murphy-lab7.R

William Murphy

Fri Nov 30 13:55:17 2018

setwd("/home/manwthglasses/Documents/IntroToDataScience/")  
  
##Activating the sqldf library  
library(sqldf)

## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite

##Making sure built-in data set airquality is available, and seeing first few values  
print(head(airquality))

## Ozone Solar.R Wind Temp Month Day  
## 1 41 190 7.4 67 5 1  
## 2 36 118 8.0 72 5 2  
## 3 12 149 12.6 74 5 3  
## 4 18 313 11.5 62 5 4  
## 5 NA NA 14.3 56 5 5  
## 6 28 NA 14.9 66 5 6

##Copies airquality into air object  
air <- airquality  
##Checks the object type of air  
class(air)

## [1] "data.frame"

##  
average\_ozone <- sqldf("Select AVG(Ozone) From air")  
print(average\_ozone)

## AVG(Ozone)  
## 1 42.12931

sqldf("Select Ozone From air Where Ozone > (Select AVG(Ozone) From air)")

## Ozone  
## 1 45  
## 2 115  
## 3 71  
## 4 135  
## 5 49  
## 6 64  
## 7 77  
## 8 97  
## 9 97  
## 10 85  
## 11 48  
## 12 61  
## 13 79  
## 14 63  
## 15 80  
## 16 108  
## 17 52  
## 18 82  
## 19 50  
## 20 64  
## 21 59  
## 22 78  
## 23 66  
## 24 122  
## 25 89  
## 26 110  
## 27 44  
## 28 65  
## 29 59  
## 30 44  
## 31 45  
## 32 168  
## 33 73  
## 34 76  
## 35 118  
## 36 84  
## 37 85  
## 38 96  
## 39 78  
## 40 73  
## 41 91  
## 42 47  
## 43 44  
## 44 46

newAQ <- sqldf("Select Ozone From air Where Ozone > (Select AVG(Ozone) From air)")  
class(newAQ)

## [1] "data.frame"

nrow(newAQ)

## [1] 44

ncol(newAQ)

## [1] 1

head(newAQ)

## Ozone  
## 1 45  
## 2 115  
## 3 71  
## 4 135  
## 5 49  
## 6 64

# (a) Exclude Missing Values from calculating "Ozone" mean and assign the result to "average\_ozone": Hint:use na.rm  
average\_ozone <- mean(air$Ozone, na.rm = TRUE)  
# (b) print the result (average\_ozone)  
print(average\_ozone)

## [1] 42.12931

# Repeat step 5  
# (c) select rows with bigger values than the average ozone value   
######wrong approach: data$Ozone > meanOzone   
#[1] FALSE FALSE FALSE FALSE NA FALSE FALSE FALSE FALSE NA FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
#[23] FALSE FALSE  
###################  
##tapply(air$Ozone, air$Ozone>average\_ozone , data.frame)$'TRUE'  
na.exclude(air[air$Ozone>average\_ozone,])

## Ozone Solar.R Wind Temp Month Day  
## 29 45 252 14.9 81 5 29  
## 30 115 223 5.7 79 5 30  
## 40 71 291 13.8 90 6 9  
## 62 135 269 4.1 84 7 1  
## 63 49 248 9.2 85 7 2  
## 66 64 175 4.6 83 7 5  
## 68 77 276 5.1 88 7 7  
## 69 97 267 6.3 92 7 8  
## 70 97 272 5.7 92 7 9  
## 71 85 175 7.4 89 7 10  
## 77 48 260 6.9 81 7 16  
## 79 61 285 6.3 84 7 18  
## 80 79 187 5.1 87 7 19  
## 81 63 220 11.5 85 7 20  
## 85 80 294 8.6 86 7 24  
## 86 108 223 8.0 85 7 25  
## 88 52 82 12.0 86 7 27  
## 89 82 213 7.4 88 7 28  
## 90 50 275 7.4 86 7 29  
## 91 64 253 7.4 83 7 30  
## 92 59 254 9.2 81 7 31  
## 99 122 255 4.0 89 8 7  
## 100 89 229 10.3 90 8 8  
## 101 110 207 8.0 90 8 9  
## 104 44 192 11.5 86 8 12  
## 106 65 157 9.7 80 8 14  
## 109 59 51 6.3 79 8 17  
## 112 44 190 10.3 78 8 20  
## 116 45 212 9.7 79 8 24  
## 117 168 238 3.4 81 8 25  
## 118 73 215 8.0 86 8 26  
## 120 76 203 9.7 97 8 28  
## 121 118 225 2.3 94 8 29  
## 122 84 237 6.3 96 8 30  
## 123 85 188 6.3 94 8 31  
## 124 96 167 6.9 91 9 1  
## 125 78 197 5.1 92 9 2  
## 126 73 183 2.8 93 9 3  
## 127 91 189 4.6 93 9 4  
## 128 47 95 7.4 87 9 5  
## 134 44 236 14.9 81 9 11  
## 139 46 237 6.9 78 9 16

# (d) Repeat step 6  
# only keep the rows in which the Ozone values are higher than the average, and write the result table into a new R data object called "newAQ2"  
newAQ2 <- na.exclude(air[air$Ozone>average\_ozone,])  
# (e) reveal what type of object newAQ2 is  
class(newAQ2)

## [1] "data.frame"

# (f) reveal the number of rows, then reveal the number of columns  
nrow(newAQ2)

## [1] 42

ncol(newAQ2)

## [1] 6

# (g) show the first few rows of "newAQ2  
head(newAQ2)

## Ozone Solar.R Wind Temp Month Day  
## 29 45 252 14.9 81 5 29  
## 30 115 223 5.7 79 5 30  
## 40 71 291 13.8 90 6 9  
## 62 135 269 4.1 84 7 1  
## 63 49 248 9.2 85 7 2  
## 66 64 175 4.6 83 7 5