# ---------- Lab2: Manipulating Data frames ----------

# copy original dataframe into a new one

myCars <- mtcars

myCars

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

# Step 1: What is the hp (hp stands for âhorse powerâ)

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

# max() function returns the maximum of the input values, which is the 'hp' column of dataframe 'myCars'

max(myCars$hp)

# 2) Which car has the highest hp? (Maserati Bora has the highest hp.)

# get the index (row number) of the maximum number in 'hp' column

index <- which.max(myCars$hp)

# get the row name of this row in datafrome 'myCars'

row.names(myCars)[index]

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

#Step 2: Explore mpg (mpg stands for âmiles per gallonâ)

# 3) What is the highest mpg? (The highest mpg is 33.9)

# return the maximum mpg

max(myCars$mpg)

# 4) Which car has the highest mpg? (Toyota Corolla has the highest mpg.)

# get the index of maximum mpg first, and then get its row name

row.names(myCars)[which.max(myCars$mpg)]

# 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),]

# display the sorted dataframe

myCars\_sorted

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

#Step 3: Which car has the best combination of mpg and hp?

# 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

# to find the maximum of this new created column

max(myCars$efficiency)

# 7) Which car?

# get the index of maximum efficiency first, and then get its row name

row.names(myCars)[which.max(myCars$efficiency)]

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

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

# scale 'mpg' by subtracting its column mean and then dividing the columns standard deviation

#

# https://stackoverflow.com/questions/20256028/understanding-scale-in-r

#

scale(myCars$mpg)

# scale 'hp' by subtracting its column mean and then dividing its columnâs standard deviation

scale(myCars$hp)

# summarize the two scaled data and save the result as a new column 'combination' in the dataframe

myCars$combination <- scale(myCars$mpg) + scale(myCars$hp)

# get the index of maximum combination first, and then get its row name

row.names(myCars)[which.max(myCars$combination)]