##Carina Del Rio ## R and RStudio Refresher Quiz

Libraries that I need.

library(tidyverse)

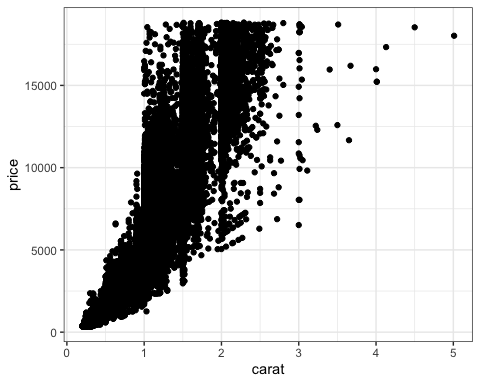
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

diamonddata = as\_tibble(diamonds)  
str(diamonddata)

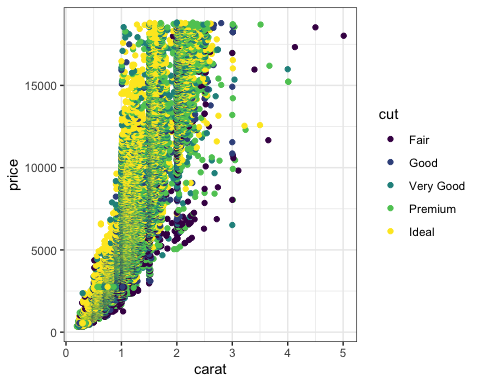
## tibble [53,940 × 10] (S3: tbl\_df/tbl/data.frame)  
## $ carat : num [1:53940] 0.23 0.21 0.23 0.29 0.31 0.24 0.24 0.26 0.22 0.23 ...  
## $ cut : Ord.factor w/ 5 levels "Fair"<"Good"<..: 5 4 2 4 2 3 3 3 1 3 ...  
## $ color : Ord.factor w/ 7 levels "D"<"E"<"F"<"G"<..: 2 2 2 6 7 7 6 5 2 5 ...  
## $ clarity: Ord.factor w/ 8 levels "I1"<"SI2"<"SI1"<..: 2 3 5 4 2 6 7 3 4 5 ...  
## $ depth : num [1:53940] 61.5 59.8 56.9 62.4 63.3 62.8 62.3 61.9 65.1 59.4 ...  
## $ table : num [1:53940] 55 61 65 58 58 57 57 55 61 61 ...  
## $ price : int [1:53940] 326 326 327 334 335 336 336 337 337 338 ...  
## $ x : num [1:53940] 3.95 3.89 4.05 4.2 4.34 3.94 3.95 4.07 3.87 4 ...  
## $ y : num [1:53940] 3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 ...  
## $ z : num [1:53940] 2.43 2.31 2.31 2.63 2.75 2.48 2.47 2.53 2.49 2.39 ...

ggplot2 (Scatterplot)

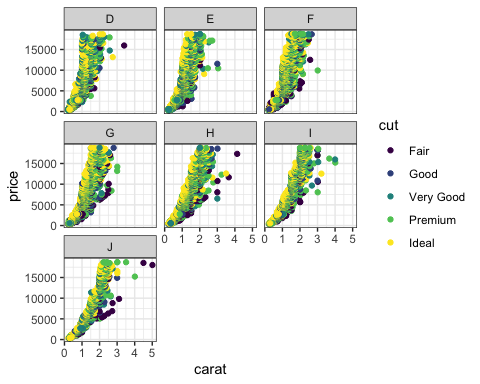
ggplot(diamonddata, aes(x = carat, y = price)) + geom\_point() + theme\_bw()

 Scatterplot Points color by Cut

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

 Facet by “color”

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

 Comma Separated Files (CSV)

Inventory = read\_csv("InventoryData.csv")

## Rows: 13561 Columns: 6  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): Supplier  
## dbl (5): Item SKU, Store, Cost per Unit ($), On Hand, Annual Demand  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

##Examining Data

str(Inventory)

## spc\_tbl\_ [13,561 × 6] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Item SKU : num [1:13561] 100 100 100 100 100 100 100 11 11 11 ...  
## $ Store : num [1:13561] 3480 1611 1611 20109 20109 ...  
## $ Supplier : chr [1:13561] "A" "B" "D" "B" ...  
## $ Cost per Unit ($): num [1:13561] 125.32 115.12 53.61 2.26 60.51 ...  
## $ On Hand : num [1:13561] 159 40 174 176 74 48 6 129 82 17 ...  
## $ Annual Demand : num [1:13561] 1693 351 1691 1559 733 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. `Item SKU` = col\_double(),  
## .. Store = col\_double(),  
## .. Supplier = col\_character(),  
## .. `Cost per Unit ($)` = col\_double(),  
## .. `On Hand` = col\_double(),  
## .. `Annual Demand` = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

summary(Inventory)

## Item SKU Store Supplier Cost per Unit ($)  
## Min. : 6 Min. : 1611 Length:13561 Min. : 0.0   
## 1st Qu.:2537 1st Qu.: 3480 Class :character 1st Qu.: 137.0   
## Median :4997 Median :20109 Mode :character Median : 377.5   
## Mean :5025 Mean :26675 Mean : 504.4   
## 3rd Qu.:7602 3rd Qu.:31779 3rd Qu.: 775.5   
## Max. :9998 Max. :80212 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

##Question 7: Mean of the “On Hand” variable: 100.5

##Question 8: How many rows are in this new data frame? A total of 3695 rows

Mutate Change Character / Filter Select Rows that meet some criteria

Inventory = Inventory %>% mutate(Supplier = as\_factor(Supplier))  
InventoryA = Inventory %>% filter(Supplier == 'A')

##Question 9: This chunk of code added a new variable called ‘OnHandRatio’

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

##Question 10

avg\_cost = InventoryA %>% select(c('Cost per Unit ($)','Item SKU'))  
avg\_cost = avg\_cost %>% mutate(KUAvgCos = 'Cost per Unit ($)')  
avg\_cost = avg\_cost %>% mutate(KUAvgCos = as\_factor(KUAvgCos))  
avg\_cost = avg\_cost %>% group\_by(`Item SKU`) %>% summarise(KUAvgCos = mean(`Cost per Unit ($)`))  
summary(avg\_cost)

## Item SKU KUAvgCos   
## Min. : 6 Min. : 0.0   
## 1st Qu.:2506 1st Qu.: 180.7   
## Median :4976 Median : 429.8   
## Mean :5005 Mean : 507.4   
## 3rd Qu.:7582 3rd Qu.: 752.8   
## Max. :9998 Max. :1969.4

str(avg\_cost)

## tibble [1,720 × 2] (S3: tbl\_df/tbl/data.frame)  
## $ Item SKU: num [1:1720] 6 7 11 13 14 30 44 46 53 55 ...  
## $ KUAvgCos: num [1:1720] 6.59 11.42 12.33 14.32 9.22 ...