Mod 1 Assign 3 Answers

Natasha Minott

1/23/2021

## Module 1 Assignment 3

## Natasha Minott

## Task 1 & 2

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

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

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

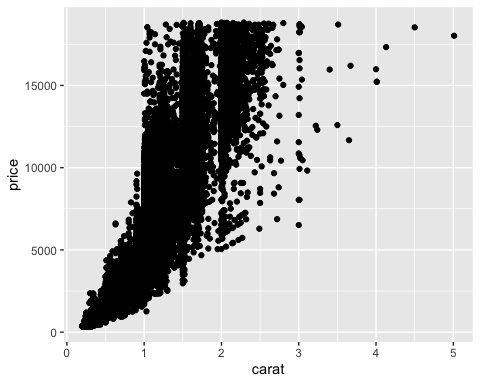
diamonddata = diamonds  
view(diamonds)  
glimpse(diamonds)

## Rows: 53,940  
## Columns: 10  
## $ carat <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0…  
## $ cut <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ve…  
## $ color <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I…  
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS1,…  
## $ depth <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4, 6…  
## $ table <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62, 5…  
## $ price <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340, 3…  
## $ x <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00, 4…  
## $ y <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05, 4…  
## $ z <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39, 2…

**Within this data set there are 53,940 rows and 10 columns.**

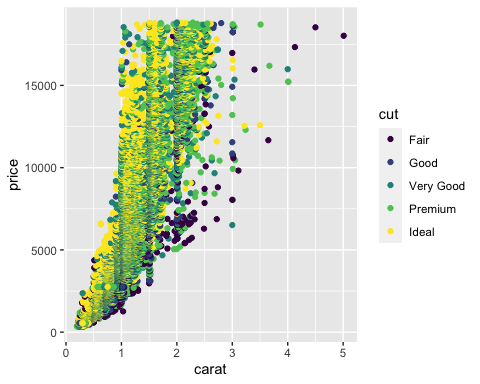
## Task 3

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

 **According to this scatter plot, as the carat size increases the price of a diamond increases as well, thus causing the relationship between carat and price to be linear.**

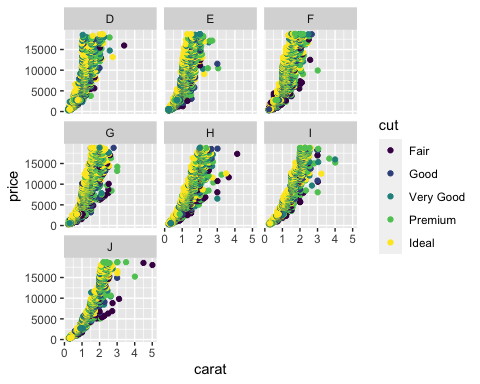
## Task 4

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

 **According to this colored scatter plot, many of the highly priced ideal and premium cut diamonds are of a smaller carat size.**

## Task 5

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



**The better color and cut of a diamond, the more expensive it is. Diamonds with a larger carat but with fair or good color and cut tend to be cheaper.**

## Task 6

inventory <- read\_csv("~/Desktop/502 Predictive Analytics/InventoryData .csv")

##   
## ── 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()  
## )

str(inventory)

## tibble [13,561 × 6] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Item SKU : chr [1:13561] "0100" "0100" "0100" "0100" ...  
## $ Store : chr [1:13561] "003480" "01611" "01611" "020109" ...  
## $ 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\_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")   
glimpse(inventoryA)

## Rows: 3,695  
## Columns: 6  
## $ `Item SKU` <chr> "0100", "011", "0113", "0113", "0122", "0122", "0…  
## $ Store <chr> "003480", "020109", "031779", "080212", "003480",…  
## $ Supplier <chr> "A", "A", "A", "A", "A", "A", "A", "A", "A", "A",…  
## $ `Cost per Unit ($)` <dbl> 125.32, 12.33, 208.19, 186.54, 68.50, 120.03, 56.…  
## $ `On Hand` <dbl> 159, 173, 166, 157, 34, 77, 133, 28, 103, 29, 143…  
## $ `Annual Demand` <dbl> 1693, 1695, 1496, 1654, 290, 680, 1239, 277, 962,…

**There are 3,695 rows in this new data frame.**

## Task 8

**This line of code is used to add a variable to the inventoryA dataframe that creates a new column called “OnHandRatio” that is the on hand variable divided by the annual demand variable.**

## Task 9

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

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

glimpse(avg\_cost)

## Rows: 1,720  
## Columns: 2  
## $ `Item SKU` <chr> "0100", "011", "0113", "0122", "013", "0133", "0137", "014…  
## $ SKUAvgCost <dbl> 125.32000, 12.33000, 197.36500, 81.70667, 14.31667, 203.47…

## Task 10

**I’ve been exposed to R/RStudio for the first time in my programming for Analytics class in the fall. Combining the group\_by and summarise functions take me a second to mentally process sometimes. More practice would make me comfortable working in R.**