

## ЛР-2 Абдул Карим Э.Н. РИ-411055

2024-10-18

Загрузка данных в датафрейм

```
url <-  
"http://people.math.umass.edu/~anna/Stat597AFall2016/rnf6080.dat"
```

Вывелось 5070 наблюдений 27 переменных

```
dim(data.df)  
[1] 5070 27
```

Имена колонок из data.df

```
colnames(data.df)  
[1] "V1" "V2" "V3" "V4" "V5" "V6" "V7" "V8" "V9" "V10" "V11"  
"V12" "V13" "V14" "V15"  
[16] "V16" "V17" "V18" "V19" "V20" "V21" "V22" "V23" "V24" "V25" "V26"  
"V27"
```

Значение из 5 строки седьмого столбца

```
data.df[5, 7]  
[1] 0
```

2 строка из data.df

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

Строка кода names(data.df) <- c("year", "month", "day", seq(0,23))

```
names(data.df) <- c("year", "month", "day", seq(0, 23))
```

```
> head(data.df)
```

```
year month day 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20  
21 22 23  
1 60 4 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0  
2 60 4 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0  
3 60 4 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0  
4 60 4 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```

0 0 0
5 60 4 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0
6 60 4 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0
tail(data.df)
      year month day 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
20 21 22 23
5065 80 11 25 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0
5066 80 11 26 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0
5067 80 11 27 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0
5068 80 11 28 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0
5069 80 11 29 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0
5070 80 11 30 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0

```

Три колонки “year” (год), “month” (месяц) и “day” (день) указывают на дату. Остальные 24 колонки, от 0 до 23, содержат информацию о количестве осадков, зафиксированных каждый час в течение суток. `head(data.df)` отображает первые строки датафрейма, что поможет разобраться в структуре данных и значениях в первых колонках (год, месяц, день). Последние 24 колонки, упоминаемые в выводе `tail(data.df)`, содержат информацию об осадках за каждый час в течение суток.

```

data.df$daily <- rowSums(data.df[, 4:27])
> hist(data.df$daily, main = "Гистограмма дневных осадков", xlab =
"Осадки (в мм)", ylab = "Частота")
> fixed.df <- data.df
> fixed.df$daily[fixed.df$daily < 0] <- 0

```

Добавьте новую колонку с названием `daily`, в которую запишите сумму крайних правых 24 колонок. Создайте новый датафрейм `fixed.df`. Диаграммы в конце файла

## Синтаксис и типизирование

`

```

v <- c("4", "8", "15", "16", "23", "42")
> max(v) # вернет "42", т.к. сравнение строк
[1] "8"
> > sort(v) # сортирует как строки
[1] "15" "16" "23" "4"  "42" "8"
> sum(v) #выдаст ошибку, потому что текст нельзя складывать как числа
Ошибка в sum(v) : неправильный 'type' (character) аргумента

```

```

v2 <- c("5",7,12)
> df3 <- data.frame(z1="5",z2=7,z3=12)
> df3[1,2] + df3[1,3] # сложение: 7 + 12 = 19
[1] 19
> l4 <- list(z1="6", z2=42, z3="49", z4=126)
> l4[[2]] + l4[[4]] # вернет 168, сумма чисел
[1] 168

```

Числа от 1 до 10000 с инкрементом 372.

```

seq(1, 10000, by = 372)
[1] 1 373 745 1117 1489 1861 2233 2605 2977 3349 3721 4093 4465
4837 5209 5581 5953 6325 6697 7069 7441 7813 8185 8557 8929 9301 9673

```

Числа от 1 до 10000 длиной 50.

```

seq(1, 10000, by = 372)
[1] 1 373 745 1117 1489 1861 2233 2605 2977 3349 3721 4093 4465
4837 5209 5581 5953 6325 6697 7069 7441 7813 8185 8557 8929 9301 9673

```

```

rep(1:5, times=3) # повторяет всю последовательность (1, 2, 3, 4, 5)
три раза
[1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5

```

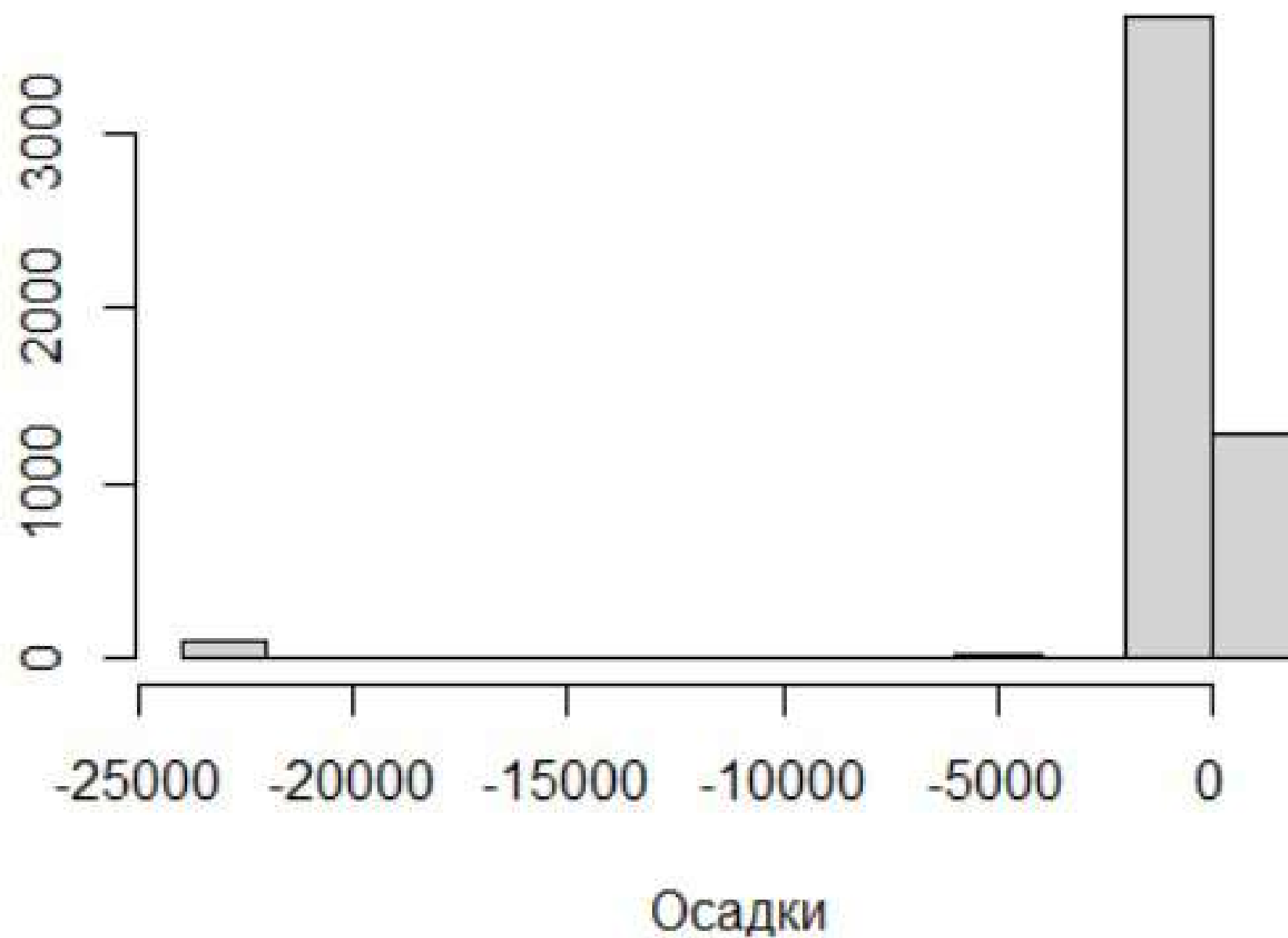
```

rep(1:5, each=3) # повторяет каждое число три раза
[1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5

```

...

## Гистограмма ежедневных осадков



## Исправленная гистограмма

