Week one

Laureen Oduor

2024-10-22

# Section One

* Load mtcars, convert it to a tibble, and explore it with summar()

The program will load a car dataset, transform it into a more readable format, and then utilizing a function to easily extract the most important data statistics.

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

data("mtcars")  
mtcarsTib <- as\_tibble(mtcars)  
summary(mtcarsTib)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb   
## Min. :0.0000 Min. :3.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000   
## Median :0.0000 Median :4.000 Median :2.000   
## Mean :0.4062 Mean :3.688 Mean :2.812   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :5.000 Max. :8.000

# Section Two

* Select all columns except qsec and vs

From the dataset mtcars, the program is supposed to select all columns except qsec and vs.

select(mtcarsTib, c(-qsec, -vs))

## # A tibble: 32 × 9  
## mpg cyl disp hp drat wt am gear carb  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 21 6 160 110 3.9 2.62 1 4 4  
## 2 21 6 160 110 3.9 2.88 1 4 4  
## 3 22.8 4 108 93 3.85 2.32 1 4 1  
## 4 21.4 6 258 110 3.08 3.22 0 3 1  
## 5 18.7 8 360 175 3.15 3.44 0 3 2  
## 6 18.1 6 225 105 2.76 3.46 0 3 1  
## 7 14.3 8 360 245 3.21 3.57 0 3 4  
## 8 24.4 4 147. 62 3.69 3.19 0 4 2  
## 9 22.8 4 141. 95 3.92 3.15 0 4 2  
## 10 19.2 6 168. 123 3.92 3.44 0 4 4  
## # ℹ 22 more rows

# Section Three

* Filter for rows with cylinder numbers not equal to 8

Filtering out 8-cylinder cars and selecting the others.

filter(mtcarsTib, cyl != 8)

## # A tibble: 18 × 11  
## mpg cyl disp hp drat wt qsec vs am gear carb  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 21 6 160 110 3.9 2.62 16.5 0 1 4 4  
## 2 21 6 160 110 3.9 2.88 17.0 0 1 4 4  
## 3 22.8 4 108 93 3.85 2.32 18.6 1 1 4 1  
## 4 21.4 6 258 110 3.08 3.22 19.4 1 0 3 1  
## 5 18.1 6 225 105 2.76 3.46 20.2 1 0 3 1  
## 6 24.4 4 147. 62 3.69 3.19 20 1 0 4 2  
## 7 22.8 4 141. 95 3.92 3.15 22.9 1 0 4 2  
## 8 19.2 6 168. 123 3.92 3.44 18.3 1 0 4 4  
## 9 17.8 6 168. 123 3.92 3.44 18.9 1 0 4 4  
## 10 32.4 4 78.7 66 4.08 2.2 19.5 1 1 4 1  
## 11 30.4 4 75.7 52 4.93 1.62 18.5 1 1 4 2  
## 12 33.9 4 71.1 65 4.22 1.84 19.9 1 1 4 1  
## 13 21.5 4 120. 97 3.7 2.46 20.0 1 0 3 1  
## 14 27.3 4 79 66 4.08 1.94 18.9 1 1 4 1  
## 15 26 4 120. 91 4.43 2.14 16.7 0 1 5 2  
## 16 30.4 4 95.1 113 3.77 1.51 16.9 1 1 5 2  
## 17 19.7 6 145 175 3.62 2.77 15.5 0 1 5 6  
## 18 21.4 4 121 109 4.11 2.78 18.6 1 1 4 2

# Section Four

* Group by gear, summarize the medians of mpg and disp, and mutate a new variable that is the mpg median divided by the disp median

Cars are grouped by the number of gears they have, and the program calculates the median mpg and disp for each group. A new column is then created that shows how fuel efficiency compares to disp for each group.

mtcarsTib %>%  
group\_by(gear) %>%  
summarize(mpgMed = median(mpg), dispMed = median(disp)) %>%  
mutate(mpgOverDisp = mpgMed / dispMed)

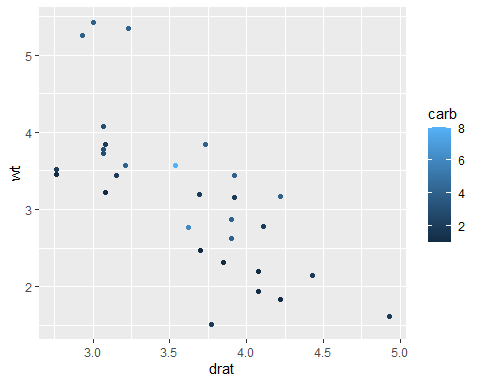
## # A tibble: 3 × 4  
## gear mpgMed dispMed mpgOverDisp  
## <dbl> <dbl> <dbl> <dbl>  
## 1 3 15.5 318 0.0487  
## 2 4 22.8 131. 0.174   
## 3 5 19.7 145 0.136

# Section 5

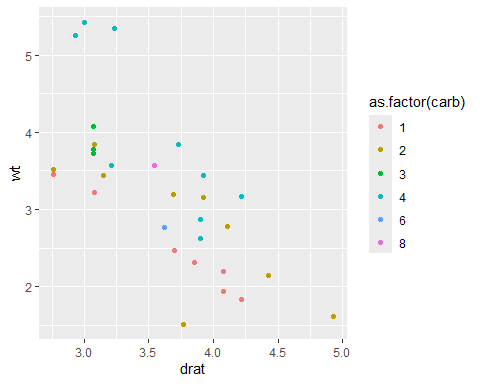
* Create a scatter plot of the drat and wt variables, and color by carb

Creating a scatter plot to show the relationship between a car’s drat and weight, while using color to indicate how many carb each automobile has.

ggplot(mtcarsTib, aes(drat, wt, col = carb)) +  
geom\_point()



ggplot(mtcarsTib, aes(drat, wt, col = as.factor(carb))) +  
geom\_point()

 # Section 6 \* Gather vs, am, gear, and carb into a single key-value pair

Reshaping the data by consolidating four different columns into a single compact format with two columns: one for variable names and one for variable values.

gather(mtcarsTib, key = "variable", value = "value", c(vs, am, gear, carb))

## # A tibble: 128 × 9  
## mpg cyl disp hp drat wt qsec variable value  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 21 6 160 110 3.9 2.62 16.5 vs 0  
## 2 21 6 160 110 3.9 2.88 17.0 vs 0  
## 3 22.8 4 108 93 3.85 2.32 18.6 vs 1  
## 4 21.4 6 258 110 3.08 3.22 19.4 vs 1  
## 5 18.7 8 360 175 3.15 3.44 17.0 vs 0  
## 6 18.1 6 225 105 2.76 3.46 20.2 vs 1  
## 7 14.3 8 360 245 3.21 3.57 15.8 vs 0  
## 8 24.4 4 147. 62 3.69 3.19 20 vs 1  
## 9 22.8 4 141. 95 3.92 3.15 22.9 vs 1  
## 10 19.2 6 168. 123 3.92 3.44 18.3 vs 1  
## # ℹ 118 more rows

# Section 7

* Iterate over each column of mtcars, returning a logical vector

The program loops through each column in the dataset, testing a condition for each value. This gives a new list of logical vectors, with each value indicating whether or not the condition was met.

map\_lgl(mtcars, ~sum(.) > 1000)

## mpg cyl disp hp drat wt qsec vs am gear carb   
## FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE