#### Module 1, Assignment 3 - Linnea Williams

**Task 1 - Libraries & Files**

#install.packages("esquisse")  
#received a error after trying to install esquisse that said I would need to install RTools?  
library(tidyverse)

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

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.4 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

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

inventory <- read\_csv("InventoryData.csv") #Task 6 file load

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## `Item SKU` = col\_double(),  
## Store = col\_double(),  
## Supplier = col\_character(),  
## `Cost per Unit ($)` = col\_double(),  
## `On Hand` = col\_double(),  
## `Annual Demand` = col\_double()  
## )

**Task 2 - Read in diamonddata**

diamonddata=diamonds  
str(diamonddata)

## tibble [53,940 x 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 ...

**Task 3 - ggplot diamonddata**

*Overall, the larger the carat, the more expensive the price. However, we can tell by the way points are clustered between the carat size 1 and 2.5, that carat is not the only variable and possibly not the strongest variable impacting the price.*

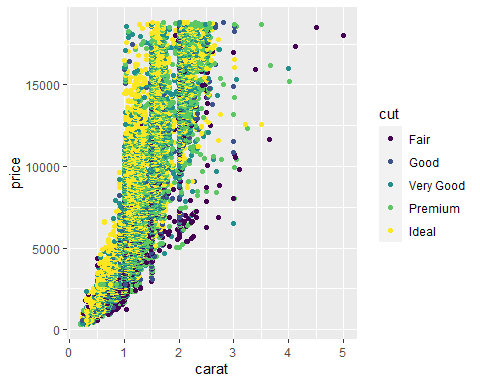
ggplot(diamonddata, aes(x=carat, y=price))+  
 geom\_point()



**Task 4 - ggplot diamnoddata w/ color**

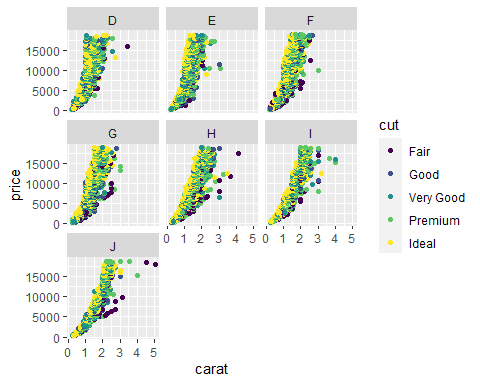
*The more ideal cut of the diamond, the more expensive the price. The more ideal cut of the diamond, trumping the number of carats (as far as impact to prices goes) when carat sizes are betwen 1 and 2.5.*

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



**Task 5 - ggplot diamonddata, facet by color**

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



**Task 6 - Examine the structure and summary of “inventory”**

#loaded inventory when libraries were loaded in Task 1  
str(inventory)

## tibble [13,561 x 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()  
## .. )

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

**Task 7 - Create new df called inventoryA**

*There are 3695 rows in this new data frame.*

inventoryA <- inventory %>%  
 filter(Supplier=="A")  
  
str(inventoryA)

## tibble [3,695 x 6] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Item SKU : num [1:3695] 100 11 113 113 122 122 122 13 13 13 ...  
## $ Store : num [1:3695] 3480 20109 31779 80212 3480 ...  
## $ Supplier : chr [1:3695] "A" "A" "A" "A" ...  
## $ Cost per Unit ($): num [1:3695] 125.3 12.3 208.2 186.5 68.5 ...  
## $ On Hand : num [1:3695] 159 173 166 157 34 77 133 28 103 29 ...  
## $ Annual Demand : num [1:3695] 1693 1695 1496 1654 290 ...  
## - 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()  
## .. )

**Task 8 - add new variable “OnHandRatio” to InventoryA**

*This line of code adds a new variable called “OnHand Ratio”*

inventoryA=mutate(inventoryA,OnHandRatio=`On Hand`/`Annual Demand`)  
  
#Note the use of backtick character to delineate the variable names with spaces in them  
  
str(inventoryA)

## tibble [3,695 x 7] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Item SKU : num [1:3695] 100 11 113 113 122 122 122 13 13 13 ...  
## $ Store : num [1:3695] 3480 20109 31779 80212 3480 ...  
## $ Supplier : chr [1:3695] "A" "A" "A" "A" ...  
## $ Cost per Unit ($): num [1:3695] 125.3 12.3 208.2 186.5 68.5 ...  
## $ On Hand : num [1:3695] 159 173 166 157 34 77 133 28 103 29 ...  
## $ Annual Demand : num [1:3695] 1693 1695 1496 1654 290 ...  
## $ OnHandRatio : num [1:3695] 0.0939 0.1021 0.111 0.0949 0.1172 ...  
## - 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()  
## .. )

**Task 9 - Create new df, “avg\_cost”**

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

## `summarise()` ungrouping output (override with `.groups` argument)

head(avg\_cost)

## # A tibble: 6 x 2  
## `Item SKU` SKUAvgCost  
## <dbl> <dbl>  
## 1 6 6.59  
## 2 7 11.4   
## 3 11 12.3   
## 4 13 14.3   
## 5 14 9.22  
## 6 30 26.3

**Task 10 - Most challenging topics**

*I’ve only taken the MIS 503 course as it relates to R/R Studio. The concepts in the course were actually pretty easy for me to follow. I think what is most challenging is just learning a new syntax/language. Many times during the course assignments I find my thoughts getting mixed in with other syntax like Excel formulas and Sql or dax syntax and have to reel myself back into r. It can get a bit frustrating (yet fun) navigating through small issues like remembering to add piping at the end of line, misspellings, or capitalization vs lowercase in variable/function names, or knowing he difference between a warning and an error message more quickly. For example in Task 10 of this assignment, I was stuck for about 20 minutes trying to fix what I thought to be an error “summarise() ungrouping output (override with .groups argument” which turned out to be just a warning message (palms forehead after quick Google search). But all concepts are easy to grasp, DataCamp this session is also great supplement to the class.*