# Ryan Timbrook

## **Applied Data Science**

## **IST687 Intro to Data Science**, Spring 2019

## **Due Date:** 04/16/2019

## **Homework:** 2

### **NetID**: RTIMBROO

### **SUID**: 386792749

## #R Code - unexecuted

## # ---------- HW2: Explore the mtcars dataset -----------

**# Homework Week 2 Objective: Explore the mtcars dataset**

## Copy the mtcars dataset into a new variable called myCars

myCars <- mtcars

str(myCars)

summary(myCars)

row.names(myCars)

**#Step 1: What is the hp**

## Q1: What is the highest hp?

maxHp <- max(myCars$hp)

maxHp

**## Q2: Which car has the highest hp?**

carMaxHp <- myCars[myCars$hp == max(myCars$hp),]

carNameMaxHp <- row.names(carMaxHp)

carNameMaxHp

---------------------------------------

**#Step 2: Explore mpg**

**## Q3: What is the highest mpg?**

maxMPG <- max(myCars$mpg)

maxMPG

**## Q4: Which car has the highest mpg?**

carMaxMPG <- myCars[myCars$mpg == max(myCars$mpg),]

carMaxMPG

carNameMaxMPG <- row.names(carMaxMPG)

carNameMaxMPG

**## Q5: Create a sorted dataframe, based on mpg**

sortedMyCars <- myCars[order(myCars$mpg),]

sortedMyCars

#---------------------------------------

**#Step 3; Which car has the "best" combination of mpg and hp?**

**## Q6: What logic did you use?**

carsByMPGAndHP <- data.frame(sortedMyCars$mpg,sortedMyCars$hp,row.names = row.names(sortedMyCars))

colnames(carsByMPGAndHP) <- c('mpg','hp')

carsByMPGAndHP$eff <- carsByMPGAndHP$mpg/carsByMPGAndHP$hp

carBestEff <- carsByMPGAndHP[carsByMPGAndHP$eff == max(carsByMPGAndHP$eff),]

carBestEff

**## Q7: Which car?**

carNameBestEff <- row.names(carBestEff)

carNameBestEff

#---------------------------------------

**#Step 4: Which car has "best" car combination of mpg and hp, where mpg and hp must be given equal weight?**

hist(carsByMPGAndHP$mpg)

mpg.z <- scale(carsByMPGAndHP$mpg)

hist(mpg.z)

hist(carsByMPGAndHP$hp)

hp.z <- scale(carsByMPGAndHP$hp)

hist(hp.z)

eff.z <- mpg.z/hp.z

hist(eff.z)

scaledBestEff <- data.frame(mpg.z,hp.z,eff.z, row.names = row.names(carsByMPGAndHP))

scaledBestEff

carScaledBestEff <- scaledBestEff[scaledBestEff$eff.z == max(scaledBestEff$eff.z),]

carScaledBestEff

carNameScaledBestEff <- row.names(carScaledBestEff)

carNameScaledBestEff

## #R Code – executed

> # Homework Week 2 Objective: Explore the mtcars dataset

> ## Copy the mtcars dataset into a new variable called myCars

> myCars <- mtcars

> str(myCars)

'data.frame': 32 obs. of 11 variables:

$ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...

$ cyl : num 6 6 4 6 8 6 8 4 4 6 ...

$ disp: num 160 160 108 258 360 ...

$ hp : num 110 110 93 110 175 105 245 62 95 123 ...

$ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...

$ wt : num 2.62 2.88 2.32 3.21 3.44 ...

$ qsec: num 16.5 17 18.6 19.4 17 ...

$ vs : num 0 0 1 1 0 1 0 1 1 1 ...

$ am : num 1 1 1 0 0 0 0 0 0 0 ...

$ gear: num 4 4 4 3 3 3 3 4 4 4 ...

$ carb: num 4 4 1 1 2 1 4 2 2 4 ...

> summary(myCars)

mpg cyl disp hp drat

Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0 Min. :2.760

1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5 1st Qu.:3.080

Median :19.20 Median :6.000 Median :196.3 Median :123.0 Median :3.695

Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7 Mean :3.597

3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0 3rd Qu.:3.920

Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0 Max. :4.930

wt qsec vs am gear

Min. :1.513 Min. :14.50 Min. :0.0000 Min. :0.0000 Min. :3.000

1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:3.000

Median :3.325 Median :17.71 Median :0.0000 Median :0.0000 Median :4.000

Mean :3.217 Mean :17.85 Mean :0.4375 Mean :0.4062 Mean :3.688

3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:4.000

Max. :5.424 Max. :22.90 Max. :1.0000 Max. :1.0000 Max. :5.000

carb

Min. :1.000

1st Qu.:2.000

Median :2.000

Mean :2.812

3rd Qu.:4.000

Max. :8.000

> row.names(myCars)

[1] "Mazda RX4" "Mazda RX4 Wag" "Datsun 710"

[4] "Hornet 4 Drive" "Hornet Sportabout" "Valiant"

[7] "Duster 360" "Merc 240D" "Merc 230"

[10] "Merc 280" "Merc 280C" "Merc 450SE"

[13] "Merc 450SL" "Merc 450SLC" "Cadillac Fleetwood"

[16] "Lincoln Continental" "Chrysler Imperial" "Fiat 128"

[19] "Honda Civic" "Toyota Corolla" "Toyota Corona"

[22] "Dodge Challenger" "AMC Javelin" "Camaro Z28"

[25] "Pontiac Firebird" "Fiat X1-9" "Porsche 914-2"

[28] "Lotus Europa" "Ford Pantera L" "Ferrari Dino"

[31] "Maserati Bora" "Volvo 142E"

>

> #Step 1: What is the hp

> ## Q1: What is the highest hp?

> maxHp <- max(myCars$hp)

> maxHp

[1] 335

>

> ## Q2: Which car has the highest hp?

> carMaxHp <- myCars[myCars$hp == max(myCars$hp),]

> carNameMaxHp <- row.names(carMaxHp)

> carNameMaxHp

[1] "Maserati Bora"

>

> #---------------------------------------

> #Step 2: Explore mpg

>

> ## Q3: What is the highest mpg?

> maxMPG <- max(myCars$mpg)

> maxMPG

[1] 33.9

>

> ## Q4: Which car has the highest mpg?

> carMaxMPG <- myCars[myCars$mpg == max(myCars$mpg),]

> carMaxMPG

mpg cyl disp hp drat wt qsec vs am gear carb

Toyota Corolla 33.9 4 71.1 65 4.22 1.835 19.9 1 1 4 1

> carNameMaxMPG <- row.names(carMaxMPG)

> carNameMaxMPG

[1] "Toyota Corolla"

>

> ## Q5: Create a sorted dataframe, based on mpg

> sortedMyCars <- myCars[order(myCars$mpg),]

> sortedMyCars

mpg cyl disp hp drat wt qsec vs am gear carb

Cadillac Fleetwood 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3 4

Lincoln Continental 10.4 8 460.0 215 3.00 5.424 17.82 0 0 3 4

Camaro Z28 13.3 8 350.0 245 3.73 3.840 15.41 0 0 3 4

Duster 360 14.3 8 360.0 245 3.21 3.570 15.84 0 0 3 4

Chrysler Imperial 14.7 8 440.0 230 3.23 5.345 17.42 0 0 3 4

Maserati Bora 15.0 8 301.0 335 3.54 3.570 14.60 0 1 5 8

Merc 450SLC 15.2 8 275.8 180 3.07 3.780 18.00 0 0 3 3

AMC Javelin 15.2 8 304.0 150 3.15 3.435 17.30 0 0 3 2

Dodge Challenger 15.5 8 318.0 150 2.76 3.520 16.87 0 0 3 2

Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5 4

Merc 450SE 16.4 8 275.8 180 3.07 4.070 17.40 0 0 3 3

Merc 450SL 17.3 8 275.8 180 3.07 3.730 17.60 0 0 3 3

Merc 280C 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4

Valiant 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3 1

Hornet Sportabout 18.7 8 360.0 175 3.15 3.440 17.02 0 0 3 2

Merc 280 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4

Pontiac Firebird 19.2 8 400.0 175 3.08 3.845 17.05 0 0 3 2

Ferrari Dino 19.7 6 145.0 175 3.62 2.770 15.50 0 1 5 6

Mazda RX4 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4 4

Mazda RX4 Wag 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4 4

Hornet 4 Drive 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3 1

Volvo 142E 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2

Toyota Corona 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1

Datsun 710 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1

Merc 230 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2

Merc 240D 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2

Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5 2

Fiat X1-9 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1

Honda Civic 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2

Lotus Europa 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2

Fiat 128 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1

Toyota Corolla 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1

>

> #---------------------------------------

> #Step 3; Which car has the "best" combination of mpg and hp?

> ## Q6: What logic did you use?

> carsByMPGAndHP <- data.frame(sortedMyCars$mpg,sortedMyCars$hp,row.names = row.names(sortedMyCars))

> colnames(carsByMPGAndHP) <- c('mpg','hp')

> carsByMPGAndHP$eff <- carsByMPGAndHP$mpg/carsByMPGAndHP$hp

> carBestEff <- carsByMPGAndHP[carsByMPGAndHP$eff == max(carsByMPGAndHP$eff),]

> carBestEff

mpg hp eff

Honda Civic 30.4 52 0.5846154

> ## Q7: Which car?

> carNameBestEff <- row.names(carBestEff)

> carNameBestEff

[1] "Honda Civic"

>

> #---------------------------------------

> #Step 4: Which car has "best" car combination of mpg and hp, where mpg and hp must be given equal weight?

> hist(carsByMPGAndHP$mpg)

> mpg.z <- scale(carsByMPGAndHP$mpg)

> #hist(mpg.z)

>

> hist(carsByMPGAndHP$hp)

> hp.z <- scale(carsByMPGAndHP$hp)

> #hist(hp.z)

>

> eff.z <- mpg.z/hp.z

> #hist(eff.z)

>

> scaledBestEff <- data.frame(mpg.z,hp.z,eff.z, row.names = row.names(carsByMPGAndHP))

> #scaledBestEff

>

> carScaledBestEff <- scaledBestEff[scaledBestEff$eff.z == max(scaledBestEff$eff.z),]

> carScaledBestEff

mpg.z hp.z eff.z

Merc 280C -0.3800638 -0.3454858 1.100085

>

> carNameScaledBestEff <- row.names(carScaledBestEff)

> carNameScaledBestEff

[1] "Merc 280C"

>

> carsByMPGAndHP[carNameScaledBestEff,]

mpg hp eff

Merc 280C 17.8 123 0.1447154