ma-xinshi-homework1

Jan 5. 2022

1

A

```
# import data
rain.df <- read.table('rnf6080.dat')
```

I used the command `read.table` to import the dat file into R.

B

```
# number of rows
nrow(rain.df)
```

```
## [1] 5070
```

```
# number of columns
ncol(rain.df)
```

```
## [1] 27
```

I used the `nrow()` function to get the number of rows and the `ncol()` function to get the number of columns.

C

```
#get names
names(rain.df)
```

```
##  [1] "V1"  "V2"  "V3"  "V4"  "V5"  "V6"  "V7"  "V8"  "V9"  "V10" "V11" "V12"
## [13] "V13" "V14" "V15" "V16" "V17" "V18" "V19" "V20" "V21" "V22" "V23" "V24"
## [25] "V25" "V26" "V27"
```

I used the function `names()` to get the column names which range from "V1" to "V27".

D

```
# select
rain.df[2,4]
```

```
## [1] 0
```

I used brackets [row, col] to get the value, it is 0.

E

```
rain.df[2,]
```

```
##   V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V21
## 2 60  4  2  0  0  0  0  0  0   0   0   0   0   0   0   0   0   0   0   0   0
##   V22 V23 V24 V25 V26 V27
## 2   0   0   0   0   0   0
```

I used brackets [row,] to get the entire row, the contents are numbers 60 (V1), 4 (V2), 2 (V3) and then all 0s for V4 to V27.

F

```
#name cols
names(rain.df) <- c("year","month","day",seq(0,23))
```

This command creates a vector of names including year, month, day and a sequence of number ranging from 0 to 24 and then assigns these names to the respective columns. Column 1 is now named "year", column 2 "month", column 3 "day" and so on.

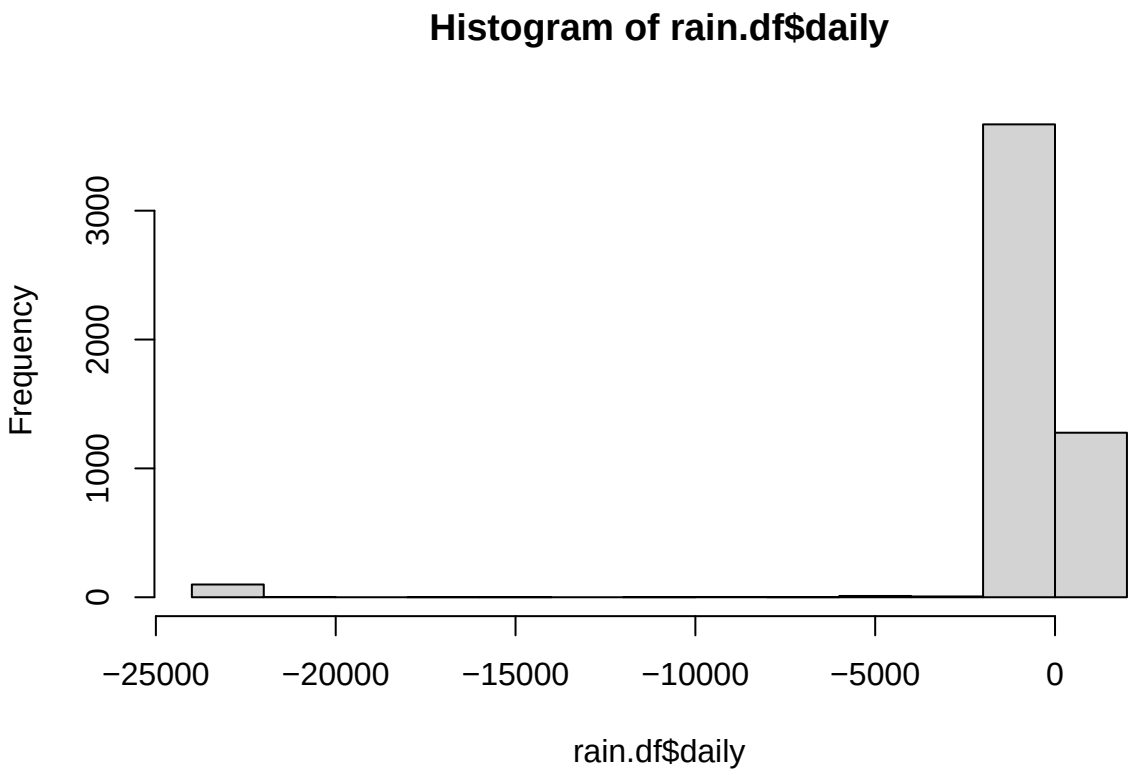
G

```
rain.df$daily = rowSums(rain.df[,4:27])
```

I used the rowSums function to add the values of each row together (leaving out year month and day) and assigned them to a new column in the dataframe called "daily".

H

```
#create histogram
hist(rain.df$daily)
```



I

There are very negative values in the amount of rain daily which is impossible (to have negative rainfall a day).

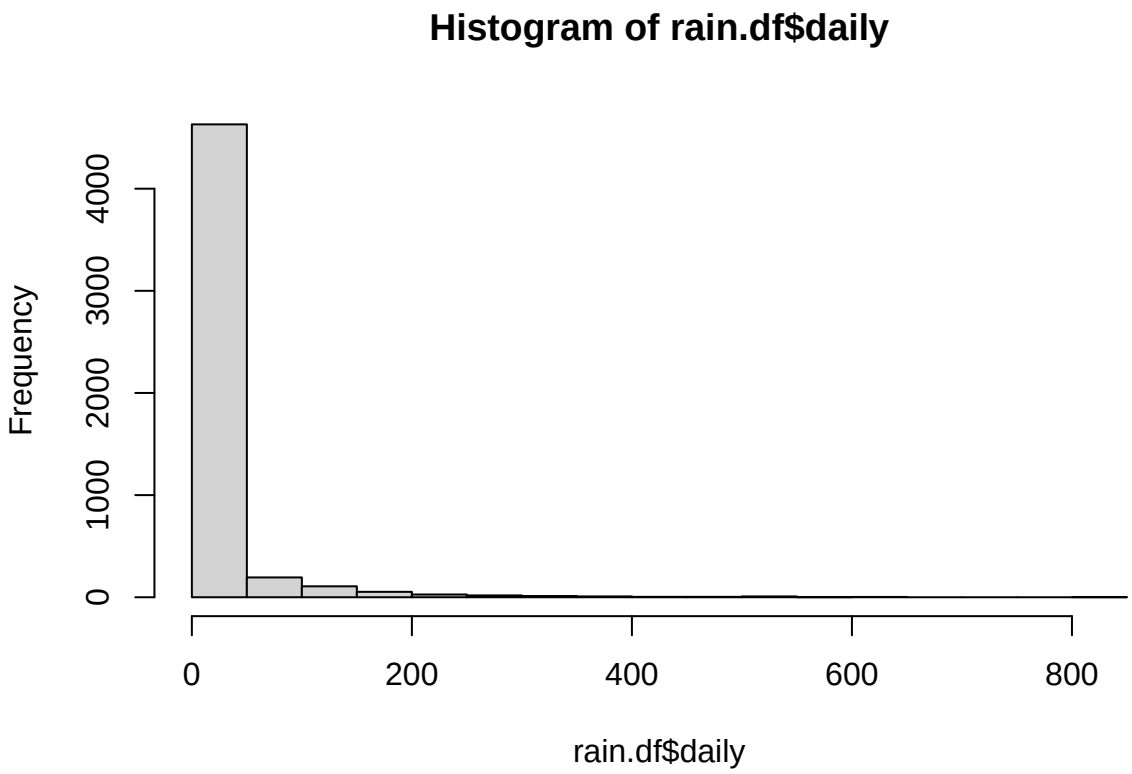
J

```
# replace negative values with 0
rain.df[rain.df < 0] <- 0
# redo sum of daily rainfall
rain.df$daily = rowSums(rain.df[,4:27])
```

We can remove the negative values of rainfall and replace them with 0 (since they are data points with error).

K

```
#draw histogram
hist(rain.df$daily)
```



There are no longer negative values of daily rainfall and the distribution of rainfall is more clearly displayed starting from 0. It makes sense that on most days rainfall is concentrated around 0: no or little rain.

2

A

```
x <- c("5","12","7")
max(x)
```

```
## [1] "7"
```

```
sort(x)
```

```
## [1] "12" "5"  "7"
```

```
#sum(x)
```

x is a list/vector of strings therefore max() takes the largest "number" by lexicographical order, comparing the first character in the string first and then the second, etc. Therefore comparing 1, 5, 7, max() returns 7. Likewise, sort() sorts the strings in order starting with the first character in ascending order so 1 is first and then 5 and then 7. sum() does not work with the string data type and will return an error.

B

```
y <- c("5",7,12)
# y[2] + y[3]
```

y is a list of "5", 7, 12. The list forces all entries to be of the same data type. Since "5" is a string, y2 and y3 are forced into "7" and "12", therefore addition y2 + y3 does not work (on strings) and returns an error.

C

```
z <- data.frame(z1="5",z2=7,z3=12)
z[1,2] + z[1,3]
```

```
## [1] 19
```

z <- data.frame(z1="5",z2=7,z3=12) creates a dataframe named z that contains 1 row and 3 columns. row 1 column 1 is "5" data type of string while row 1 col 2 and row 1 col 3 is 7 and 12 respectively (numeric data type). z[1,2] + z[1,3] adds row 1 col 2 and row 1 col 3: 7+12 = 19.

3

A

It is important that code can be reproduced and checked/used by other individuals (or yourself) trying to accomplish the same task or build on top of what is already done. Reproducible code is also very useful for collaboration on a particular project.

B

Making my code reproducible can help me go back and look over what I have done in the past and build on it for other assignments or new concepts, instead of starting from scratch every time. It also makes it easier to ask for help, if I send the code to others, they should obtain the same results.

C

4