Cara Arena Refresher Assignment

2025-05-18

Load Packages

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4   
## ── 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

Load Diamonds Dataset

diamonddata = diamonds

How many rows are in the Diamonds Dataset?

nrow(diamonddata)

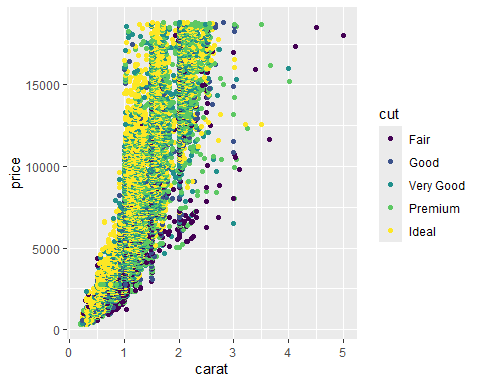
## [1] 53940

Basic Scatterplot: Price by Carat

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

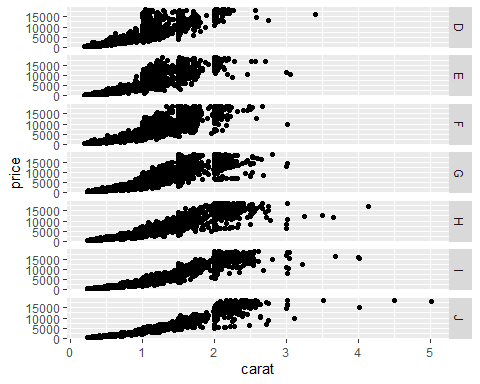
 Scatterplot: Price by Carat & Color

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

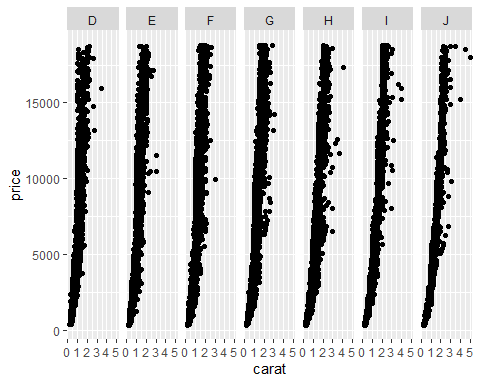


dd <- ggplot(diamonddata, aes(x = carat, y = price)) + geom\_point()

dd + facet\_grid(color ~ .)



dd + facet\_grid(. ~ color)



Basic R Functions

InventoryData <- read\_csv("InventoryData.csv")

## Rows: 13561 Columns: 6  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (3): Item SKU, Store, Supplier  
## dbl (3): 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.

summary(InventoryData)

## 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

inventoryA = InventoryData %>% filter(Supplier == "A")

nrow(inventoryA)

## [1] 3695

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

avg\_cost = inventoryA %>% group\_by(`Item SKU`) %>% summarise(avgcpu = mean(`Cost per Unit ($)`))

What is the average cost per unit for Item 011?

avg\_cost %>% filter(`Item SKU` == "011")

## # A tibble: 1 × 2  
## `Item SKU` avgcpu  
## <chr> <dbl>  
## 1 011 12.3