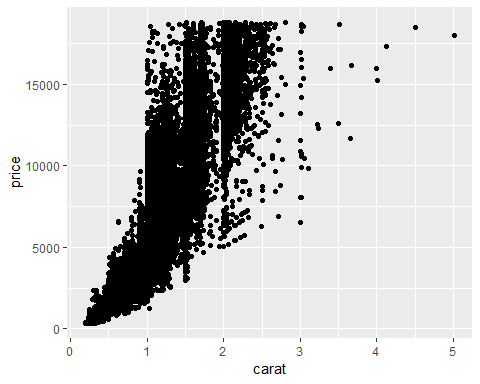
library(ggplot2)  
  
select(diamonds)

## # A tibble: 53,940 x 0

diamonddata = diamonds  
str(diamonddata)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 53940 obs. of 10 variables:  
## $ carat : num 0.23 0.21 0.23 0.29 0.31 0.24 0.24 0.26 0.22 0.23 ...  
## $ cut : Ord.factor w/ 5 levels "Fair"<"Good"<..: 5 4 2 4 2 3 3 3 1 3 ...  
## $ color : Ord.factor w/ 7 levels "D"<"E"<"F"<"G"<..: 2 2 2 6 7 7 6 5 2 5 ...  
## $ clarity: Ord.factor w/ 8 levels "I1"<"SI2"<"SI1"<..: 2 3 5 4 2 6 7 3 4 5 ...  
## $ depth : num 61.5 59.8 56.9 62.4 63.3 62.8 62.3 61.9 65.1 59.4 ...  
## $ table : num 55 61 65 58 58 57 57 55 61 61 ...  
## $ price : int 326 326 327 334 335 336 336 337 337 338 ...  
## $ x : num 3.95 3.89 4.05 4.2 4.34 3.94 3.95 4.07 3.87 4 ...  
## $ y : num 3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 ...  
## $ z : num 2.43 2.31 2.31 2.63 2.75 2.48 2.47 2.53 2.49 2.39 ...

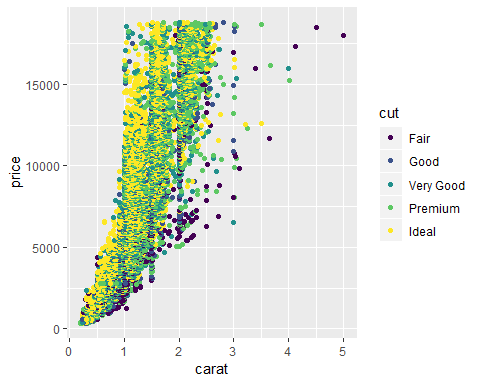
ggplot(data=diamonddata, mapping=aes(x=carat, y=price))+   
 geom\_point()

 Based on the above scatter plot as the carat size increases so does the price.

select(diamonddata)

## # A tibble: 53,940 x 0

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

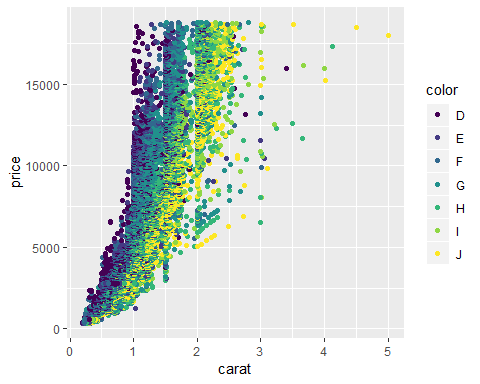


By adding in the factor of cut to the scatter plot the ideal cut seems to be more predominate as the carat size and price rise.

select(diamonddata)

## # A tibble: 53,940 x 0

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

 By adding in the factor of color to the scatter plot, as the carat size increases and the color quality increases so does the price.

## Inventory Data

library(readr)  
InventoryData <- 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(InventoryData)  
  
inventory=InventoryData

select(inventory)

## # A tibble: 13,561 x 0

inventoryA=filter(inventory,Supplier=='A')  
  
  
nrow(inventoryA)

## [1] 3695

The number of suppliers in the data set is 3695

select(inventoryA)

## # A tibble: 3,695 x 0

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

The about r Task 8 used mutate to create a new variable using existing variables in a ration.

select(inventory)

## # A tibble: 13,561 x 0

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

**Task 10: Given your previous course experience with R/R Studio, what topics/concepts did you find to be most challenging?** \* I am still new to R/R studio and I currently do not use it in my everyday job duties. This refresher has been great to help me remember some of the concepts from last semester. I am constantly learning, so I am open to learning any and all concepts. \*