Task1:

#install.packages(tidyverse)  
library(tidyverse)

## Registered S3 methods overwritten by 'ggplot2':  
## method from   
## [.quosures rlang  
## c.quosures rlang  
## print.quosures rlang

## Registered S3 method overwritten by 'rvest':  
## method from  
## read\_xml.response xml2

## -- Attaching packages ---------------------- tidyverse 1.2.1 --

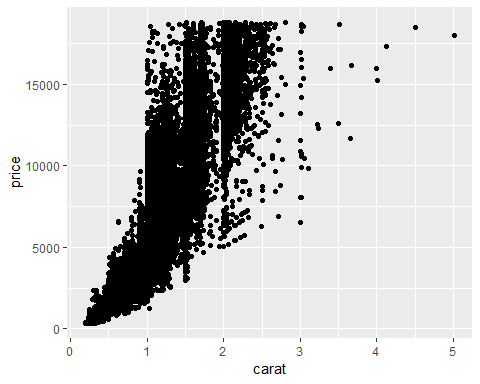
## v ggplot2 3.1.1 v purrr 0.3.2   
## v tibble 2.1.1 v dplyr 0.8.0.1  
## v tidyr 0.8.3 v stringr 1.4.0   
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

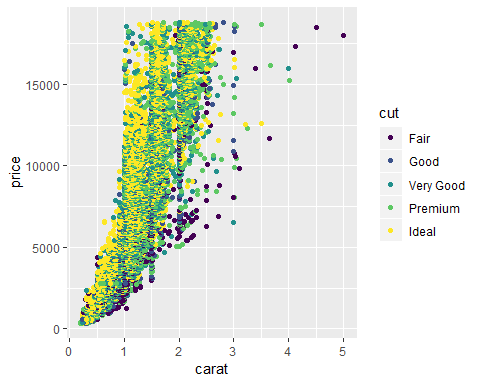
diamondsdata = diamonds  
  
diamonds

## # A tibble: 53,940 x 10  
## carat cut color clarity depth table price x y z  
## <dbl> <ord> <ord> <ord> <dbl> <dbl> <int> <dbl> <dbl> <dbl>  
## 1 0.23 Ideal E SI2 61.5 55 326 3.95 3.98 2.43  
## 2 0.21 Premium E SI1 59.8 61 326 3.89 3.84 2.31  
## 3 0.23 Good E VS1 56.9 65 327 4.05 4.07 2.31  
## 4 0.290 Premium I VS2 62.4 58 334 4.2 4.23 2.63  
## 5 0.31 Good J SI2 63.3 58 335 4.34 4.35 2.75  
## 6 0.24 Very Good J VVS2 62.8 57 336 3.94 3.96 2.48  
## 7 0.24 Very Good I VVS1 62.3 57 336 3.95 3.98 2.47  
## 8 0.26 Very Good H SI1 61.9 55 337 4.07 4.11 2.53  
## 9 0.22 Fair E VS2 65.1 61 337 3.87 3.78 2.49  
## 10 0.23 Very Good H VS1 59.4 61 338 4 4.05 2.39  
## # ... with 53,930 more rows

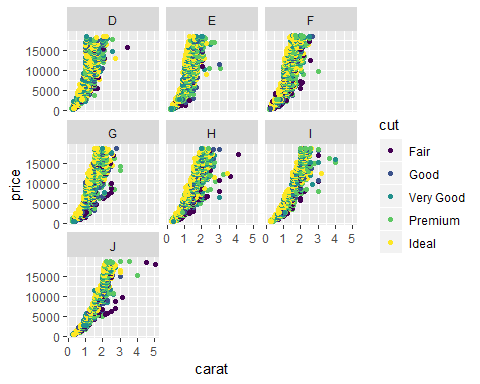
ggplot(diamondsdata, aes(x=carat, y=price)) + geom\_point()



ggplot(diamondsdata, aes(x=carat, y=price, colour= cut)) + geom\_point()



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



Task 2: There are 53,940 rows in the diamonds dataset.

Task 3: The relationship between carat and price is positive. As the number of carats increases, the price also increases.

Task 4: The majority of the premium cut diamonds are of higher carat and of higher price.

Task 5: All color of diamonds are experiencing the same trend, as the carat increases so does the price. Color D seems to be the most expensive. There are some instances where just 1 carat can cost as much as $20,000.

Task 6:

library(readr)  
InventoryData <- 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()  
## )

View(InventoryData)  
  
inventoryA <- InventoryData %>%   
 filter(Supplier == "A")  
  
  
inventoryA = mutate(inventoryA, OnHandRatio = `On Hand` / `Annual Demand`)

Task 7: In the inventoryA dataset, there are 3,695 rows.

Task 8: The code shown on the assignment sheet adds a column called “OnHandRatio” to the “inventoryA” dataset that is consists of dividing the “OnHand” amount by the “Annual Demand” amount.

Task 9:

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

Task 10: Personally, I didn’t find the concepts covered in MIS 503 to be very challenging. One thing I would like clarification on is when the would it be more appropriate to create a tibble instead of a data frame and vice versa.