MIS 506 Module 1 Assignment 1: R and R Studio

Tabitha Hagen

2023-03-27

# MIS 506 Module 1 - Assignment 1

## R and R Studio Refresher Assigment

Task 1: Install and load the “tidyverse” package using the “install.packages” and “library” commands. Note: It is strongly advised that you do not actually run the “install.packages” command when knitting your Markdown document. I strongly suggest running this function (by running the R code code chunk) to allow the “tidyverse” package to be installed. After doing this, comment out the “install.packages” command before knitting your Markdown document. Recall that placing a # before a line of R code comments it out and prevents it from being executed.

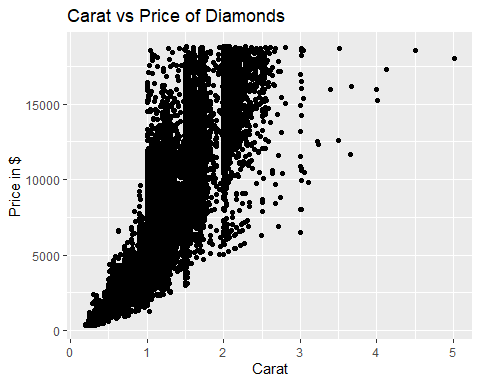
#install.packages("tidyverse")  
#install.packages("rlang")  
#install.packages("cli")  
library(tidyverse)

Task 2: The “ggplot2” package (part of the tidyverse set of packages) includes a dataset containing data on diamonds. Use the line of code below to read in this dataset into a data frame called “diamonddata”.

# Read default dataset Diamonds  
diamonddata <-diamonds

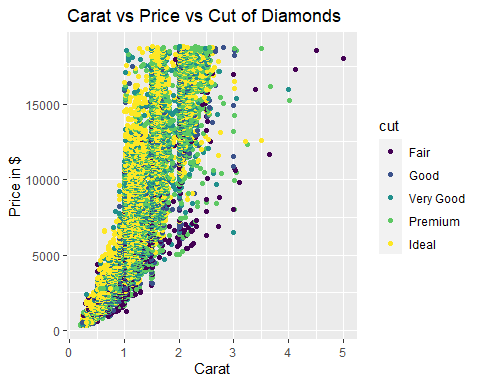
Task 3: Using ggplot, create a scatterplot of carat (x axis) versus price (y axis). Briefly describe the relationship between these two variables. Carat and Price share a direct relationship where the larger the carat, the more it cost.

# scatterplot Task 3  
ggplot(diamonddata,mapping=aes(carat, price)) +   
 # aes(x variable, y variable)  
 geom\_point()+  
 labs(title = "Carat vs Price of Diamonds", # Plot Title of Scatterplot  
 x = "Carat", # x axis label  
 y = "Price in $" # y axis label  
 )



Task 4: Repeat Task 3, but in this plot color the scatterplot points by the “cut” variable. Briefly describe the relationship between these three variables (carat, price, and cut). Again, they are in a direct relationship. When the Carat goes up, the price goes up; when the cut goes up, the price also goes up. However, the larger carats usually have less quality cuts.

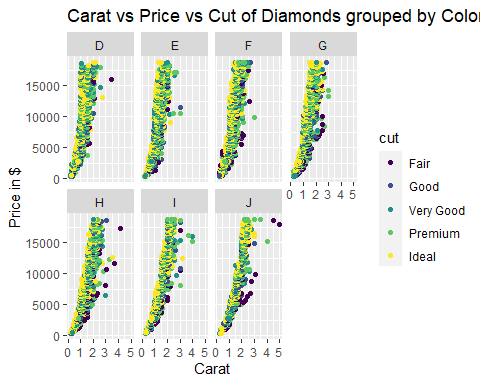
# scatterplot Task 4  
ggplot(diamonddata,mapping=aes(carat, price, color=cut)) +   
 # aes(x variable, y variable)  
 geom\_point()+  
 labs(title = "Carat vs Price vs Cut of Diamonds", # Plot Title of Scatterplot  
 x = "Carat", # x axis label  
 y = "Price in $" # y axis label  
 )



Task 5: Repeat Task 4, but in this plot facet by “color”. You can show the plots in four rows: nrow = 4. Briefly describe the relationship between the four variables (carat, price, cut, and color).

The higher color grades have higher carats, but there are many more lower color grades below 3 carats.

# scatterplot Task 5  
ggplot(diamonddata,mapping=aes(carat, price, color=cut)) +   
 # aes(x variable, y variable)  
 geom\_point()+  
 facet\_wrap(~color, nrow = 2)+  
 labs(title = "Carat vs Price vs Cut of Diamonds grouped by Color Grade", # Plot Title of Scatterplot  
 x = "Carat", # x axis label  
 y = "Price in $" # y axis label  
 )



Task 6: Use the “read\_csv” function to read-in the “InventoryData.csv” file as a data frame called “inventory”. Examine the structure and summary of the data frame. Hint: Recall the head() function.

# Read\_csv Task 6  
 library(readr)  
inventory <- read\_csv("~/UNCW/MIS 506 Text and Unstructured Data/MIS 506 week (1)/InventoryData.csv")  
head(inventory, n=10)

## # A tibble: 10 × 6  
## `Item SKU` Store Supplier `Cost per Unit ($)` `On Hand` `Annual Demand`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 0100 003480 A 125. 159 1693  
## 2 0100 01611 B 115. 40 351  
## 3 0100 01611 D 53.6 174 1691  
## 4 0100 020109 B 2.26 176 1559  
## 5 0100 020109 C 60.5 74 733  
## 6 0100 020109 D 53.7 48 496  
## 7 0100 080212 B 107. 6 58  
## 8 011 003480 B 1.33 129 1106  
## 9 011 003480 C 12.9 82 771  
## 10 011 01611 C 5.16 17 172

Task 7: Use a filter to create a new data frame called “inventoryA” containing only inventory from Supplier A. How many rows are in this new data frame? ’

There are 3695 rows for Supplier A.

# Filter inventory dataset Task 7  
inventoryA <- inventory %>% # piping reduces redundant commands   
 filter(Supplier=="A") # filter by Supplier A  
nrow(inventoryA)

## [1] 3695

Task 8: What does the line of code shown below do? Note the use of the backtick character (on the tilde key on your computer’s keyboard) to delineate the variable names with spaces in them. ### inventoryA = mutate(inventoryA, OnHandRatio = On Hand / Annual Demand)

It creates a new column called “OnHandRatio” which equals “OnHand” divided by “Annual Demand”.

# Task 8 Add a calculated column  
inventoryA = mutate(inventoryA, OnHandRatio = `On Hand` / `Annual Demand`)  
head(inventoryA, n=10)

## # A tibble: 10 × 7  
## `Item SKU` Store Supplier `Cost per Unit ($)` `On Hand` Annual Dem…¹ OnHan…²  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 0100 003480 A 125. 159 1693 0.0939  
## 2 011 020109 A 12.3 173 1695 0.102   
## 3 0113 031779 A 208. 166 1496 0.111   
## 4 0113 080212 A 187. 157 1654 0.0949  
## 5 0122 003480 A 68.5 34 290 0.117   
## 6 0122 020109 A 120. 77 680 0.113   
## 7 0122 031779 A 56.6 133 1239 0.107   
## 8 013 003480 A 19.1 28 277 0.101   
## 9 013 020109 A 22.7 103 962 0.107   
## 10 013 031779 A 1.13 29 297 0.0976  
## # … with abbreviated variable names ¹​`Annual Demand`, ²​OnHandRatio

Task 9: Create a new data frame called “avg\_cost” that contains the average “Cost per Unit ($)” by each “Item SKU” (let this quantity be in a variable called “SKUAvgCost. Hint: Recall the summarize and group\_by functions and the use of the backtick character from Task 8. Your data frame should have only two columns:”Item SKU" and “SKUAvgCost”.

# Task 9 Add a new dataframe with 2 columns  
avg\_cost <- select(inventory)  
  
  
avg\_cost <-inventory%>%  
 group\_by(`Item SKU`) %>%  
 summarize(SKUAvgCost = mean(inventory$`Cost per Unit ($)`, na.rm = TRUE)) #remove N/A's  
avg\_cost

## # A tibble: 1,817 × 2  
## `Item SKU` SKUAvgCost  
## <chr> <dbl>  
## 1 0100 504.  
## 2 011 504.  
## 3 0113 504.  
## 4 0122 504.  
## 5 013 504.  
## 6 0133 504.  
## 7 0137 504.  
## 8 014 504.  
## 9 0151 504.  
## 10 0152 504.  
## # … with 1,807 more rows

Task 10: Given your previous course experience with R/RStudio, what topics/concepts did you find to be most challenging?

First, I found the update to RTools a challenge because i had to go into R, create a file, in characters which I did not understand. It was “just type this” but it was for the outdated versions. Then I had to piecemeal which packages were outdated. You can’t update packages that are in use. It was 2 hours I’ll never get back.

Secondly, the last part of Task 9, I could not get. I am not as familiar with the make a table skills as most of what we did in R programming was creating and analyzing visualizations. The average portion is a Total Average not average per Item SKU.