# R and RStudio Refreasher Assignment

## Zac Coffey

Question 1

#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()

Question 2

diamonddata <- diamonds #renamed table from ggplot2  
nrow(diamonddata)

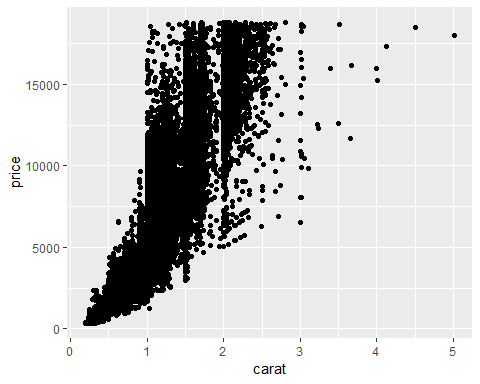
## [1] 53940

ncol(diamonddata)

## [1] 10

Question 3

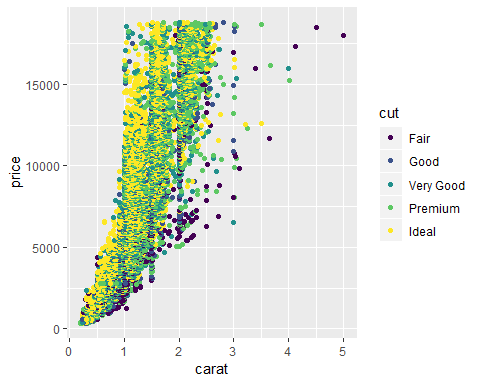
ggplot(diamonddata,aes(carat,price))+geom\_point()



#There is a postive linear relationship between these variables carat and price. As the carat size increases so does the price.

Question 4

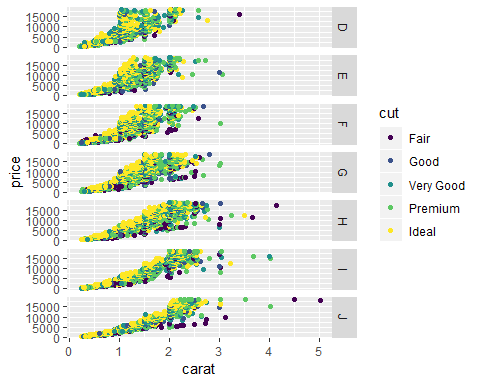
ggplot(diamonddata,aes(carat,price,color=cut))+geom\_point()



#This scatterplot shows that there is a relationship between price and cut. For instance a high quality cut has a higher price than a low quailty cut. The carat size appears to have no correlation to the quality of cut. For instance a size 5 carat has fair quality not premium or ideal

Question 5

ggplot(diamonddata,aes(carat,price,color=cut))+geom\_point()+facet\_grid (color ~ .)



#This scatterplot shows that all diamond colors have a postive linear relationship with carat size, price and cut. For instance color J increase in price and quality of cut as carat size increases. Important to note that there are outliers.

Question 6

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()  
## )

View(Inventory)   
#The dataset has 13561 rows and 6 columns

Question 7

InventoryA <- Inventory %>% filter(Supplier == "A") #Filtered supplier column from Inventory   
  
nrow(InventoryA)

## [1] 3695

Question 8

InventoryA = mutate(InventoryA, OnHandRatio = `On Hand` / `Annual Demand`)   
#This line of code created a new column that shows (On Hand) being divided by (Annual Demand). This new column would help measure the efficiently of the company.

Question 9

Avg\_Cost <- InventoryA %>% group\_by(`Item SKU`) %>%   
 summarise(SKUAvgCost= mean(`Cost per Unit ($)`))   
  
View(Avg\_Cost) #display the newly grouped and summarized data

Question 10

I find the dyplr package to be difficult to not confuse the arrange and filter functions. Also I found the facet\_grid to be difficult.