ForteLab3.R

andre

2022-05-25

myCars <- mtcars  
head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

head(myCars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

# 1) What is the highest hp? (The highest hp is 335)  
max(myCars$hp)

## [1] 335

# 2) Which car has the highest hp?  
index <- which.max(myCars$hp)  
row.names(myCars)[index]

## [1] "Maserati Bora"

# 3) What is the highest mpg?  
max(myCars$mpg)

## [1] 33.9

# 4) Which car has the highest mpg? (Toyota Corolla has the highest mpg.)  
row.names(myCars)[which.max(myCars$mpg)]

## [1] "Toyota Corolla"

# 5) Create a sorted dataframe, based on mpg.   
# Sort the dataframe by mpg, in descending order, and store the sorted dataframe in 'myCars\_sorted'   
myCars\_sorted <- myCars[order(-myCars$mpg), ]  
  
# 6) One method to pick an efficient car: divide the mpg value by hp, and pick the car with highest result  
# add a new column 'efficiency' in the dataframe to store the division result  
myCars$efficiency <- myCars$mpg/myCars$hp  
  
# 7) Which car has the best combination of mpg and hp?   
# get the index of maximum efficiency first, and then get its row name  
row.names(myCars)[which.max(myCars$efficiency)]

## [1] "Honda Civic"

# scale 'mpg' by subtracting its column mean and then dividing the column standard deviation  
#scale mpg first:  
scale(myCars$mpg)

## [,1]  
## [1,] 0.15088482  
## [2,] 0.15088482  
## [3,] 0.44954345  
## [4,] 0.21725341  
## [5,] -0.23073453  
## [6,] -0.33028740  
## [7,] -0.96078893  
## [8,] 0.71501778  
## [9,] 0.44954345  
## [10,] -0.14777380  
## [11,] -0.38006384  
## [12,] -0.61235388  
## [13,] -0.46302456  
## [14,] -0.81145962  
## [15,] -1.60788262  
## [16,] -1.60788262  
## [17,] -0.89442035  
## [18,] 2.04238943  
## [19,] 1.71054652  
## [20,] 2.29127162  
## [21,] 0.23384555  
## [22,] -0.76168319  
## [23,] -0.81145962  
## [24,] -1.12671039  
## [25,] -0.14777380  
## [26,] 1.19619000  
## [27,] 0.98049211  
## [28,] 1.71054652  
## [29,] -0.71190675  
## [30,] -0.06481307  
## [31,] -0.84464392  
## [32,] 0.21725341  
## attr(,"scaled:center")  
## [1] 20.09062  
## attr(,"scaled:scale")  
## [1] 6.026948

# scale 'hp' (scale is subtracting its column mean and then dividing its column’s standard deviation.   
# But you just use scale function like this:)  
scale(myCars$hp)

## [,1]  
## [1,] -0.53509284  
## [2,] -0.53509284  
## [3,] -0.78304046  
## [4,] -0.53509284  
## [5,] 0.41294217  
## [6,] -0.60801861  
## [7,] 1.43390296  
## [8,] -1.23518023  
## [9,] -0.75387015  
## [10,] -0.34548584  
## [11,] -0.34548584  
## [12,] 0.48586794  
## [13,] 0.48586794  
## [14,] 0.48586794  
## [15,] 0.85049680  
## [16,] 0.99634834  
## [17,] 1.21512565  
## [18,] -1.17683962  
## [19,] -1.38103178  
## [20,] -1.19142477  
## [21,] -0.72469984  
## [22,] 0.04831332  
## [23,] 0.04831332  
## [24,] 1.43390296  
## [25,] 0.41294217  
## [26,] -1.17683962  
## [27,] -0.81221077  
## [28,] -0.49133738  
## [29,] 1.71102089  
## [30,] 0.41294217  
## [31,] 2.74656682  
## [32,] -0.54967799  
## attr(,"scaled:center")  
## [1] 146.6875  
## attr(,"scaled:scale")  
## [1] 68.56287

# You just created two new variable using mpg and hp using scale function. Add the two scaled data and save the   
# result as a new column 'combination' in the dataframe. Populate XXXX below;  
myCars$combination <- scale(myCars$mpg) + scale(myCars$hp)  
  
# Get the index of maximum combination first, and then get its row name. HINT: use which.max function AND row.names function  
row.names(myCars)[which.max(myCars$combination)]

## [1] "Maserati Bora"