murphy-week5lab2.R

William Murphy

Wed Oct 3 12:48:16 2018

getwd()

## [1] "/home/manwthglasses/Documents/IntroToDataScience"

setwd("/home/manwthglasses/Documents/IntroToDataScience/")  
# Task 1  
# copy original dataframe into a new one: my\_mtcars  
# mtcars object is one of many built-in data sets in R. So you do not need to worry about creating mtcars.  
my\_mtcars <- mtcars  
# 1: investigate my\_mtcars using str function. How many variables and observations are included in this dataframe?  
str(my\_mtcars)

## '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 ...

# 2: calculate engine displacement per cylinder and save it as a new variable 'UnitEngine' in the dataframe. Populate the two XXXX below  
my\_mtcars$UnitEngine <- my\_mtcars$disp/my\_mtcars$cyl  
# 3. summarize the new variable 'UnitEngine': use summary function  
summary(my\_mtcars)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb UnitEngine   
## Min. :0.0000 Min. :3.000 Min. :1.000 Min. :17.77   
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:26.92   
## Median :0.0000 Median :4.000 Median :2.000 Median :34.48   
## Mean :0.4062 Mean :3.688 Mean :2.812 Mean :35.03   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:43.19   
## Max. :1.0000 Max. :5.000 Max. :8.000 Max. :59.00

# Task 2  
# 4. create a numeric vector 'Pets' with this numbers (1,1,1,0,0)  
Pets <- c(1,1,1,0,0)  
# 5. create a numeric vector 'Order' with these numbers (3,1,2,3,3)  
Order <- c(3,1,2,3,3)  
# create a numeric vector 'Siblings'  
Siblings <- c(0,3,5,0,0)  
# create a numeric vector 'IDs'  
IDs <- c(1,2,3,4,5)  
# 6. Combine those four numeric vectors together into a dataframe called 'myFriends'. You must use data.frame function  
myFriends <- data.frame(Pets, Order, Siblings, IDs)  
# 7. report the structure of the dataframe  
str(myFriends)

## 'data.frame': 5 obs. of 4 variables:  
## $ Pets : num 1 1 1 0 0  
## $ Order : num 3 1 2 3 3  
## $ Siblings: num 0 3 5 0 0  
## $ IDs : num 1 2 3 4 5

# 8. summarize the dataframe. Use summary function  
summary(myFriends)

## Pets Order Siblings IDs   
## Min. :0.0 Min. :1.0 Min. :0.0 Min. :1   
## 1st Qu.:0.0 1st Qu.:2.0 1st Qu.:0.0 1st Qu.:2   
## Median :1.0 Median :3.0 Median :0.0 Median :3   
## Mean :0.6 Mean :2.4 Mean :1.6 Mean :3   
## 3rd Qu.:1.0 3rd Qu.:3.0 3rd Qu.:3.0 3rd Qu.:4   
## Max. :1.0 Max. :3.0 Max. :5.0 Max. :5

# list (or print) all of the values for 'IDs' variable in the dataframe  
myFriends$IDs

## [1] 1 2 3 4 5

# list all of the values for 'Pets' variable in the dataframe  
myFriends$Pets

## [1] 1 1 1 0 0

# 9. list all of the values for 'Order' variable in the dataframe  
myFriends$Order

## [1] 3 1 2 3 3

# list all of the values for 'Siblings' variable in the dataframe  
myFriends$Siblings

## [1] 0 3 5 0 0

# 10. write a code to print the values in the fifth observation of the Pets variable  
print(myFriends$Pets[5])

## [1] 0

# 11. add a vector called 'age' to 'myFriends' using cbind function. \*\*\* YOU MUST USE cbind FUNCTION to receive full grades.  
age <-c(23, 21, 45, 21, 18)  
cbind(myFriends, age)

## Pets Order Siblings IDs age  
## 1 1 3 0 1 23  
## 2 1 1 3 2 21  
## 3 1 2 5 3 45  
## 4 0 3 0 4 21  
## 5 0 3 0 5 18

# 12. define a vector called 'names' by including all the names in a vector. Add a vector 'names' to 'myFriends' using cbind function. Print the structure of 'myFriends'. What is the data type (among: factor, numeric, logical, string) of the 'names'?  
names <- c("Pets", "Order", "Siblings", "IDs", "age")  
cbind(names, myFriends)

## names Pets Order Siblings IDs  
## 1 Pets 1 3 0 1  
## 2 Order 1 1 3 2  
## 3 Siblings 1 2 5 3  
## 4 IDs 0 3 0 4  
## 5 age 0 3 0 5