# R Refresher Assignment

## Dan Sanchez

## BAN 502

## Dr. Stephen Hill

## May 18th, 2020

### Task 1

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

## -- Attaching packages ------------------------------------------ tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.4  
## v tibble 3.0.1 v dplyr 0.8.5  
## v tidyr 1.0.3 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

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

### Task 2

diamonddata = diamonds  
glimpse(diamonddata)

## Rows: 53,940  
## Columns: 10  
## $ carat <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23,...  
## $ cut <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, ...  
## $ color <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J,...  
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS...  
## $ depth <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4,...  
## $ table <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62,...  
## $ price <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340,...  
## $ x <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00,...  
## $ y <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05,...  
## $ z <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39,...

### Task 3

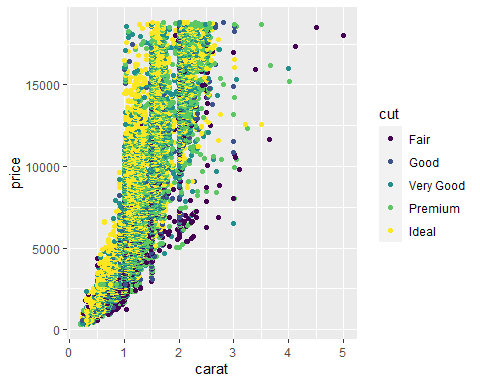
ggplot(diamonddata,aes(x=carat,y=price))+  
 geom\_point()



This scatterplot shows that diamond price tends to increase along with carat size. The majority of diamonds with less than 1 carat are $5,000 or less. However, there is much variability in the price of diamonds between 1 and 3 carats. This implies that, for diamonds in this 1-3 carat range, factors other than carat size tend to drive the price. Carat size beyond 4 tends to align with a high price.

### Task 4

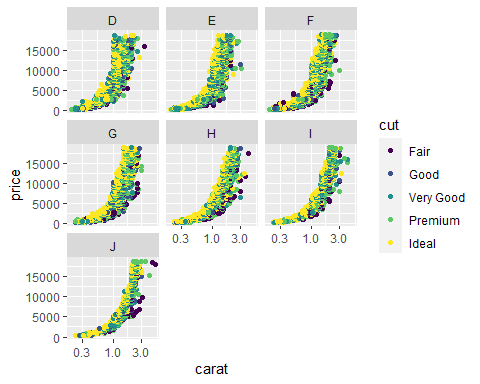
ggplot(diamonddata,aes(x=carat,y=price,color=cut))+  
 geom\_point()



The cut of diamonds plays an important role in determining price for diamonds under 2 carats. Most noticebly, ideal cuts of diamond seem to be more expensive in that range of carat size. Fair cuts tend to have the lowest price for the small carat sizes. Cut is not as indicative of price beyond the 2-carat size. Ideal cuts are also less represented in this data set from 3 carats and beyond. 4- and 5-carat diamonds tend to have a higher price, regardless of cut.

### Task 5

ggplot(diamonddata,aes(x=carat,y=price,color=cut))+  
 geom\_point()+  
 scale\_x\_log10()+  
 facet\_wrap("color")



When moving up in carat size, the price of D, E, F, and G-color diamonds tends to rise earlier than it does for the H, I, and J colors. While D, E, F, and G have a large cluster of diamonds between 1.0 and 2.5 carats, H, I, and J-color diamonds has a more concentrated distribution between 2.0 and 3.0.

### Task 6

library(readr)  
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\_double(),  
## `Annual Demand` = col\_double()  
## )

str(inventory)

## tibble [13,561 x 6] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Item SKU : chr [1:13561] "0100" "0100" "0100" "0100" ...  
## $ Store : chr [1:13561] "003480" "01611" "01611" "020109" ...  
## $ Supplier : chr [1:13561] "A" "B" "D" "B" ...  
## $ Cost per Unit ($): num [1:13561] 125.32 115.12 53.61 2.26 60.51 ...  
## $ On Hand : num [1:13561] 159 40 174 176 74 48 6 129 82 17 ...  
## $ Annual Demand : num [1:13561] 1693 351 1691 1559 733 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. `Item SKU` = col\_character(),  
## .. Store = col\_character(),  
## .. Supplier = col\_character(),  
## .. `Cost per Unit ($)` = col\_double(),  
## .. `On Hand` = col\_double(),  
## .. `Annual Demand` = col\_double()  
## .. )

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")

There are 3,695 rows in inventoryA

### Task 8

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

The above line of code will create the “OnHandRatio” variable in a new column in the dataframe inventoryA. The result will show the calculation of the OnHand variable divided by Annual Demand.

### Task 9

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

### Task 10

**Given your previous course experience with R/RStudio, what topics/concepts did you find to be most challenging?**

This week has been an excellent refresher on R for me. I took MIS 503 in the Fall 2019 semester. During this class, I was first introduced to R. I have been pleasantly surprised to find that I am able to recall much of the content from my previous R course.

I still find myself needing to fiddle around with the syntax to make R do exactly what I’m intending. The hints in R are really helpful for finding where there might be an error in my code. A lot of the time, I’m simply missing a parenthesis, or I have an extra pipe operator or plus sign somewhere. I plan on using the cheat sheets and continuing to simply practice, practice, practice.