## R and RStudio Refresher Assignment

### BAN502 - Module 1 / Assignment 2

### Nick Barrett, 1/27/19

#### Task 1. Install and load “tidyverse”.

#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 0.7.7  
## v tidyr 0.8.2 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

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

#### Task 2: Read in the “diamonddata” data frame.

diamonddata = diamonds

nrow(diamonds)

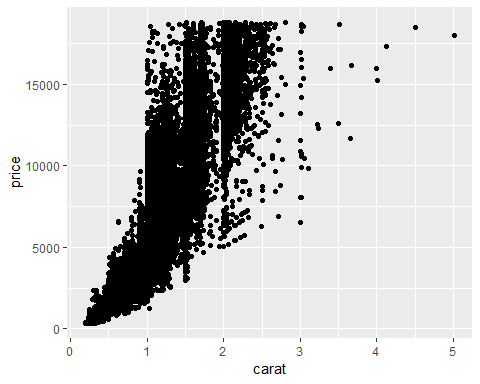
## [1] 53940

ncol(diamonds)

## [1] 10

#### Task 3: Create a scatterplot of carat versus price with ggplot.

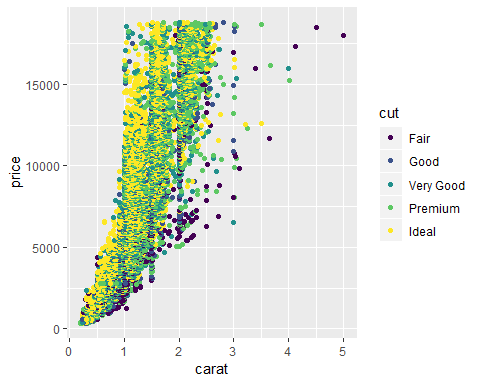
ggplot(diamonds,aes(carat,price))+  
geom\_point()



Considering the above scatterplot comparing carat to price, it is observed that generally the larger the carat size the higher the price.

#### Task 4: Repeat Task 3, using “cut” to color the scatterplot. Describe the relationship between the three variables (carat, price and cut)

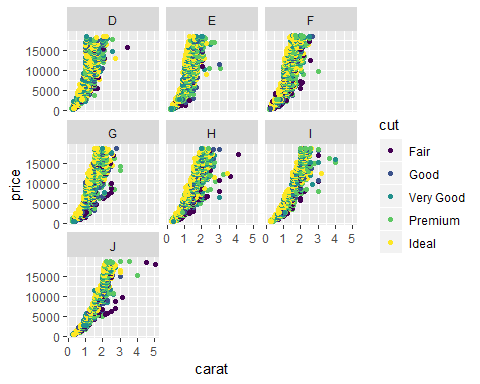
ggplot(diamonds,aes(carat,price,color=cut))+  
geom\_point()



Observing the above scatterplot with the three variables of carat, price and cut, the majority of the diamonds are 1 to 2 carats. Of these 1 to 2 carat diamonds, the majority of those diamonds are of cuts from “very good” to “ideal”. The number of “fair” and “good” cuts are far less less than the other cuts. Regarding price, carat size appears to have a positive relationship to pricing - bigger carat sizes are more expensive.

#### Task 5: Repeat Task 4 using facet to color the scatterplot. Describe the relationship between the four variables (carat, price, cut and color)

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



#### Task 6: Use “readr” to read-in the “InventoryData.csv” as a dataframe called “inventory”. Examine the structure and summary of data-frame.

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\_integer(),  
## `Annual Demand` = col\_integer()  
## )

summary(Inventory)

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

#### Task 7: Create a new data frame called “inventoryA” containing inventory from Supplier A. Determine the number of rows in the new data frame.

inventoryA <-filter(Inventory,Supplier=="A")

nrow(inventoryA)

## [1] 3695

The new data frame, “inventoryA”, has 3695 rows and 6 variables.

#### Task 8: Use the mutute code below and determine what it does.

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

The above mutate code adds a column, ‘OnHandRatio’, to the “inventoryA” data frame that calculates the ratio of ‘on hand’ inventory in relation to the ‘annual demand’.

#### Task 9: Create a new data frame called “avg\_cost” that contains the average “Cost per Unit ($)” by each “Item SKU” creating a new variable “SKUAvgCost”.

avg\_cost <- select(Inventory, "Item SKU", "Cost per Unit ($)")

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

avg\_cost <-select(avg\_cost,"Item SKU","SKUAvgCost")

#### Task 10: Given my previous experience with R/RStudio, the topics that I found most challenging.

I had Programming for Analytics last semester, so it is still somewhat fresh; however, I do not in anyway feel like I speak the language yet. I have to have a lot of reference points and examples to work my way through to solutions. I tend to figure it out but I get stuck often. I look forward to getting more comfortable in application.