lab1

m.pryidun

2022-11-10

#Task 1. Exploraring a dataset

airquality <- read.csv("D:/KNU/DataProcSoft/Data-processing-software/lab1/airquality.csv")  
View(airquality)

#Question 1. What are the column names of the data frame?

cols <- colnames(airquality)   
cols

## [1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"

#Question 2. What are the row names of the data frame?

rows <- rownames(airquality)  
rows

## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12"   
## [13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22" "23" "24"   
## [25] "25" "26" "27" "28" "29" "30" "31" "32" "33" "34" "35" "36"   
## [37] "37" "38" "39" "40" "41" "42" "43" "44" "45" "46" "47" "48"   
## [49] "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"   
## [61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72"   
## [73] "73" "74" "75" "76" "77" "78" "79" "80" "81" "82" "83" "84"   
## [85] "85" "86" "87" "88" "89" "90" "91" "92" "93" "94" "95" "96"   
## [97] "97" "98" "99" "100" "101" "102" "103" "104" "105" "106" "107" "108"  
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"  
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"  
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"  
## [145] "145" "146" "147" "148" "149" "150" "151" "152" "153"

#Question 3. Extract the first 6 rows of the data frame and print them to the console

head6 <- head(airquality,n=6)  
print(head6)

## 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

#Question 4. How many observations (i.e. rows) are in this data frame?

nrow(airquality)

## [1] 153

#Question 5. Extract the last 6 rows of the data frame and print them to the console

tail6 <- tail(airquality,n=6)  
print(tail6)

## Ozone Solar.R Wind Temp Month Day  
## 148 14 20 16.6 63 9 25  
## 149 30 193 6.9 70 9 26  
## 150 NA 145 13.2 77 9 27  
## 151 14 191 14.3 75 9 28  
## 152 18 131 8.0 76 9 29  
## 153 20 223 11.5 68 9 30

#Question 6. How many missing values are in the “Ozone” column of this data frame?

which(is.na(airquality$Ozone))

## [1] 5 10 25 26 27 32 33 34 35 36 37 39 42 43 45 46 52 53 54  
## [20] 55 56 57 58 59 60 61 65 72 75 83 84 102 103 107 115 119 150

#Question 7. What is the mean of the “Ozone” column in this dataset? Exclude missing values (coded as NA) from this calculation.

mean ( airquality$Ozone, na.rm = TRUE )

## [1] 42.12931

#Question 8. Extract the subset of rows of the data frame where Ozone values are above 31 and Temp values are above 90.

identify\_rows <- airquality$Ozone > 31 & airquality$Ozone < 90  
identify\_rows

## [1] TRUE TRUE FALSE FALSE NA FALSE FALSE FALSE FALSE NA FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  
## [25] NA NA NA FALSE TRUE FALSE TRUE NA NA NA NA NA  
## [37] NA FALSE NA TRUE TRUE NA NA FALSE NA NA FALSE TRUE  
## [49] FALSE FALSE FALSE NA NA NA NA NA NA NA NA NA  
## [61] NA FALSE TRUE TRUE NA TRUE TRUE TRUE FALSE FALSE TRUE NA  
## [73] FALSE FALSE NA FALSE TRUE TRUE TRUE TRUE TRUE FALSE NA NA  
## [85] TRUE FALSE FALSE TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE TRUE  
## [97] TRUE TRUE FALSE TRUE FALSE NA NA TRUE FALSE TRUE NA FALSE  
## [109] TRUE FALSE FALSE TRUE FALSE FALSE NA TRUE FALSE TRUE NA TRUE  
## [121] FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE TRUE FALSE FALSE FALSE  
## [133] FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE  
## [145] FALSE TRUE FALSE FALSE FALSE NA FALSE FALSE FALSE

#Question 9. Use a for loop to create a vector of length 6 containing the mean of each column in the data frame (excluding all missing values).

means<-c()  
for(i in 1:ncol(airquality)) {  
 means<-append(means, mean ( airquality[,i], na.rm = TRUE ))  
}  
means

## [1] 42.129310 185.931507 9.957516 77.882353 6.993464 15.803922

#Question 10. Use the apply function to calculate the standard deviation of each column in the data frame (excluding all missing values).

apply(airquality, 2, sd)

## Ozone Solar.R Wind Temp Month Day   
## NA NA 3.523001 9.465270 1.416522 8.864520

#Question 11. Calculate the mean of “Ozone” for each Month in the data frame and create a vector containing the monthly means (exclude all missing values).

tapply(airquality$Ozone, airquality$Month, mean, na.rm = TRUE)

## 5 6 7 8 9   
## 23.61538 29.44444 59.11538 59.96154 31.44828

#Question 12. Draw a random sample of 5 rows from the data frame

set.seed(1)  
airquality[sample(nrow(airquality), 5), ]

## Ozone Solar.R Wind Temp Month Day  
## 68 77 276 5.1 88 7 7  
## 129 32 92 15.5 84 9 6  
## 43 NA 250 9.2 92 6 12  
## 14 14 274 10.9 68 5 14  
## 51 13 137 10.3 76 6 20