# Refresher Assignment

## David Cleveland

### Task 1

#install.packages(tidyverse)  
  
library(tidyverse)

#### Task 2

diamonddata = diamonds  
  
#head(diamonddata)

#### There are 53940 observations, 10 variables, 6 rows and 10 columns in this dataset.

#### Task 3

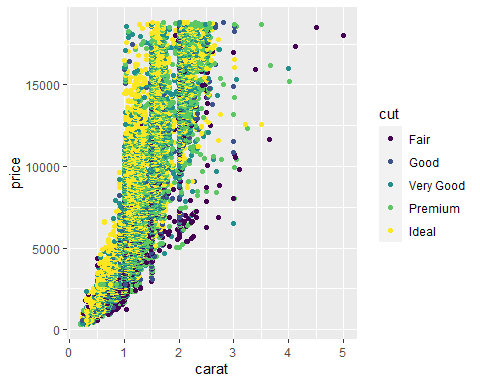
ggplot(diamonddata, aes(x = carat, y = price)) +  
 geom\_point()



#### Generally, the larger the carat the more the diamond will cost.

#### Task 4

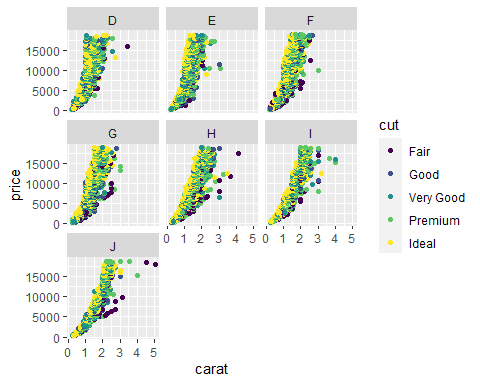
ggplot(diamonddata, aes(x = carat, y = price, color = cut)) +  
 geom\_point()



#### Generally, the larger the carat the more the diamond will cost. Additionally, the more the diamond cost the more ideal the cut appears. Conversely, the lease expensive the diamond seem to be fair.

#### Task 5

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



#### The relationship between carat and price remain the same. Generally, the larger the carat the higher the price. However, as you begin to add the factor of color one begins to see differences in the graphs. It appears that colors “D, E, F and G” tend to have smaller carats but still have higher cost. There are less occurrences of diamonds above 3 carats in color “D,E,F and G.”

#### Task 6

InventoryData <- read\_csv("InventoryData.csv")  
  
inventory = InventoryData  
  
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")  
  
head(inventoryA)

## # A tibble: 6 x 6  
## `Item SKU` Store Supplier `Cost per Unit ($)` `On Hand` `Annual Demand`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 0100 003480 A 125. 159 1693  
## 2 011 020109 A 12.3 173 1695  
## 3 0113 031779 A 208. 166 1496  
## 4 0113 080212 A 187. 157 1654  
## 5 0122 003480 A 68.5 34 290  
## 6 0122 020109 A 120. 77 680

#### There are 3,695 observations and 6 variables (rows) in this data set.

#### Task 8

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

## # A tibble: 6 x 7  
## `Item SKU` Store Supplier `Cost per Unit ~ `On Hand` `Annual Demand`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 0100 0034~ A 125. 159 1693  
## 2 011 0201~ A 12.3 173 1695  
## 3 0113 0317~ A 208. 166 1496  
## 4 0113 0802~ A 187. 157 1654  
## 5 0122 0034~ A 68.5 34 290  
## 6 0122 0201~ A 120. 77 680  
## # ... with 1 more variable: OnHandRatio <dbl>

#### The line of code mutates the dataset to add a variable (row), which is labeled “OnHandRatio.” The code is telling the table divide “On Hand” by “Annual Demand” and call the variable “OnHandRatio.”

#### Task 9

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

## # A tibble: 1,720 x 2  
## `Item SKU` SKUAvgCost  
## <chr> <dbl>  
## 1 0100 125.   
## 2 011 12.3   
## 3 0113 197.   
## 4 0122 68.5   
## 5 013 19.1   
## 6 0133 203.   
## 7 0137 165.   
## 8 014 9.22  
## 9 0151 75.7   
## 10 0152 269.   
## # ... with 1,710 more rows

#### Task 10

#### I found task 9 to be the most challenging. I was not able to arrive with the average of “Cost per Unit ($).”