# BAN 502

## Module 1, Assignment 3

### Susan Lawhead

## Task 1

#install.packages("esquisse") commented out after initial install  
library(tidyverse)  
library(esquisse)

## Task 2

diamonddata <- diamonds

**There are 53940 rows of data with 10 columns representing the variables**

## Task 3

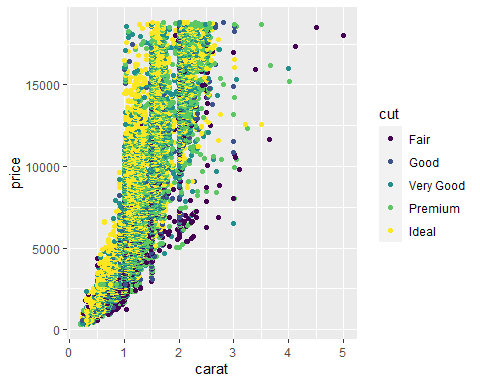
ggplot(diamonddata, aes(x = carat, y = price))+  
 geom\_point()



**When looking at price and carat, we see that typically, as the carat size increases, so does the price of the diamond**

## Task 4

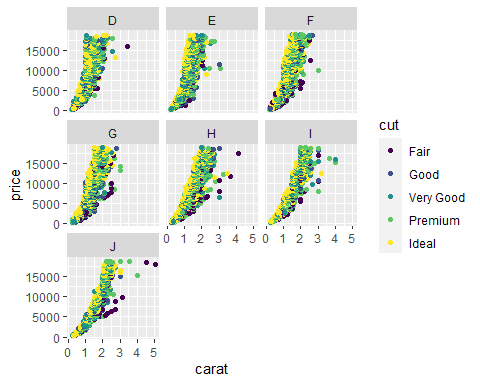
ggplot(diamonddata, aes (x = carat, y = price, color = cut))+  
 geom\_point()



**In this graph, we see that there are more large diamonds in the lower quality cuts. We see most of the ideal diamonds are in between 0-2 carats in size, with some up to 3, and few above that. In the lower carats, the higher quality appears more expensive, however, ultimately, size has the larger impact with fair diamonds of large size being the most expensive.**

## Task 5

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



**The cut and size of diamonds still have a large impact on the price of diamonds, even across the various colors. There are fewer large diamonds in D-G, with more large diamonds of lower cut quality stating at H and going through J. The largest, most expensive diamonds are in color J.**

## Task 6

**Inventory is a tibble with 6 variables for 13,561 observations. 3 of the variables are character while 3 are numeric. The mean cost per unit is $504.4, with a mean of 100.5 on hand. The mean average demand is 966.2.**

## Task 7

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

**After filtering for supplier A, we have a total of 3695**

## Task 8

The following code creates a new variable, adding a new column to the data set. This new variable will show the ratio of items on hand compared to the annual demand.

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

## Task 9

The following code will create a new data frame with only the desired information.

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

## Task 10

Overall, I did not find this assignment difficult. There have been some differences in preferred syntax between MIS 503 and the DataCamp course, however, it has been easy to identify the desired outcome and still perform the desired task. I do get a bit confused of when to use the backtick vs. apostrophe, but I think I am getting that figured out (Selecting vs. mutating/changing).