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 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── 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

options(repos = c(CRAN = "https://cran.r-project.org"))

diamond\_data = diamonds  
ncol(diamond\_data)

## [1] 10

nrow(diamond\_data)

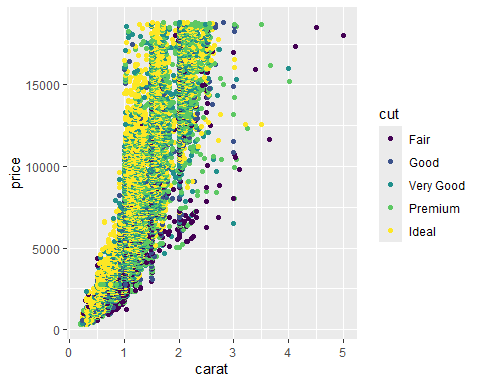
## [1] 53940

# Q1 How many rows in the dataset? Answer: 53940  
# Q2 How many columns in the dataset? Answer: 10

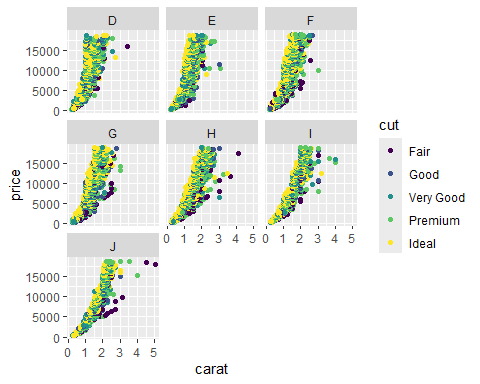
ggplot(diamond\_data, aes(carat, price)) + geom\_point()



ggplot(diamond\_data, aes(carat, price, color=cut)) + geom\_point()



ggplot(diamond\_data, aes(carat, price, color=cut)) + geom\_point() + facet\_wrap(~color)



library(readr)  
Inventory <- 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.

head(Inventory)

## # A tibble: 6 × 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

class(Inventory$Supplier)

## [1] "character"

mean(Inventory$`On Hand`, na.rm = TRUE)

## [1] 100.5373

InventoryA = filter(Inventory, Supplier == "A")  
nrow(InventoryA)

## [1] 3695

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

## # A tibble: 6 × 7  
## `Item SKU` Store Supplier `Cost per Unit ($)` `On Hand` `Annual Demand`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 0100 003480 A 125. 159 1693  
## 2 011 020109 A 12.3 173 1695  
## 3 0113 031779 A 208. 166 1496  
## 4 0113 080212 A 187. 157 1654  
## 5 0122 003480 A 68.5 34 290  
## 6 0122 020109 A 120. 77 680  
## # ℹ 1 more variable: OnHandRatio <dbl>

avg\_cost <- InventoryA %>%  
 group\_by(`Item SKU`) %>%  
 summarize(SKUAvgCost = mean(`Cost per Unit ($)`, na.rm = TRUE))  
  
avg\_cost %>% filter(`Item SKU` == "011")

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