Auto-Lab-2.R

saiup

2022-10-07

library(ggplot2)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.2 ✔ forcats 0.5.2   
## ✔ purrr 0.3.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(ISLR2)  
?ISLR2::Auto

## starting httpd help server ... done

summary(Auto)

## mpg cylinders displacement horsepower weight   
## Min. : 9.00 Min. :3.000 Min. : 68.0 Min. : 46.0 Min. :1613   
## 1st Qu.:17.00 1st Qu.:4.000 1st Qu.:105.0 1st Qu.: 75.0 1st Qu.:2225   
## Median :22.75 Median :4.000 Median :151.0 Median : 93.5 Median :2804   
## Mean :23.45 Mean :5.472 Mean :194.4 Mean :104.5 Mean :2978   
## 3rd Qu.:29.00 3rd Qu.:8.000 3rd Qu.:275.8 3rd Qu.:126.0 3rd Qu.:3615   
## Max. :46.60 Max. :8.000 Max. :455.0 Max. :230.0 Max. :5140   
##   
## acceleration year origin name   
## Min. : 8.00 Min. :70.00 Min. :1.000 amc matador : 5   
## 1st Qu.:13.78 1st Qu.:73.00 1st Qu.:1.000 ford pinto : 5   
## Median :15.50 Median :76.00 Median :1.000 toyota corolla : 5   
## Mean :15.54 Mean :75.98 Mean :1.577 amc gremlin : 4   
## 3rd Qu.:17.02 3rd Qu.:79.00 3rd Qu.:2.000 amc hornet : 4   
## Max. :24.80 Max. :82.00 Max. :3.000 chevrolet chevette: 4   
## (Other) :365

str(Auto)

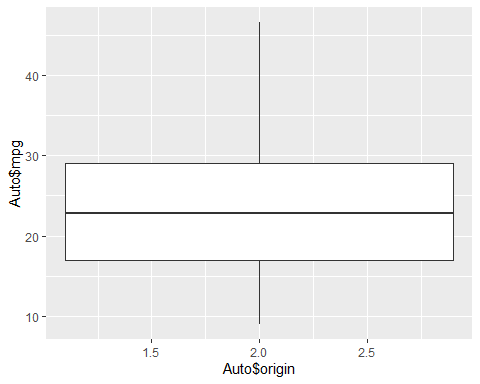
## 'data.frame': 392 obs. of 9 variables:  
## $ mpg : num 18 15 18 16 17 15 14 14 14 15 ...  
## $ cylinders : int 8 8 8 8 8 8 8 8 8 8 ...  
## $ displacement: num 307 350 318 304 302 429 454 440 455 390 ...  
## $ horsepower : int 130 165 150 150 140 198 220 215 225 190 ...  
## $ weight : int 3504 3693 3436 3433 3449 4341 4354 4312 4425 3850 ...  
## $ acceleration: num 12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...  
## $ year : int 70 70 70 70 70 70 70 70 70 70 ...  
## $ origin : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ name : Factor w/ 304 levels "amc ambassador brougham",..: 49 36 231 14 161 141 54 223 241 2 ...  
## - attr(\*, "na.action")= 'omit' Named int [1:5] 33 127 331 337 355  
## ..- attr(\*, "names")= chr [1:5] "33" "127" "331" "337" ...

Auto %>% group\_by(origin)

## # A tibble: 392 × 9  
## # Groups: origin [3]  
## mpg cylinders displacement horsepower weight accelerat…¹ year origin name   
## <dbl> <int> <dbl> <int> <int> <dbl> <int> <int> <fct>  
## 1 18 8 307 130 3504 12 70 1 chev…  
## 2 15 8 350 165 3693 11.5 70 1 buic…  
## 3 18 8 318 150 3436 11 70 1 plym…  
## 4 16 8 304 150 3433 12 70 1 amc …  
## 5 17 8 302 140 3449 10.5 70 1 ford…  
## 6 15 8 429 198 4341 10 70 1 ford…  
## 7 14 8 454 220 4354 9 70 1 chev…  
## 8 14 8 440 215 4312 8.5 70 1 plym…  
## 9 14 8 455 225 4425 10 70 1 pont…  
## 10 15 8 390 190 3850 8.5 70 1 amc …  
## # … with 382 more rows, and abbreviated variable name ¹​acceleration

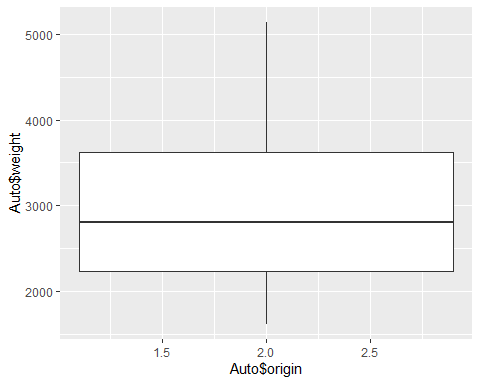
Auto %>% ggplot(aes(x = Auto$origin, y = Auto$mpg)) + geom\_boxplot()

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?



view(Auto)  
Auto %>% ggplot(aes(x = Auto$origin, y = Auto$weight)) + geom\_boxplot()

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?



origin\_model <- lm(mpg ~ origin, data = Auto)  
coef(origin\_model)

## (Intercept) origin   
## 14.811974 5.476547

larger\_model <- lm(mpg ~ weight + origin, data = Auto)  
coef(larger\_model)

## (Intercept) weight origin   
## 42.49081746 -0.00700711 1.15402780

predictionData <- data.frame(origin=c(1, 3), weight=c(3000, 3000))  
predict(larger\_model,predictionData)

## 1 2   
## 22.62352 24.93157