## R and RStudio Refresher Assignment

### Anthony Sumter

#### Task 1

**Install and load the “tidyverse” package using the “install.packages” and “library” commands.**

#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()

#### Task 2

**The ggplot2 package (part of the tidyverse set of packages) includes a dataset containing data on diamonds. Use the line of code below to read in this dataset into a data frame called diamonddata**

diamonddata = diamonds

#### Task 3

**Using ggplot, create a scatterplot of caret (x axis) versus price (y axis). Briefly describe the relationship between these two variables.**

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

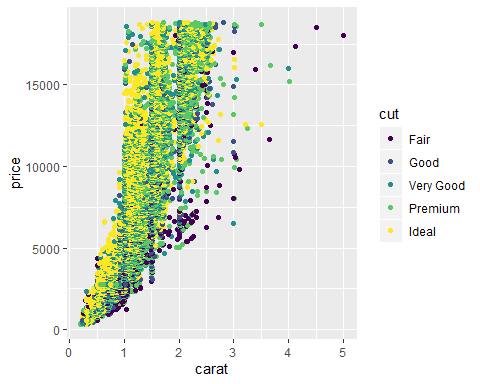


The relationship between these two variables is that as carat size increases so does the price of the diamond as well.

#### Task 4

**Repeat Task 3, but in this plot color the scatterplot points by the “cut” variable. Briefly describe the relationship between these three variables (carat, price, and cut).**

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



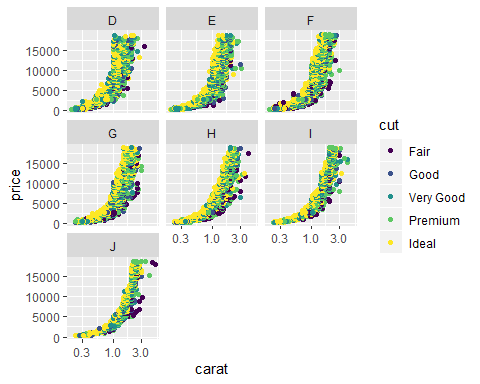
It appears that the relationship between the three variables of carat, cut, and price is that the smaller the carat is the more ideal the cut is leading to a more affordable price. Also, vice versa the price increases along with a larger carat and worse cut.

#### 

#### Task 5

**Repeat Task 4, but in this plot facet by “color”. Briefly describe the relationship between the four variables (carat, price, cut, and color).**

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



Looking at this data as the color moves further up the alphabet the price and carat size goes up all the while the quality of the cut of the diamond decreases.

#### Task 6

**Use the “readr” package (part of the tidyverse) to read-in the “InventoryData.csv” file as a data frame called “inventory”. Examine the structure and summary of the data frame.**

library(readr)  
Inventory <- read\_csv("InventoryData(1).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()  
## )

View(Inventory)

The Inventory data frame has a total of 13,561 observations and 6 variables with data about the inventory of various suppliers.

#### Task 7

**Use a filter to create a new data frame called “inventoryA” containing only inventory from Supplier**

*A. How many rows are in this new data frame?*

There is a total of 3,695 rows (observations).

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

#### Task 8

**What does the line of code shown below do?**

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

The line of code adds an additional column (variable) to the inventoryA data frame of OnHandRatio. This new variable is the result of dividing the “On Hand” variable by the “Annual Demand” variable.

#### Task 9

**Create a new data frame called “avg\_cost” that contains the average “Cost per Unit ($)” by each “Item SKU”.**

avg\_cost <- inventoryA %>%  
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?**

Based on the previous course I would say the topics that were the most challenging was dealing with proper usage of functions such as select, rename, gather, etc.