Язык R
и
его применение в биоинформатике

Анастасия Жарикова
Дмитрий Пензар

Занятие 2

9 сентября 2019

Встроенные наборы данных

```
data()
```

mtcars

```
head(mtcars,4)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
```

```
?mtcars
```

mtcars

```
str(mtcars)
```

```
## 'data.frame':     32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

mtcars

```
rownames(mtcars)
```

```
##  [1] "Mazda RX4"           "Mazda RX4 Wag"       "Datsun 710"
##  [4] "Hornet 4 Drive"      "Hornet Sportabout"   "Valiant"
##  [7] "Duster 360"          "Merc 240D"           "Merc 230"
## [10] "Merc 280"            "Merc 280C"           "Merc 450SE"
## [13] "Merc 450SL"          "Merc 450SLC"         "Cadillac Fleetwood"
## [16] "Lincoln Continental" "Chrysler Imperial"   "Fiat 128"
## [19] "Honda Civic"         "Toyota Corolla"      "Toyota Corona"
## [22] "Dodge Challenger"    "AMC Javelin"         "Camaro Z28"
## [25] "Pontiac Firebird"    "Fiat X1-9"           "Porsche 914-2"
## [28] "Lotus Europa"        "Ford Pantera L"      "Ferrari Dino"
## [31] "Maserati Bora"       "Volvo 142E"
```

mtcars

```
colnames(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

mtcars

```
table(mtcars$cyl)
```

```
##
##  4  6  8
## 11  7 14
```

Выбор строк, столбцов, ячеек

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
tail(mtcars,7)
```

```
##                 mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Fiat X1-9      27.3   4  79.0  66 4.08 1.935 18.9  1  1    4    1
## Porsche 914-2  26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

Выбор строк, столбцов, ячеек

```
mtcars[12,2]
```

```
## [1] 8
```

```
mtcars[8,]
```

```
##               mpg cyl  disp hp drat   wt qsec vs am gear carb
## Merc 240D 24.4    4 146.7 62 3.69 3.19   20  1  0    4    2
```

```
mtcars[1:3,]
```

```
##               mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

Выбор строк, столбцов, ячеек

```
mtcars[,2]
```

```
##  [1] 6 6 4 6 8 6 8 4 4 6 6 8 8 8 8 8 8 4 4 4 4 8 8 8 8 4 4 4 8 6 8 4
```

```
mtcars[c(1,13),]
```

```
##             mpg cyl  disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4  21.0   6 160.0 110 3.90 2.62 16.46  0  1    4    4
## Merc 450SL 17.3   8 275.8 180 3.07 3.73 17.60  0  0    3    3
```

```
mtcars[c(1,3,7),2]
```

```
## [1] 6 4 8
```

Выбор строк, столбцов, ячеек

```
mtcars['Mazda RX4',c('mpg','cyl','disp','am')]
```

```
##           mpg cyl disp am
## Mazda RX4  21   6  160  1
```

Добавить столбец к data frame

```
dim(mtcars)
```

```
## [1] 32 11
```

```
num <- 1:32
mtnew <- cbind(mtcars,num)
dim(mtnew)
```

```
## [1] 32 12
```

```
mtnew[30:32,]
```

```
##                mpg cyl disp  hp drat   wt qsec vs am gear carb num
## Ferrari Dino  19.7   6  145 175 3.62 2.77 15.5  0  1    5    6  30
## Maserati Bora 15.0   8  301 335 3.54 3.57 14.6  0  1    5    8  31
## Volvo 142E    21.4   4  121 109 4.11 2.78 18.6  1  1    4    2  32
```

Задание

Изменить имя столбца num

Выбор данных по условию

```
mtcars[mtcars$cyl>4 & mtcars$cyl<8,]
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
```

Выбор данных по условию - %in%

```
table(mtcars$cyl)
```

```
##
##  4  6  8
## 11  7 14
```

```
vec <- c(4,8)
vec
```

```
## [1] 4 8
```

```
mt <- mtcars[mtcars$cyl %in% vec,]
table(mt$cyl)
```

```
##
##  4  8
## 11 14
```

Выбор данных по условию - !(%in%)

```
table(mtcars$cyl)
```

```
##
##  4  6  8
## 11  7 14
```

```
vec <- c(4,8)
vec
```

```
## [1] 4 8
```

```
mt <- mtcars[!(mtcars$cyl %in% vec),]
table(mt$cyl)
```

```
##
## 6
## 7
```

which

```
which(mtcars$mpg == 21)
```

```
## [1] 1 2
```

```
which(mtcars$wt == 3.215)
```

```
## [1] 4
```

```
which(rownames(mtcars) == 'Valiant')
```

```
## [1] 6
```

which

```
which.min(mtcars$qsec)
```

```
## [1] 29
```

```
which.max(mtcars$qsec)
```

```
## [1] 9
```

which

```
qsec_max <- which.max(mtcars$qsec)
qsec_max
```

```
## [1] 9
```

```
mtcars[which.max(mtcars$qsec),]
```

```
##               mpg cyl  disp hp drat   wt qsec vs am gear carb
## Merc 230 22.8    4 140.8 95 3.92 3.15 22.9  1  0    4    2
```

```
mtcars[qsec_max,]
```

```
##               mpg cyl  disp hp drat   wt qsec vs am gear carb
## Merc 230 22.8    4 140.8 95 3.92 3.15 22.9  1  0    4    2
```

Добавить строку к data frame

```
dim(mtnew)
```

```
## [1] 32 12
```

```
mtnew[1,]
```

```
##               mpg cyl disp  hp drat   wt  qsec vs am gear carb num
## Mazda RX4      21   6  160 110  3.9 2.62 16.46  0  1    4    4   1
```

```
newcar <- data.frame(mpg=21, cyl=4, disp=100, hp=80,
                     drat=1, wt=2,qsec=16, vs=1,am=0, gear=4, carb=1, num=33)
mtnew<-rbind(mtnew, newcar)
rownames(mtnew)[33]<-"Lada"
mtnew[30:33,]
```

```
##                mpg cyl disp  hp drat   wt qsec vs am gear carb num
## Ferrari Dino  19.7   6  145 175 3.62 2.77 15.5  0  1    5    6  30
## Maserati Bora 15.0   8  301 335 3.54 3.57 14.6  0  1    5    8  31
## Volvo 142E    21.4   4  121 109 4.11 2.78 18.6  1  1    4    2  32
## Lada          21.0   4  100  80 1.00 2.00 16.0  1  0    4    1  33
```

```
dim(mtnew)
```

```
## [1] 33 12
```

Сортировка

```
mtcars[order(mtcars$drat),]
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
```

Работа с переменными

```r
ls()
```

```
## [1] "mtnew"   "newcar" "num"
```

```r
rm(list=ls())
ls()
```

```
## character(0)
```

Сохранение данных

```r
write.table(mtcars,file='../mtnew.tab',quote=F, col.names = T,row.names = T,sep='\t')
write.csv(mtcars,file='mtnew.csv')
save(mtcars, file="mtnew.RData")
```

Чтение данных

```
mt<-read.table("../mtnew.tab",sep="\t",header=T)
head(mt)
```

```
##                     mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4          21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag      21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710         22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive     21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout  18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant            18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Чтение данных

```
mt<-read.table("../mtnew.tab",sep="\t",skip = 2)
head(mt)
```

```
##                        V1   V2 V3  V4  V5   V6    V7    V8 V9 V10 V11 V12
## 1      Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0   1   4   4
## 2         Datsun 710 22.8   4 108  93 3.85 2.320 18.61  1   1   4   1
## 3     Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1   0   3   1
## 4  Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0   0   3   2
## 5            Valiant 18.1   6 225 105 2.76 3.460 20.22  1   0   3   1
## 6         Duster 360 14.3   8 360 245 3.21 3.570 15.84  0   0   3   4
```

Чтение данных

```
mt<-read.table("../mtnew.tab",sep="\t",header=T,skip = 2)
head(mt)
```

```
##              Mazda.RX4.Wag  X21 X6   X160 X110 X3.9 X2.875 X17.02 X0 X1 X4 X4.1
## 1             Datsun 710 22.8   4 108.0    93 3.85  2.320  18.61  1  1  4     1
## 2         Hornet 4 Drive 21.4   6 258.0   110 3.08  3.215  19.44  1  0  3     1
## 3      Hornet Sportabout 18.7   8 360.0   175 3.15  3.440  17.02  0  0  3     2
## 4                Valiant 18.1   6 225.0   105 2.76  3.460  20.22  1  0  3     1
## 5             Duster 360 14.3   8 360.0   245 3.21  3.570  15.84  0  0  3     4
## 6              Merc 240D 24.4   4 146.7    62 3.69  3.190  20.00  1  0  4     2
```

```
load('mtnew.RData')
```

Данные о качестве воздуха

```
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
```

```
dim(airquality)
```

```
## [1] 153   6
```

В чем проблема?

```
mean (airquality$Ozone)
```

```
## [1] NA
```

Работа с отсутствующими данными

NA - пропущенное значение: is.na()

NaN - результат недопустимой арифметической операции: is.nan()

NULL - отсутствие субъекта: is.null()

Работа с отсутствующими данными

Не учитывать отсутствующие данные

```
mean(airquality$Ozone, na.rm = T)
```

```
## [1] 42.12931
```

Удалить строки с отсутствующими данными

```
air <- na.omit (airquality)
dim (air)
```

```
## [1] 111    6
```

```
mean(air$Ozone)
```

```
## [1] 42.0991
```

Работа с отсутствующими данными

```
is.na(airquality$Ozone)
```

```
##    [1] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE
##   [12] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [23] FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
##   [34]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE
##   [45]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
##   [56]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE
##   [67] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE
##   [78] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
##   [89] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [100] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
##  [111] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
##  [122] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [144] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
```

Работа с отсутствующими данными

```r
sum(is.na(airquality$Ozone))
```

```
## [1] 37
```

Работа с отсутствующими данными

```
which(is.na(airquality$Ozone))
```

```
##  [1]   5  10  25  26  27  32  33  34  35  36  37  39  42  43  45  46  52
## [18]  53  54  55  56  57  58  59  60  61  65  72  75  83  84 102 103 107
## [35] 115 119 150
```

```
anyNA(airquality$Ozone)
```

```
## [1] TRUE
```

NaN

```
0/0
```

```
## [1] NaN
```

Удаление столбцов

```
head(air,3)
```

```
##   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
```

```
air$Ozone = NULL
head(air,3)
```

```
##   Solar.R Wind Temp Month Day
## 1     190  7.4   67     5   1
## 2     118  8.0   72     5   2
## 3     149 12.6   74     5   3
```