Alonso\_Week 2 Homework Assignment

Martin Alonso

7/17/2018

IST687 Introduction to Data Science: Week 2 Homework.

Explore the mtcars dataset (which is already included in R).

### Part 1: The Code

#For this exercise we'll use the dplyr and scales packages.  
require(dplyr)  
require(scales)  
  
#Copy the mtcars dataset into a new variable (called it myCars), so that if you mess up, you can start again very easily (by copying mtcars into myCars again).  
cars <- rownames(mtcars)  
myCars <- cbind(cars, mtcars)  
  
#Step 1: What is the hp (hp stands for "horse power".)  
#1. What is the highest hp?  
max(myCars$hp)  
  
#2. Which car has the highest hp?  
myCars %>% filter(hp == max(hp)) #'%>%' is known as a pipe operator that chains functions.   
  
#Step 2: Explore mpg (mpg stands for "miles per gallon".)  
#3. What is the highest mpg?  
max(myCars$mpg)  
  
#4. Which car has the highest mpg?  
myCars %>% filter(mpg == max(mpg))  
  
#5. Create a sorted dataframe, based on mpg  
myCars %>% arrange(mpg) %>% head()  
  
#Step 3: Which car has the "best" combination of mpg and hp?  
#6. The rescale package will scale mpg and hp to 100, so both columns can be averaged on an equal footing.  
myCarsBest <- myCars %>% mutate(  
 mpg\_rescale = rescale(mpg, to = c(0,100)), #Rescale mpg and hp from 0 to 100.  
 hp\_rescale = rescale(hp, to = c(0,100)),   
 rank = (mpg\_rescale + hp\_rescale)/2 #Calculate the average.   
)  
  
myCarsBest %>% head(n=3) #Show the best 3 cars.  
  
myCarsBest %>% filter(rank == max(rank))  
  
#7. The "best" car using this formula is the Maserati Bora.  
  
#Step 4: Which car has the "best" combination of mpg and hp, where mpg and hp must be given an equal weight?  
myCarsEqual <- myCars %>%   
 mutate(weighted\_mean = mpg/mean(mpg) \* hp/mean(hp)) %>% #calculates how far above/below the mean the value is.   
 ungroup()  
  
myCarsEqual %>% head(n=3)  
  
myCarsEqual %>%   
 select(cars, mpg, hp, weighted\_mean) %>%   
 filter(weighted\_mean == max(weighted\_mean))  
  
#Giving similar weights to both mpg and hp, the Maserati Bora is still the best car available.

### Part 2: Running the Code

#For this exercise we'll use the dplyr and scales packages.  
require(dplyr)  
require(scales)  
  
#Copy the mtcars dataset into a new variable (called it myCars), so that if you mess up, you can start again very easily (by copying mtcars into myCars again).  
cars <- rownames(mtcars)  
myCars <- cbind(cars, mtcars)  
  
#Step 1: What is the hp (hp stands for "horse power".)  
#1. What is the highest hp?  
max(myCars$hp)

## [1] 335

#2. Which car has the highest hp?  
myCars %>% filter(hp == max(hp)) #'%>%' is known as a pipe operator that chains functions.

## cars mpg cyl disp hp drat wt qsec vs am gear carb  
## 1 Maserati Bora 15 8 301 335 3.54 3.57 14.6 0 1 5 8

#Step 2: Explore mpg (mpg stands for "miles per gallon".)  
#3. What is the highest mpg?  
max(myCars$mpg)

## [1] 33.9

#4. Which car has the highest mpg?  
myCars %>% filter(mpg == max(mpg))

## cars mpg cyl disp hp drat wt qsec vs am gear carb  
## 1 Toyota Corolla 33.9 4 71.1 65 4.22 1.835 19.9 1 1 4 1

#5. Create a sorted dataframe, based on mpg  
myCars %>% arrange(mpg) %>% head()

## cars mpg cyl disp hp drat wt qsec vs am gear carb  
## 1 Cadillac Fleetwood 10.4 8 472 205 2.93 5.250 17.98 0 0 3 4  
## 2 Lincoln Continental 10.4 8 460 215 3.00 5.424 17.82 0 0 3 4  
## 3 Camaro Z28 13.3 8 350 245 3.73 3.840 15.41 0 0 3 4  
## 4 Duster 360 14.3 8 360 245 3.21 3.570 15.84 0 0 3 4  
## 5 Chrysler Imperial 14.7 8 440 230 3.23 5.345 17.42 0 0 3 4  
## 6 Maserati Bora 15.0 8 301 335 3.54 3.570 14.60 0 1 5 8

#Step 3: Which car has the "best" combination of mpg and hp?  
#6. The rescale package will scale mpg and hp to 100, so both columns can be averaged on an equal footing.  
myCarsBest <- myCars %>% mutate(  
 mpg\_rescale = rescale(mpg, to = c(0,100)), #Rescale mpg and hp from 0 to 100.  
 hp\_rescale = rescale(hp, to = c(0,100)),   
 rank = (mpg\_rescale + hp\_rescale)/2 #Calculate the average.   
)  
  
myCarsBest %>% head(n=3) #Show the best 3 cars.

## cars mpg cyl disp hp drat wt qsec vs am gear carb  
## 1 Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## 2 Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## 3 Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## mpg\_rescale hp\_rescale rank  
## 1 45.10638 20.49470 32.80054  
## 2 45.10638 20.49470 32.80054  
## 3 52.76596 14.48763 33.62679

myCarsBest %>% filter(rank == max(rank))

## cars mpg cyl disp hp drat wt qsec vs am gear carb  
## 1 Maserati Bora 15 8 301 335 3.54 3.57 14.6 0 1 5 8  
## mpg\_rescale hp\_rescale rank  
## 1 19.57447 100 59.78723

#7. The "best" car using this formula is the Maserati Bora.  
  
#Step 4: Which car has the "best" combination of mpg and hp, where mpg and hp must be given an equal weight?  
myCarsEqual <- myCars %>%   
 mutate(weighted\_mean = mpg/mean(mpg) \* hp/mean(hp)) %>% #calculates how far above/below the mean the value is.   
 ungroup()  
  
myCarsEqual %>% head(n=3)

## cars mpg cyl disp hp drat wt qsec vs am gear carb  
## 1 Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## 2 Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## 3 Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## weighted\_mean  
## 1 0.7838364  
## 2 0.7838364  
## 3 0.7195007

myCarsEqual %>%   
 select(cars, mpg, hp, weighted\_mean) %>%   
 filter(weighted\_mean == max(weighted\_mean))

## cars mpg hp weighted\_mean  
## 1 Maserati Bora 15 335 1.705099

#Giving similar weights to both mpg and hp, the Maserati Bora is still the best car available.

### Sources

#### 1. Stack Exchange, <https://stats.stackexchange.com/questions/25894/changing-the-scale-of-a-variable-to-0-100>, 2018-07-17