# Язык R
# и его применение в биоинформатике

Анастасия Александровна Жарикова

Дмитрий Дмитриевич Пензар

1 сентября 2020

# Правила игры

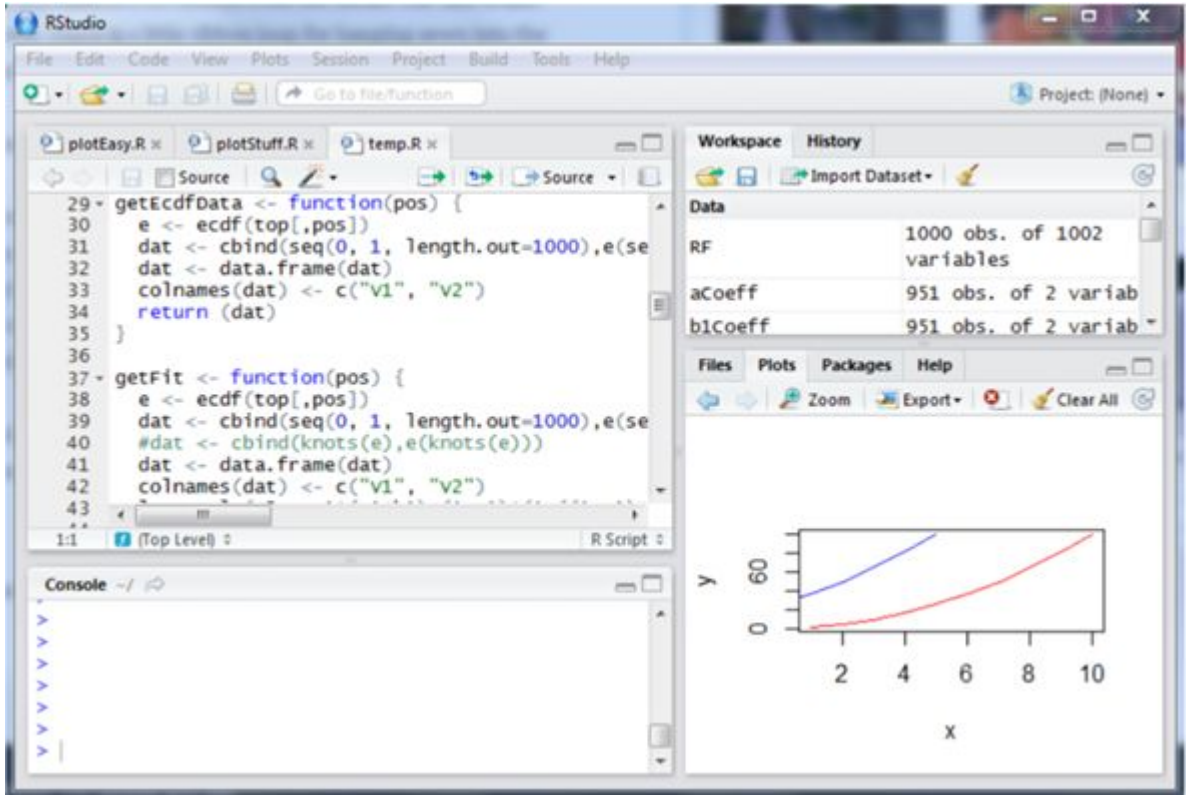
- Занятия очные до иных распоряжений
- Ходить можно только со своей группой
- Презентации, домашние задания и прочие материалы – на странице курса на kodomo
- Набор баллов: (см. документ с правилами!)
  - Самостоятельные работы
  - Домашние задания (есть сроки сдачи)
  - Контрольные работы
  - Проект
- Переписываний контрольных и самостоятельных работ не будет
- Мы считаем, что вы усвоили программу прошлого семестра по статистике
- Материалы лекций прорабатываются вами дома
- Можно и нужно (!!!) задавать вопросы

# ~~За что~~ Почему R?

- Бесплатный
- Очень простой и быстрый в изучении
- Форумы, поддержка, развитие
- Быстрая и удобная статистическая обработка данных
- Построение красивых графиков
- Множество дополнительных пакетов
- Обработка данных RNA-seq
- Создание отчетов
- Некоторые лаборатории используют только R

# R и RStudio

Язык R - https://cloud.r-project.org/
RStudio - https://rstudio.com/



# Демонстрация работы RStudio

RStudio есть на kodomo

https://kodomo.fbb.msu.ru/rstudio

# R - рекомендации к работе

- Можно запускать без RStudio
- Можно запускать из командной строки
- RStudio - настоятельно рекомендуется (особенно на контрольных работах!)
- Создавайте проекты в RStudio
- Сохраняйте свой код в виде скриптов script.R (на контрольных работах - это обязательное требование) или script.Rmd
- Сохраняйте полученные таблицы и графики в отдельные файлы
- Использовать Google можно и нужно!
- Если в задании требуется прислать файл, то называть его следует по принципу: Petrov_R_quiz2.R

# Markdown

https://rmarkdown.rstudio.com/index.html

https://rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf

tinytex::install_tinytex() - для сохранения отчета в pdf

Запускается целиком, сразу видны ошибки и проблемы

Фрагменты кода записывают в “chunk”:

```{r}
your_code
```

Переменные и функции, инициализированные в одном chunk, будут работать во всем документе

# Markdown

https://rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf

**syntax**

```
Plain text
End a line with two spaces to start a new paragraph.
*italics* and _italics_
**bold** and __bold__
superscript^2^
~~strikethrough~~
[link](www.rstudio.com)

# Header 1

## Header 2

### Header 3

#### Header 4

##### Header 5

###### Header 6

endash: --
emdash: ---
ellipsis: ...
inline equation: $A = \pi*r^{2}$
image: ![](path/to/smallorb.png)

horizontal rule (or slide break):

***
```

**becomes**

Plain text
End a line with two spaces to start a new paragraph.
*italics* and *italics*
**bold** and **bold**
superscript$^2$
~~strikethrough~~
link

# Header 1

## Header 2

### Header 3

#### Header 4

##### Header 5

###### Header 6

endash: –
emdash: —
ellipsis: …
inline equation: $A = \pi * r^2$
image: 
horizontal rule (or slide break):

# Markdown

```
### Отображать и код и результат его выполнения

```{r}
dim (mtcars)
```

### Отображать только результат выполнения кода

```{r, echo = F}
dim (mtcars)
```

### Код отображается, но не выполняется

```{r, eval = F}
dim (mtcars)
```

### Выполнение кода внутри строки текста

Two plus two equals `r 2+2`
```

### Отображать и код и результат его выполнения

```
dim (mtcars)
```

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

### Отображать только результат выполнения кода

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

### Код отображается, но не выполняется

```
dim (mtcars)
```

### Выполнение кода внутри строки текста

Two plus two equals 4

# Markdown

В случае ошибки в коде итоговый документ не будет создан, появится сообщение об ошибке


Решение:

```
```{r, error = T}
mean(a)
```
```

```
mean(a)
```

```
## Error in mean(a): object 'a' not found
```

# Markdown

https://rpruim.github.io/s341/S19/from-class/MathinRmd.html

$$\int x^2 dx = \frac{x^3}{3} + C.$$

$$\log(x)$$

$$a = b$$

$$X \sim \mathsf{Norm}(10, 3)$$

$$5 \leq 10$$

$$X \sim \mathsf{Binom}(n, \pi)$$

$$\sum_{n=1}^{10} n^2$$

# Markdown

```
---
title: "Домашнее задание - тест - 1.09.2020"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

### 1. Сколько строк и столбцов в наборе данных mtcars?

```{r}
dim(mtcars)
```

**Ответ:** 32 строки и 11 столбцов

### 2. Постройте график зависимости mpg от wt по данным mtcars?

```{r}
library(ggplot2)
ggplot(mtcars, aes(x=wt, y=mpg)) + geom_point()
```
```

## Домашнее задание - тест - 1.09.2020

1. Сколько строк и столбцов в наборе данных mtcars?

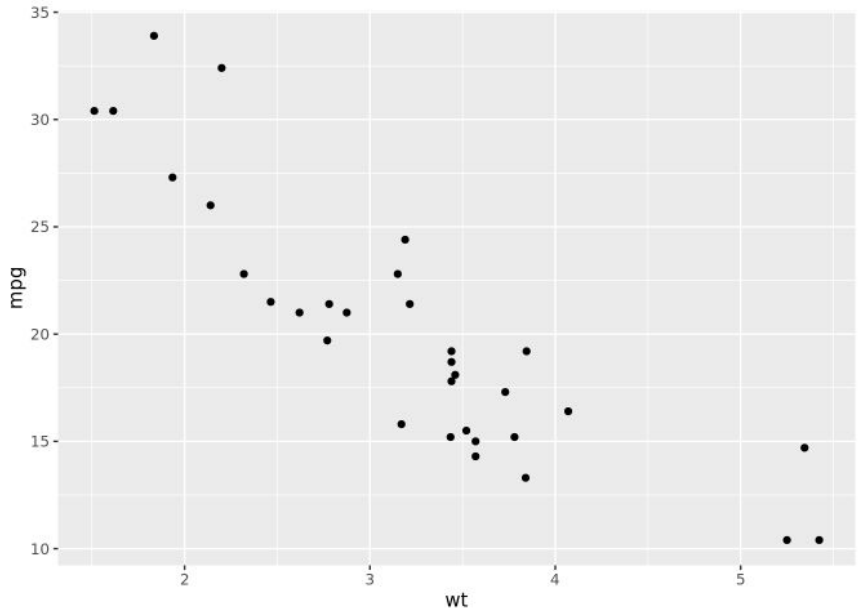
```
dim(mtcars)
```

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

**Ответ:** 32 строки и 11 столбцов

2. Постройте график зависимости mpg от wt по данным mtcars?

```
library(ggplot2)
ggplot(mtcars, aes(x=wt, y=mpg)) + geom_point()
```



# Калькулятор

$$\log_5 60 - \log_5 12$$

```
log(60, base = 5) - log (12, base = 5)
```

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

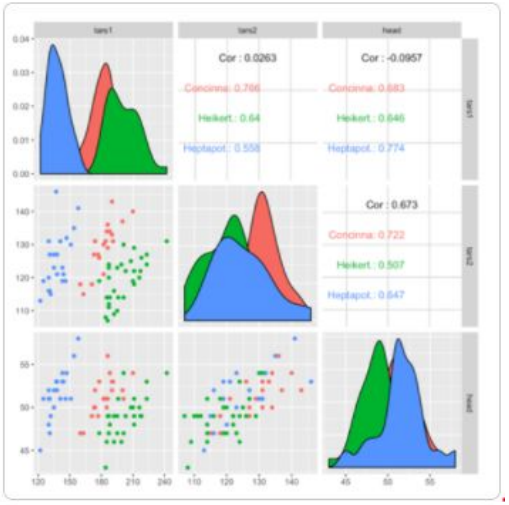
$$\frac{(\sqrt{17} - 2)(\sqrt{(34)} + \sqrt{8} + \sqrt{17} + 2)}{\sqrt{2} + 1}$$

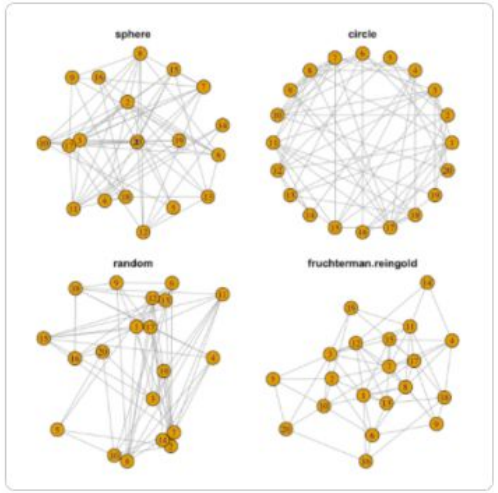
```
(sqrt(17) - 2) * (sqrt(34) + sqrt(8) + sqrt(17) + 2) / (sqrt(2) + 1)
```

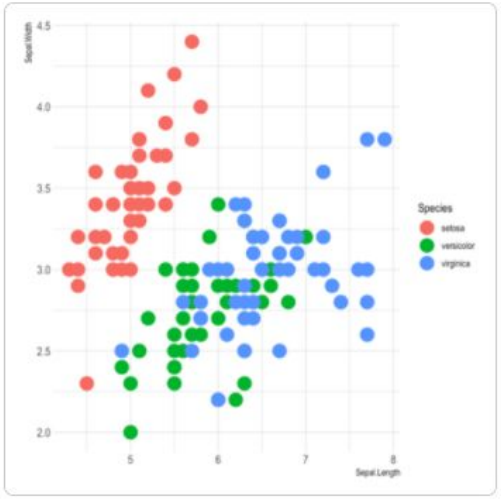
```
## [1] 13
```

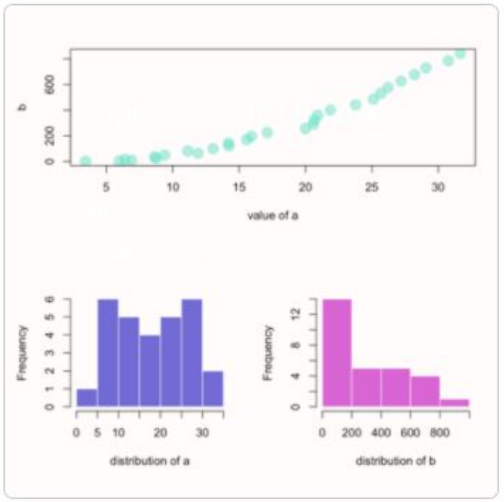
# Графики

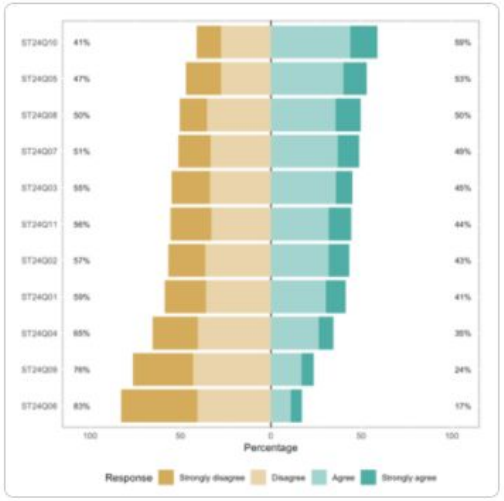
https://www.r-graph-gallery.com/

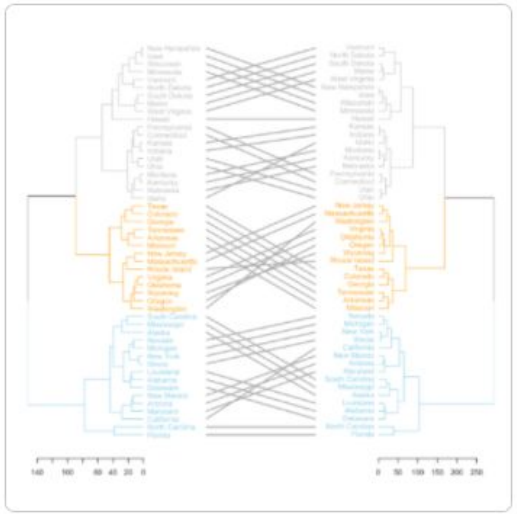


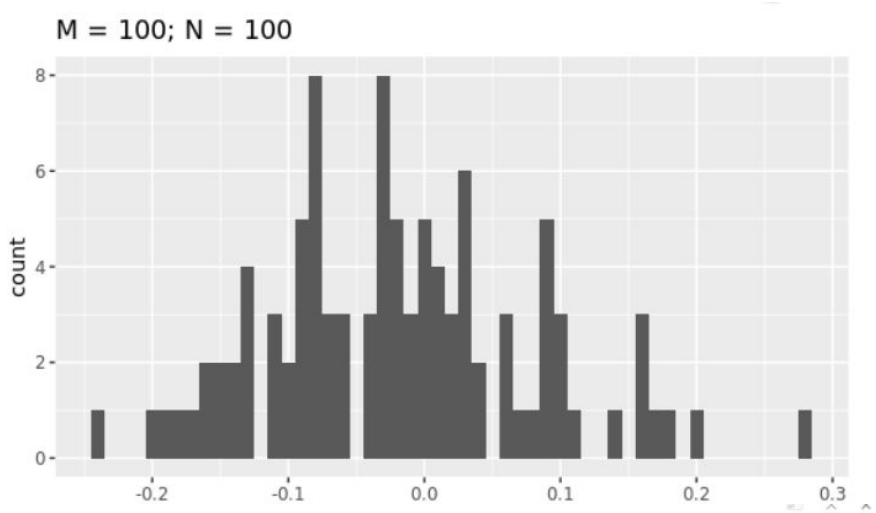



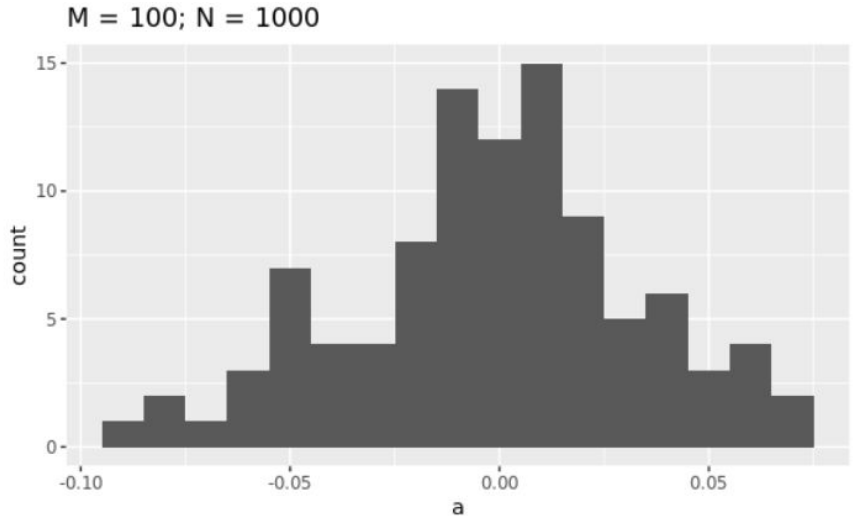



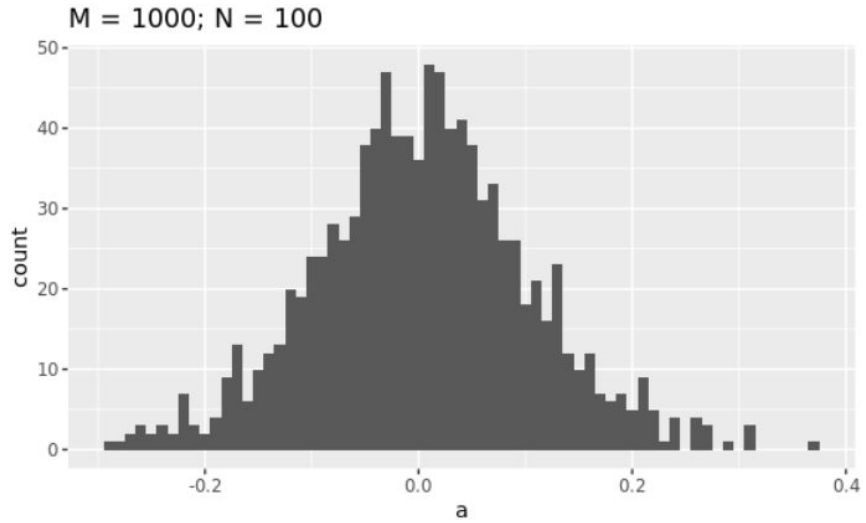



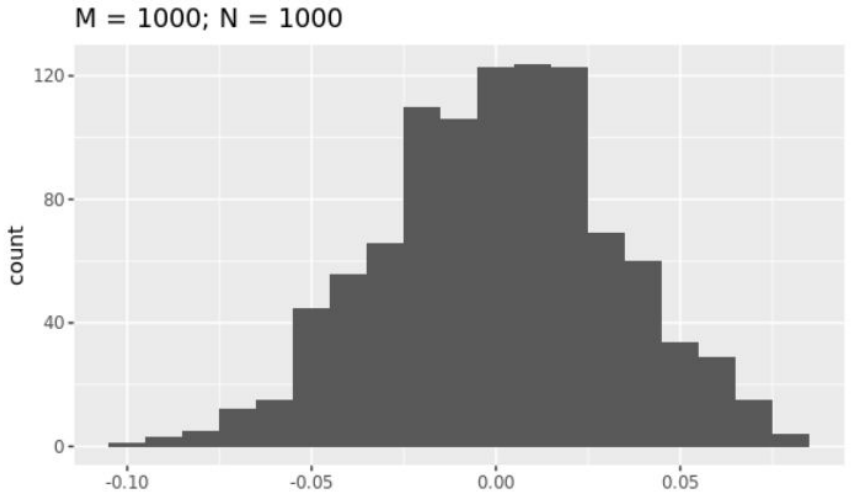
# Симуляции

Возьмем M выборок из нормального распределения размером N. Построим распределение средних этих выборок.



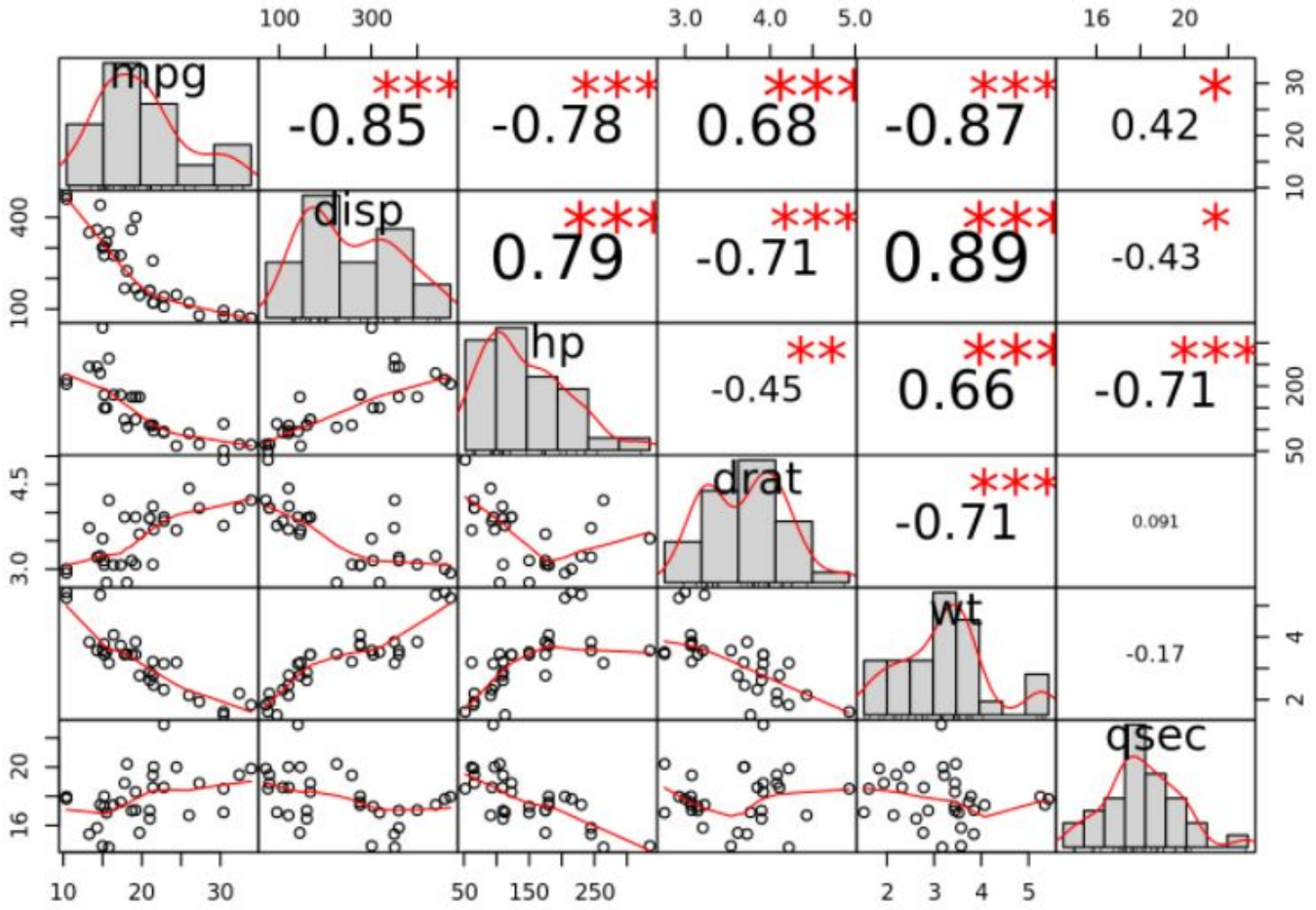



# Статистика

```
library("PerformanceAnalytics")
my_data <- mtcars[, c(1,3,4,5,6,7)]
chart.Correlation(my_data, histogram=TRUE, pch=19)
```



# Статистика

```
library(MASS)
t.test(Prob ~ So, data = UScrime)
```

```
##
##  Welch Two Sample t-test
##
## data:  Prob by So
## t = -3.8954, df = 24.925, p-value = 0.0006506
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03852569 -0.01187439
## sample estimates:
## mean in group 0 mean in group 1
##      0.03851265      0.06371269
```

# Полезные команды

```
getwd()  # узнать рабочую директорию

setwd('Newdir')  # задать рабочую директорию

dir()  # список файлов в рабочей директории
```

# Основной тип данных - вектор

```
x <- 1:5
x
```

```
## [1] 1 2 3 4 5
```

```
y <- 6:10
y
```

```
## [1]  6  7  8  9 10
```

```
length(x)
```

```
## [1] 5
```

# Основной тип данных - вектор

```
x <- 1:5
x
```

```
## [1] 1 2 3 4 5
```

# Основной тип данных - вектор

```
x <- 1:5
x
```

```
## [1] 1 2 3 4 5
```


```
vec <- 100:150
vec
```

```
##  [1] 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116
## [18] 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133
## [35] 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
```

# Операции с векторами

```
x + y
```

```
## [1]  7  9 11 13 15
```

```
x^2
```

```
## [1]  1  4  9 16 25
```

```
x + 3
```

```
## [1] 4 5 6 7 8
```

```
-x
```

```
## [1] -1 -2 -3 -4 -5
```

# Операции с векторами

```
z <- 11:12
```

```
z
```

```
## [1] 11 12
```

```
x
```

```
## [1] 1 2 3 4 5
```

```
x + z
```

# Операции с векторами

```
z
```

```
## [1] 11 12
```

```
x
```

```
## [1] 1 2 3 4 5
```

```
x + z
```

```
## Warning in x + z: longer object length is not a multiple of shorter object
## length
```

```
## [1] 12 14 14 16 16
```

# Операции с векторами

```
a <- 1:6
b <- 10:11
a
```

```
## [1] 1 2 3 4 5 6
```

```
b
```

```
## [1] 10 11
```

# Операции с векторами

```r
a <- 1:6
b <- 10:11
a
```

```
## [1] 1 2 3 4 5 6
```

```r
b
```

```
## [1] 10 11
```

```r
a + b
```

```
## [1] 11 13 13 15 15 17
```

# Операции с векторами

```
x
```

```
## [1] 1 2 3 4 5
```

```
x > 4
```

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

```
x == 4
```

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

# Операции с векторами

```
x
```

```
## [1] 1 2 3 4 5
```

```
x = 4
x
```

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

# Способы создания вектора

Оператор c()

```
c(1,2,3)
```

```
## [1] 1 2 3
```

# Способы создания вектора

## Последовательности

```
1:10
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```
seq(from = 1,to = 8,by = 2)
```

```
## [1] 1 3 5 7
```

```
seq(3,4,length.out = 5)
```

```
## [1] 3.00 3.25 3.50 3.75 4.00
```

# Вызов справки

```
?seq
```

```
help (mtcars)
```

# Способы создания вектора

## Объединение

```
x <- 1:3
x
```

```
## [1] 1 2 3
```

```
x <- c(x, 5:7)
x
```

```
## [1] 1 2 3 5 6 7
```

# Способы создания вектора

## Повторы

```
rep(1:3, times = 3)
```

```
## [1] 1 2 3 1 2 3 1 2 3
```

```
rep(1:3, each = 3)
```

```
## [1] 1 1 1 2 2 2 3 3 3
```

```
rep(1:3, length.out = 5)
```

```
## [1] 1 2 3 1 2
```

# Способы создания вектора

Взять 10 случайных чисел от 1 до 30

```
sample(1:30, 10, replace = T)
```

```
##  [1] 11 25 13  2  4  5  7  4 15 23
```

# Способы создания вектора

## Из распределения

Нужно сгенерировать заданное количество чисел из известного распределения

- rnorm(n,mean,sd) – нормальное распределение
- runif(n,min,max) – равномерное распределение
- rbinom(n,size,prob) – биномиальное распределение
- rpois(n,lambda) – распределение Пуассона

# Способы создания вектора

## Из распределения

```
rpois(20, 10)
```

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

# Способы создания вектора

## Из распределения

```
set.seed(123)
```

```
rpois(20, 10)
```

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

# Немного описательной статистики

```
set.seed(123)
a <- rpois(20, 10)
a
```

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

```
sum(a)
```

```
## [1] 189
```

```
max(a)
```

```
## [1] 15
```

```
sd(a)
```

```
## [1] 3.410124
```

```
mean(a)
```

```
## [1] 9.45
```

# Вектор - данные одного типа

```
x <- c(T,F,F,T)
typeof(x)
```

```
## [1] "logical"
```

```
x <- 1:5
typeof(x)
```

```
## [1] "integer"
```

```
x <- c(0.5, 1.2, 3.6)
typeof(x)
```

```
## [1] "double"
```

```
x <- c('a', 'b', "c")
typeof(x)
```

```
## [1] "character"
```

# Вектор - данные одного типа

```
x <- c(F,1,2,T)
x
```

```
## [1] 0 1 2 1
```

```
typeof(x)
```

```
## [1] "double"
```

```
x <- c(1, 2.8)
x
```

```
## [1] 1.0 2.8
```

```
typeof(x)
```

```
## [1] "double"
```

```
x <- c(1, 'a', F, 5.5)
x
```

```
## [1] "1"     "a"     "FALSE" "5.5"
```

```
typeof(x)
```

```
## [1] "character"
```

# Срезы

```
x <- c(5, 16, 8, 32, 56, 2)
x
```

```
## [1]  5 16  8 32 56  2
```

```
x[1]
```

```
## [1] 5
```

```
x[2:4]
```

```
## [1] 16  8 32
```

```
x[c(2, 5)]
```

```
## [1] 16 56
```

# Срезы

```
x <- c(5, 16, 8, 32, 56, 2)
x
```

```
## [1]  5 16  8 32 56  2
```

```
x[-1]
```

```
## [1] 16  8 32 56  2
```

```
x[x > 10]
```

```
## [1] 16 32 56
```

```
x[x >= 5 & x < 10]
```

```
## [1] 5 8
```

# Информация

```
x <- 1:5
x
```

```
## [1] 1 2 3 4 5
```

```
typeof(x)
```

```
## [1] "integer"
```

```
is.vector(x)
```

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

```
is.logical(x)
```

```
## [1] FALSE
```

```
is.integer(x)
```

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

```
str(x)
```

```
##  int [1:5] 1 2 3 4 5
```

# Data frame

```
##                   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
## Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360         14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230           22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## 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
## 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
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
```

# Data frame. Создание

```r
a <- c(5, 4, 8)
b <- c("aa", "bb", "cc")
h <- c(T, F, T)
df <- data.frame(a,b,h)
df
```

```
##   a  b     h
## 1 5 aa  TRUE
## 2 4 bb FALSE
## 3 8 cc  TRUE
```

# Data frame. Создание

```
df <- data.frame(a = c(5, 4),
                 b = c("aa", "bb"),
                 h = c(T, F))
df
```

```
##   a  b     h
## 1 5 aa  TRUE
## 2 4 bb FALSE
```

# Data frame. Основные операции

```
##   a  b     h
## 1 5 aa  TRUE
## 2 4 bb FALSE
```

```
str(df)
```

```
## 'data.frame':    2 obs. of  3 variables:
##  $ a: num  5 4
##  $ b: Factor w/ 2 levels "aa","bb": 1 2
##  $ h: logi  TRUE FALSE
```

```
dim(df)
```

```
## [1] 2 3
```

# Data frame. Основные операции

```
##   a  b     h
## 1 5 aa  TRUE
## 2 4 bb FALSE
```

```
str(df)
```

```
## 'data.frame':    2 obs. of  3 variables:
##  $ a: num  5 4
##  $ b: Factor w/ 2 levels "aa","bb": 1 2
##  $ h: logi  TRUE FALSE
```

строки

```
dim(df)
```

```
## [1] 2 3
```

столбцы

# Data frame. Основные операции

```
colnames(df)
```

```
## [1] "a" "b" "h"
```

```
rownames(df)
```

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