## Task 1

#install.packages("tidyverse")  
library(tidyverse)

## -- Attaching packages -------------------------------- tidyverse 1.3.0 --

## 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()

## Task2

diamonddata = diamonds  
# get number of rows  
nrow(diamonddata)

## [1] 53940

# get number pof columns  
ncol(diamonddata)

## [1] 10

## Task3

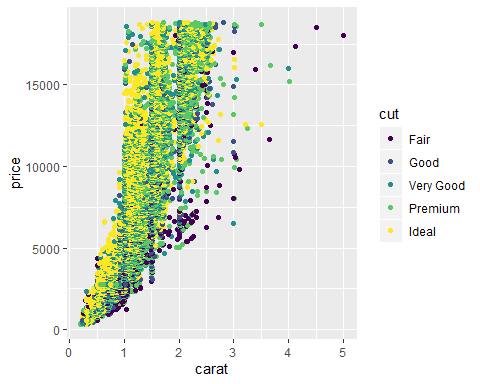
ggplot(diamonddata,aes(x=carat,y=price))+  
 geom\_point()



Carat and price have a positive correlation, Price increases with carat size increase.

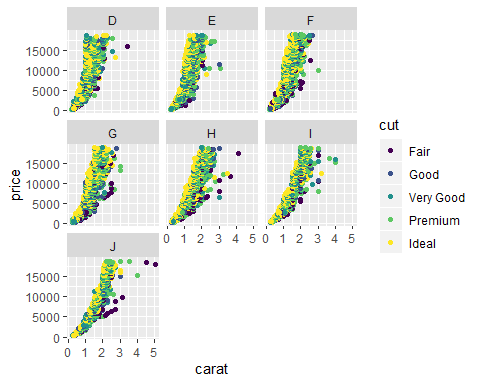
## Task 4

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

 Carat and price have a positive correlation in all cut sizes, Price increases with carat size increase in all cut sizes, diamonds with high cut are available more with small carat, while lower diamonds with cut have more carat distributions.

## Task 5

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



Carat and price have a positive correlation in all cut sizes and colors, Price increases with carat size increase in all cut sizes and colors, Colors D, E, F and G have more explorational increases while there are more variation distributions in J,H and J

## Task 6

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

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")

How many rows are in this new data frame? 3695

## Task 8

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

the above line of code, added a new column OnHandRatio, which has the value of on hand divided by annual deman.

## Task 9

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

## Task 10

Task 9 was a little challenging since the column names have spaces.