patient-inflammation.R

Jeffrey Horwitz

2022-12-01

# step 5-6

url <- "https://github.com/Jefe238/r-novice-inflammation.git"  
download.file(url, "r-novice-inflammation-data.zip")  
unzip("r-novice-inflammation-data.zip")

## Warning in unzip("r-novice-inflammation-data.zip"): error 1 in extracting from  
## zip file

# step 7

install.packages("readr")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.2'  
## (as 'lib' is unspecified)

library(readr)  
data01 <- read.csv("/cloud/project/data/data/inflammation-01.csv", header = FALSE, sep = ',')

# step 8A

class(data01)

## [1] "data.frame"

# the outpout of this command is a data frame  
typeof(data01)

## [1] "list"

# the type is a list

# setp 8B

dim(data01)

## [1] 60 40

# It means that the dataset has 60 observations and 40 variables

# step 8C

data01[30, 20]

## [1] 16

data01[c(1, 3, 5), c(10, 20)]

## V10 V20  
## 1 3 18  
## 3 9 10  
## 5 4 17

#How do we select the first ten columns of values for the first four rows?   
data01[c(1:4), c(1:10)]

## V1 V2 V3 V4 V5 V6 V7 V8 V9 V10  
## 1 0 0 1 3 1 2 4 7 8 3  
## 2 0 1 2 1 2 1 3 2 2 6  
## 3 0 1 1 3 3 2 6 2 5 9  
## 4 0 0 2 0 4 2 2 1 6 7

#How do we select the first ten columns of rows 5 to 10?  
data01[c(5:10), c(1:10)]

## V1 V2 V3 V4 V5 V6 V7 V8 V9 V10  
## 5 0 1 1 3 3 1 3 5 2 4  
## 6 0 0 1 2 2 4 2 1 6 4  
## 7 0 0 2 2 4 2 2 5 5 8  
## 8 0 0 1 2 3 1 2 3 5 3  
## 9 0 0 0 3 1 5 6 5 5 8  
## 10 0 1 1 2 1 3 5 3 5 8

# step 8D

data01[, 'V16']

## [1] 4 4 15 8 10 15 13 9 11 6 3 8 12 3 5 10 11 4 11 13 15 5 14 13 4  
## [26] 9 13 6 7 6 14 3 15 4 15 11 7 10 15 6 5 6 15 11 15 6 11 15 14 4  
## [51] 10 15 11 6 13 8 4 13 12 9

data01$V16

## [1] 4 4 15 8 10 15 13 9 11 6 3 8 12 3 5 10 11 4 11 13 15 5 14 13 4  
## [26] 9 13 6 7 6 14 3 15 4 15 11 7 10 15 6 5 6 15 11 15 6 11 15 14 4  
## [51] 10 15 11 6 13 8 4 13 12 9

# first row, all of the columns  
patient\_1 <- data01[1, ]  
  
# max inflammation for patient 1  
max(patient\_1)

## [1] 18

# sum inflammation for patient 1  
sum(patient\_1)

## [1] 218

# step 8E

# max inflammation for patient 1   
max(data01[1, ])

## [1] 18

# the minimum inflammation for patient 2 on day 7  
min(data01[2,"V7"]) #or min(data01[2,7])

## [1] 3

# The minimum inflammation can be found using the code min(). The result is 3.

# step 8F

max(data01[1, ])

## [1] 18

mean(data01 [1, ])

## Warning in mean.default(data01[1, ]): argument is not numeric or logical:  
## returning NA

## [1] NA

mean(as.numeric(data01[1, ]))

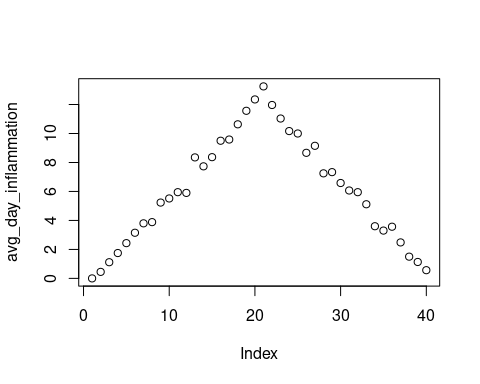
## [1] 5.45

# step 8G

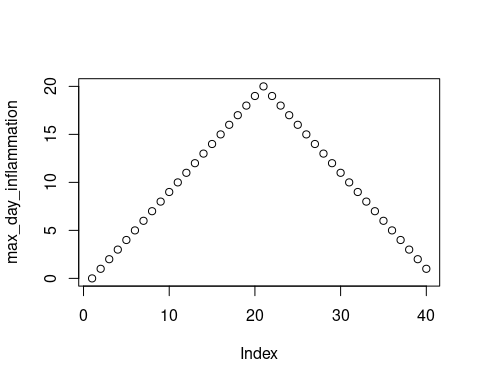
avg\_patient\_inflammation <- apply(data01, 1, mean)  
avg\_day\_inflammation <- apply(data01, 2, mean)

# step 9

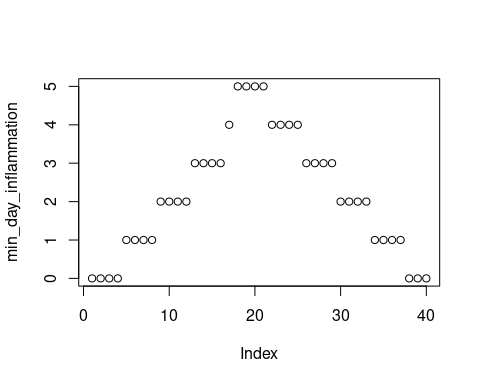
plot(avg\_day\_inflammation)



# the maximum inflammation per day  
max\_day\_inflammation <- apply(data01, 2, max)  
plot(max\_day\_inflammation)



# The data looks this way because it's only taking one variable (max day inflammation) vs the average.   
  
# the minimum inflammation per day  
min\_day\_inflammation <- apply(data01, 2, min)  
plot(min\_day\_inflammation)



# I believe that this data is correct, or else R would have given an error message. The reason this is correct is because our code is right and R is interpreting exactly what we expect it to - min inflammation data across 40 days.