Alonso\_Week 2 Lab Assignment

Martin Alonso

7/17/2018

IST687 Introduction to Data Science: Week 2 Lab.

The “dataframe” is one of the most essential data structures used in R. It is conceptually equivalent to a database “relation” and to the typical rectangular dataset with variables as columns and cases as rows. For this activity, you will gain some skill with manipulating a dataframe.

# Task 1

Using the mtcars dataset, create a new column that measures engine displacement per cylinder in cubic inches.

mt\_mycars <- mtcars  
  
mt\_mycars$disp\_per\_cyl <- mt\_mycars$disp/mt\_mycars$cyl  
  
summary(mt\_mycars$disp\_per\_cyl)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 17.77 26.92 34.48 35.03 43.19 59.00

# Task 2

Create four vectors of friends or family members which include the person ID, the number of Pets, the order of their birth, and the number of siblings. Group the vectors into a data frame called myFriends.

personID <- c('Wife','Dad','Mum','Daughter','Mother-in-law')  
Pets <- c(1, 7, 0, 0, 1)  
Order <- c(1, 3, 1, 1, 2)  
Siblings <- c(0, 2, 3, 0, 1)  
  
myFriends <- data.frame(personID, Pets, Order, Siblings)

Show the structure of myFriends…

str(myFriends)

## 'data.frame': 5 obs. of 4 variables:  
## $ personID: Factor w/ 5 levels "Dad","Daughter",..: 5 1 4 2 3  
## $ Pets : num 1 7 0 0 1  
## $ Order : num 1 3 1 1 2  
## $ Siblings: num 0 2 3 0 1

…and the summary.

summary(myFriends)

## personID Pets Order Siblings   
## Dad :1 Min. :0.0 Min. :1.0 Min. :0.0   
## Daughter :1 1st Qu.:0.0 1st Qu.:1.0 1st Qu.:0.0   
## Mother-in-law:1 Median :1.0 Median :1.0 Median :1.0   
## Mum :1 Mean :1.8 Mean :1.6 Mean :1.2   
## Wife :1 3rd Qu.:1.0 3rd Qu.:2.0 3rd Qu.:2.0   
## Max. :7.0 Max. :3.0 Max. :3.0

Finally, display the values using the $ sign.

myFriends$personID

## [1] Wife Dad Mum Daughter Mother-in-law  
## Levels: Dad Daughter Mother-in-law Mum Wife

myFriends$Pets

## [1] 1 7 0 0 1

myFriends$Order

## [1] 1 3 1 1 2

myFriends$Siblings

## [1] 0 2 3 0 1