

Лаборатория экологии морского бентоса (гидробиологии)

Кандалакшский государственный заповедник

Санкт-Петербургский государственный университет

Знакомство с R

М.А. Варфоломеева, PhD

В.М. Хайтов, к.б.н.

Почему R это круто?

М.А. Варфоломеева, PhD

В.М. Хайтов, к.б.н.

Новые впечатления и сообщество единомышленников



- R – это инструмент познания мира.
Вы получите много новых впечатлений и
встречите большое и дружественное
сообщество единомышленников.

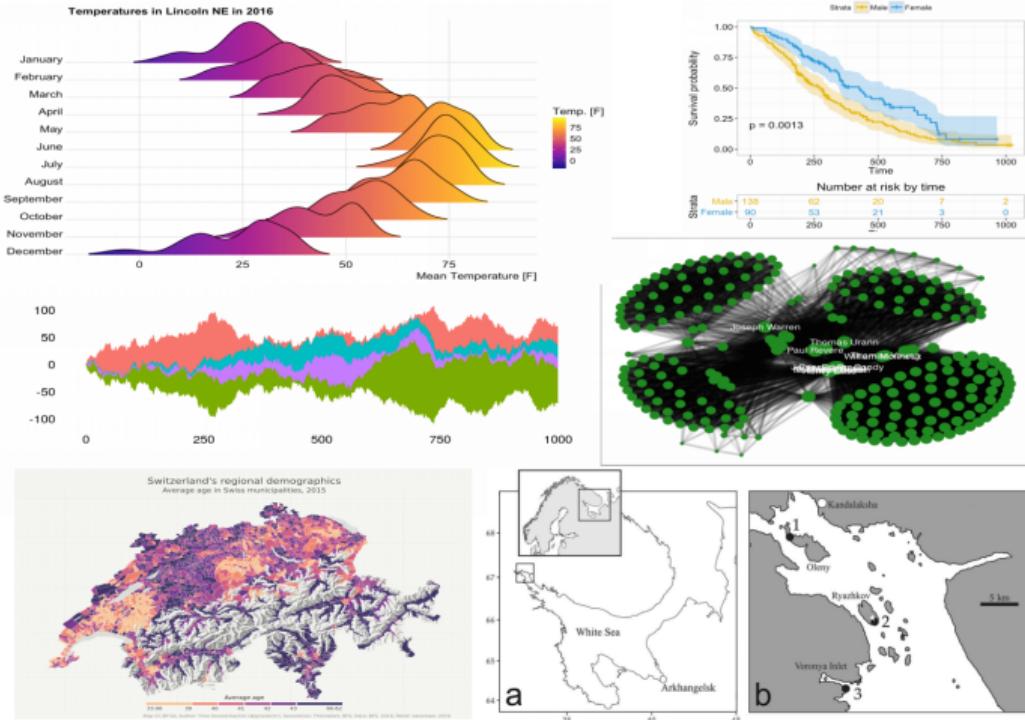
{<https://www.shutterstock.com/ru/image-photo/businessman-touching-global-network-data-exchanges-566877226>}

Огромный набор методов



- R – это открытый кроссплатформенный язык статистического программирования.
- В R реализованы практически все статистические методы, как классические, так и ультрасовременные.
- Доступно более 10 000 пакетов и их число растет. К пакетам есть подробные и понятные справки.

Визуализация данных



В R можно создавать практически любые иллюстрации:

- графики
- карты
- сложные комбинированные рисунки

<https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/>

<http://www.ggplot2-exts.org/gallery/>

Khaitov et al, 2018

Гибкий и воспроизводимый код для анализа данных



- Функции языка R подобны магическим заклинаниям, они совершают чудеса при обработке данных.
- Ход анализа данных может быть задокументирован. Вы напишете собственные магические тексты в виде кода на R, с помощью которого можно воспроизвести все этапы анализа.

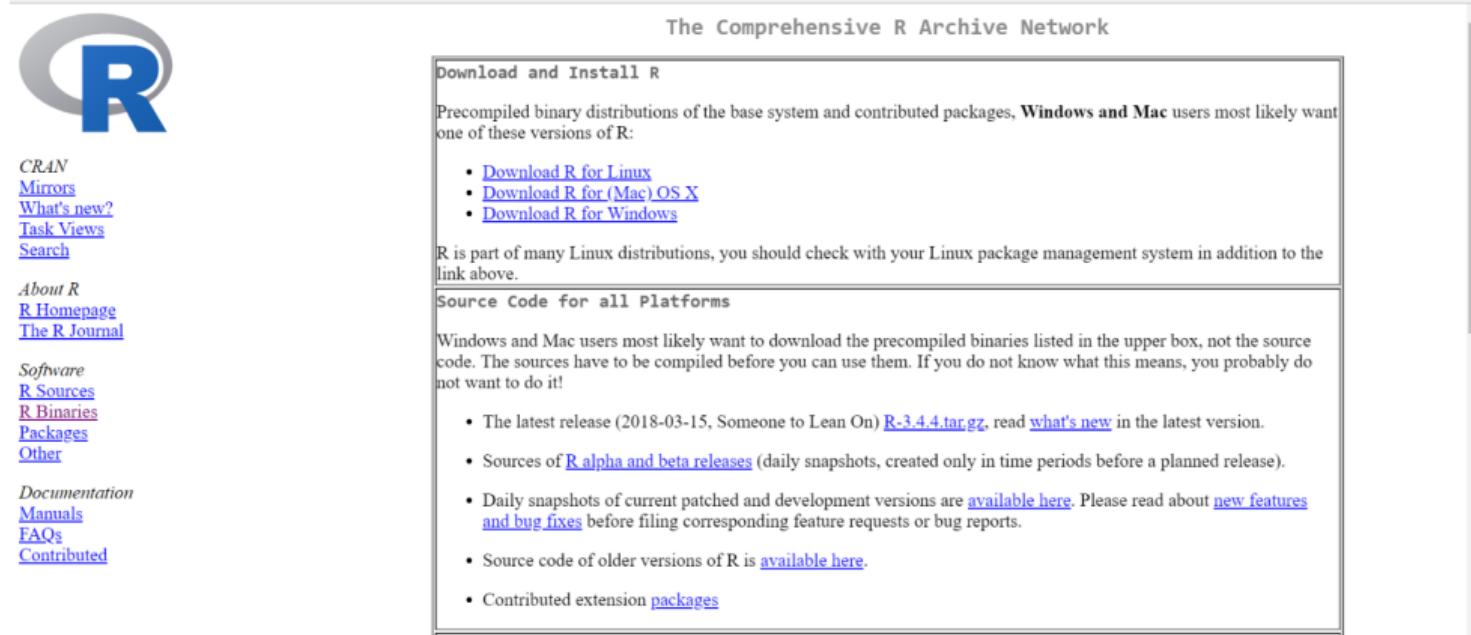
{<https://www.shutterstock.com/ru/image-photo/toy-magician-casting-spell-not-person-425596252?src=9JFUpl7OtfjD9EEo8ZPfQ-1-18>}

Установка и настройка R и RStudio

В.М. Хайтов, к.б.н.

Загрузка R

Сайт CRAN <https://cran.r-project.org/>



The screenshot shows the CRAN homepage with a sidebar on the left containing links for CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, and The R Journal. The main content area is titled "The Comprehensive R Archive Network". It features a large "R" logo and two main sections: "Download and Install R" and "Source Code for all Platforms". The "Download and Install R" section provides links for Linux, Mac OS X, and Windows. The "Source Code for all Platforms" section explains that users should check their Linux package management system and lists release tarballs, alpha/beta releases, daily snapshots, and contributed packages.

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2018-03-15, Someone to Lean On) [R-3.4.4.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Загрузка RStudio

<https://www.rstudio.com/products/rstudio/download/>

The screenshot shows the RStudio download page with the following information:

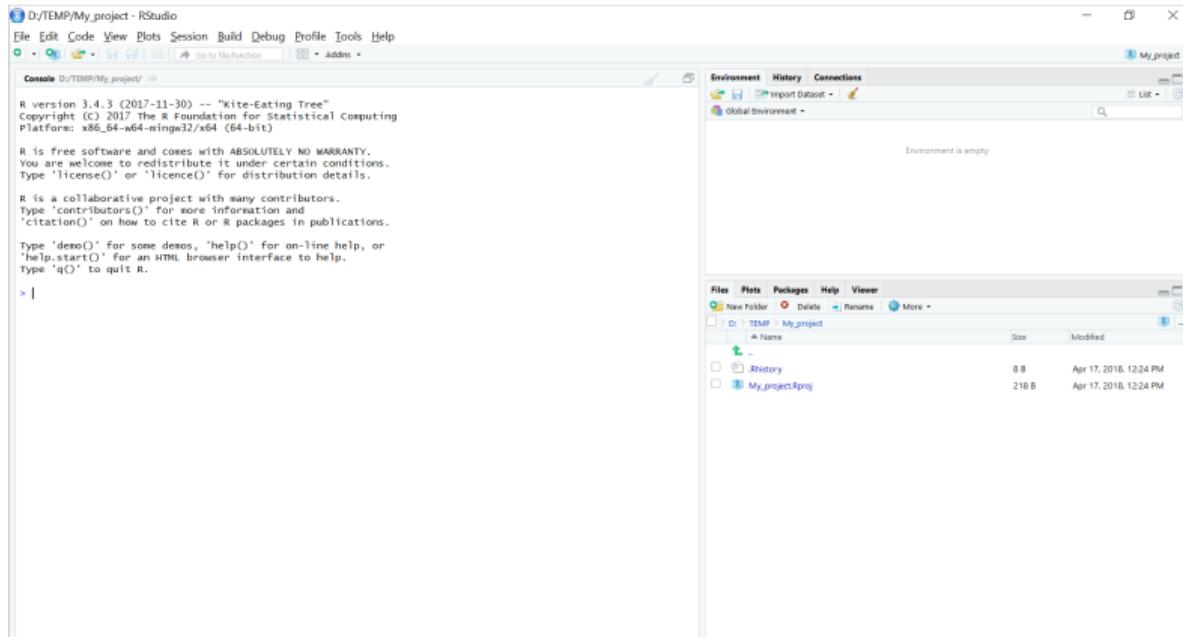
Version	License	Cost
RStudio Desktop	Open Source License	FREE
RStudio Desktop	Commercial License	\$995 per year
RStudio Server	Open Source License	FREE
RStudio Server Pro	Commercial License	\$9,995 per year
RStudio Server Pro + RStudio Connect	Commercial License	\$29,995 per year

Below the table, there are five green buttons with white text: "DOWNLOAD Learn More", "BUY Learn More", "DOWNLOAD Learn More", "DOWNLOAD Learn More", and "TALK Learn More".

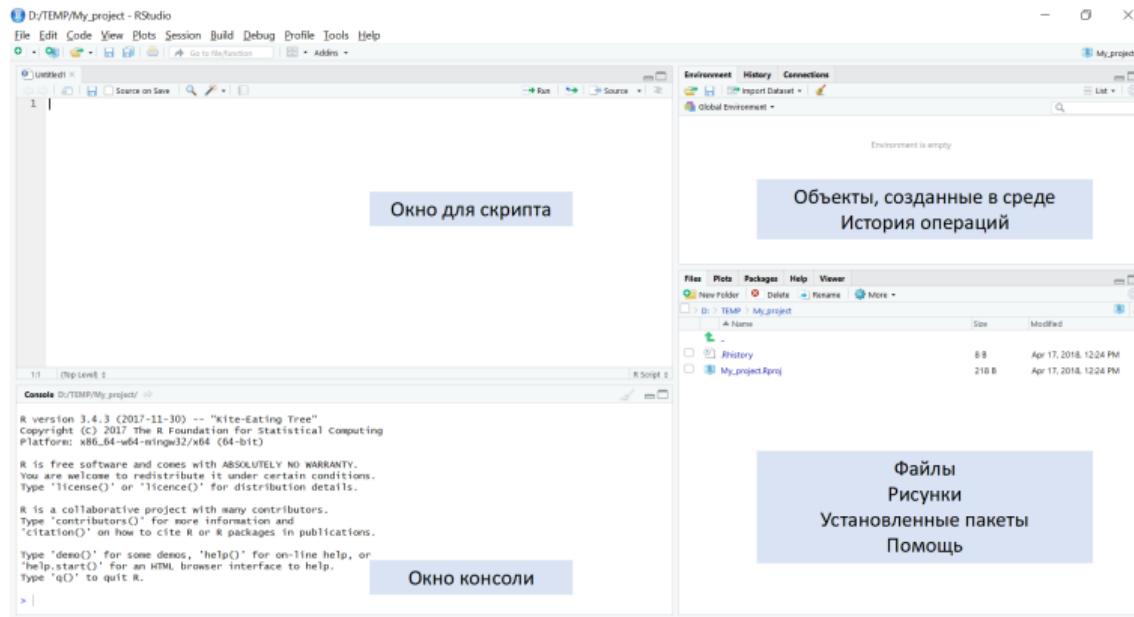
At the bottom left, there is a sidebar with two sections: "Integrated Tools for R" and "Priority".

Начальный вид окон RStudio

Сразу после открытия создаем окно для скрипта: File -> New File -> R Script



Четыре основных окна RStudio

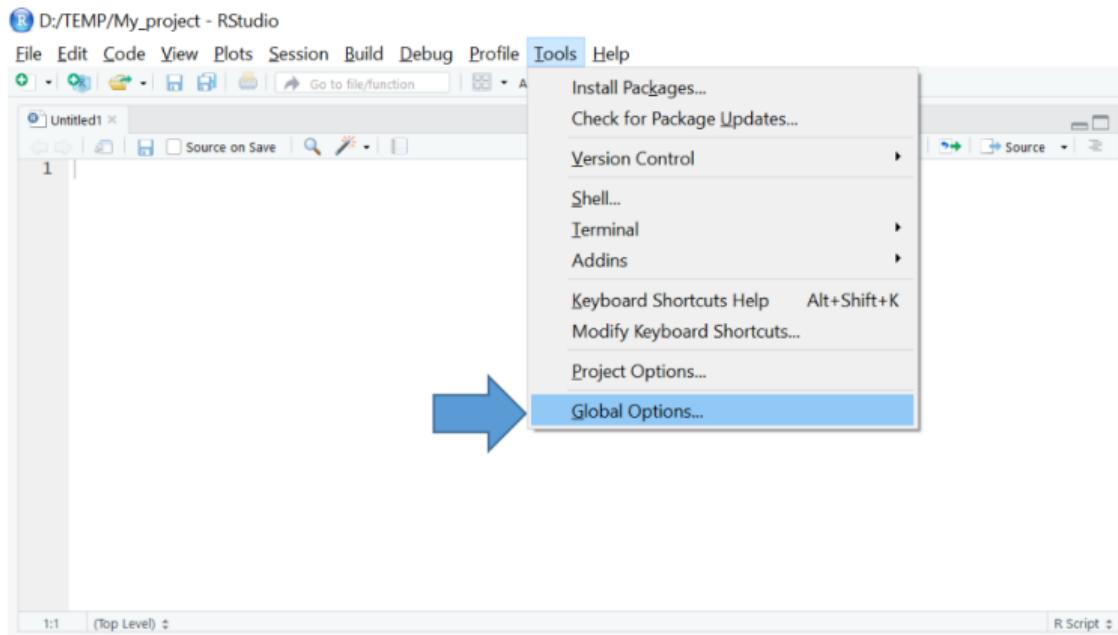


Вид окон во время работы

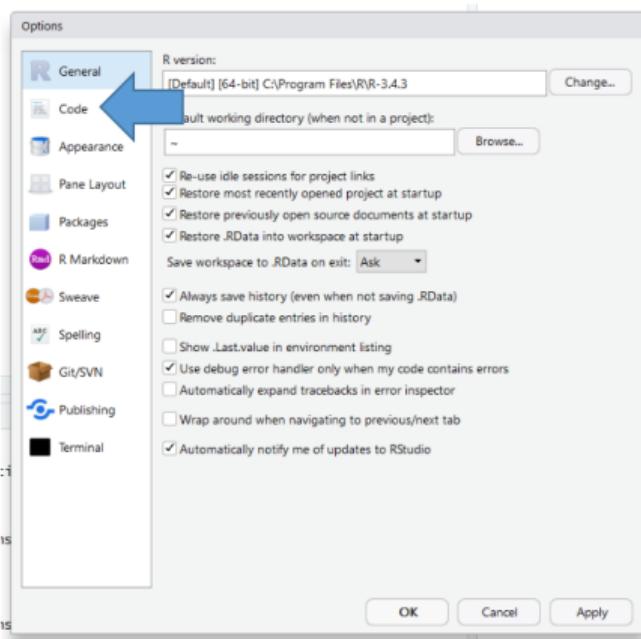
The screenshot shows the RStudio interface with four main windows:

- Source Editor:** Shows an R script named "Untitled.R" with code for generating a plot. A blue arrow points from the text "Наш код" (Our code) to the script window.
- Environment:** Displays the "dat" dataset with 21 observations and 2 variables. A blue arrow points from the text "Объекты" (Objects) to the environment window.
- Console:** Shows the R command history and output. It includes error messages about ggplot2 not being found, and the resulting numerical data and plot command. A blue box highlights the text "Численные и текстовые результаты; Сообщения (в том числе об ошибках)" (Numerical and textual results; Messages (including errors)).
- Plots:** Displays a parabolic plot of $y = 2x^2 - 3x + 5$ for $x \in [-10, 10]$. A blue box highlights the text "Изображения, полученные при исполнении кода" (Images obtained during code execution).

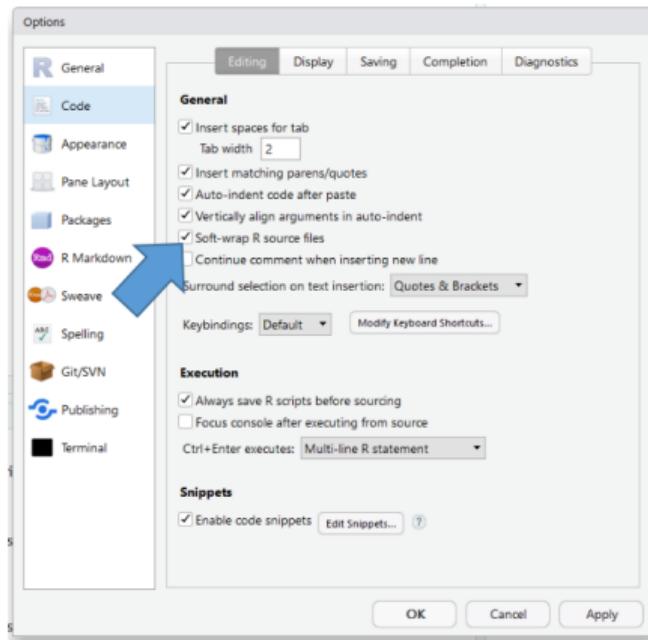
Важные настройки RStudio



Настройка окна скрипта



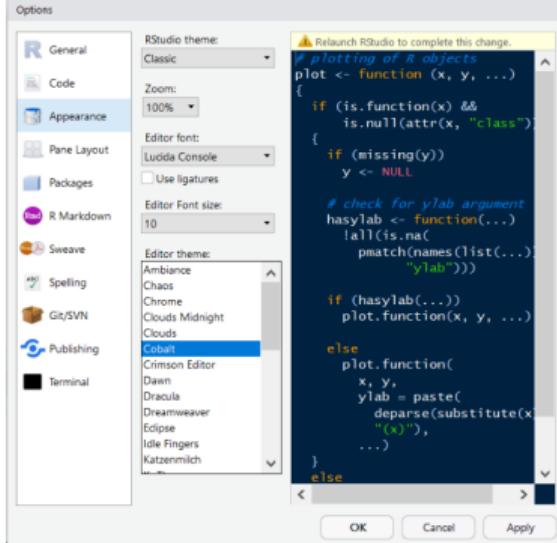
Перенос в отображении длинных строк



Ставим галочку рядом с опцией

Soft-wrap R source files

Вид RStudio по своему вкусу



Tools -> Global options -> Appearance

Организация рабочего пространства

В.М. Хайтов, к.б.н.

До начала работы

1. Создайте папку, где будут храниться ВСЕ материалы этого курса. Эта папка будет нашей рабочей директорией.

Например:

D:\Literature_programming\

В эту папку помещайте ВСЕ файлы со скриптами.

Отсюда же запускайте эти файлы из проводника или иного файл-менеджера.

2. Внутри папки Literature_programming создайте папку data.

В нее помещайте все файлы с данными для анализа.

D:\Literature_programming\data\

Открываем готовый скрипт

Скачайте с сайта файл `first_steps.R` и поместите его в рабочую директорию.

Откройте этот скрипт в RStudio

`File -> Open File`

Кракозябры вместо русских букв?

Нужно поменять кодировку

File -> Reopen with Encoding

Из предложенного списка выберете UTF-8.

В том же окне поставьте галочку "Save as default encoding for source files"

Рабочая директория

R всегда “нацелен” на определенную папку в вашей операционной системе – эта папка называется рабочая директория.

Это удобно: все файлы рабочей директории можно открыть, указав путь не от корневой папки системы, а от этой рабочей директории. Это называется относительный путь.

Проверка пути к рабочей директории

По умолчанию R считает, что рабочая директория – это основная пользовательская папка в вашей системе.

Для проверки правильности установки рабочей директории в окне скрипта выполните команду `getwd()` (get work directory).

```
getwd()
```

Для этого поставьте текстовый курсор на строчку с этой командой и нажмите `Ctrl + Enter`.

Установка пути к рабочей директории

Допустим, нашей рабочей директорией должна быть D:\Coursera\Linmodr

Способ 1:

Выберите в меню Session -> Set working Directory -> Choose Directory...

Способ 2:

Используем команду setwd()

```
setwd('D:/Coursera/Linmodr/')
```

Как получить помощь

В.М. Хайтов, к.б.н.

Просить помощи – это нормально



{<https://www.shutterstock.com/ru/image-photo/woman-drowns-sea-698646385>}

Работая в среде R, необходимо постоянно обращаться за помощью.

Три основных способа получения помощи:

- Локальная справка R
- Google
- Stack Overflow

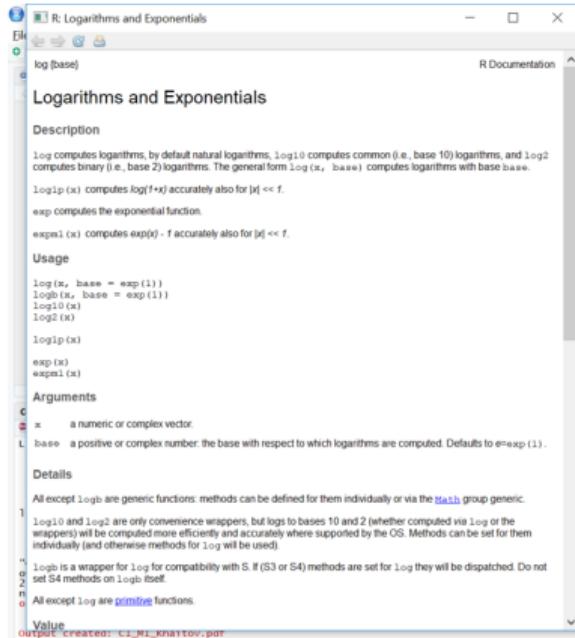
Локальная справка R

Предположим, вам необходимо получить справку по функции `log()`

```
?log      # Вызов справки через "?"
help(log) # Вызов справки с помощью функции help()
```

То же самое можно сделать, если поставить текстовый курсор на название функции и нажать “F1”.

Окно справки



Разделы:

- **Description** – общее описание функции
- **Usage** – шаблон использования функции
- **Arguments** – аргументы (параметры) функции
- **Details** – Дополнительные детали

Окно справки



Разделы:

- **Value** – результаты работы функции
- **References** – ссылки на первоисточники
- **See Also** – похожие функции
- **Examples** – примеры использования функции

Код из раздела Examples можно выполнять

Строки из раздела “Examples” можно скопировать в буфер обмена, вставить в скрипт и выполнить.

```
log(exp(3))  
# [1] 3  
  
log10(1e7) # = 7  
# [1] 7
```

Форма записи чисел $1e7 = 1 \cdot 10^7$ – это так называемая экспоненциальная запись (scientific notation).

Google – ваш лучший друг

Google search results for "r function log".

Search bar: r function log

Filter buttons: Все, Картинки, Карты, Видео, Новости, Ещё, Настройки, Инструменты

Results count: Результатов: примерно 64 800 000 (0,29 сек.)

1. **log function | R Documentation**
https://www.rdocumentation.org/packages/base/.../3.4.../log ▾ Перевести эту страницу
log computes logarithms, by default natural logarithms, log10 computes common (i.e., base 10) logarithms, and log2 computes binary (i.e., base 2) logarithms. The general form log(x, base) computes logarithms with base base . log1p(x) computes log(1+x) accurately also for |x|<<1. exp computes the exponential function.

2. **Quick-R: Built-in Functions**
https://www.statmethods.net/management/functions.html ▾ Перевести эту страницу
Function, Description. abs(x), absolute value. sqrt(x), square root. ceiling(x), ceiling(3.475) is 4. floor(x), floor(3.475) is 3. trunc(x), trunc(5.99) is 5. round(x, digits=n), round(3.475, digits=2) is 3.48. signif(x,

Сообщество Stack Overflow

Google search results for "r function log stackoverflow".

Search query: r function log stackoverflow

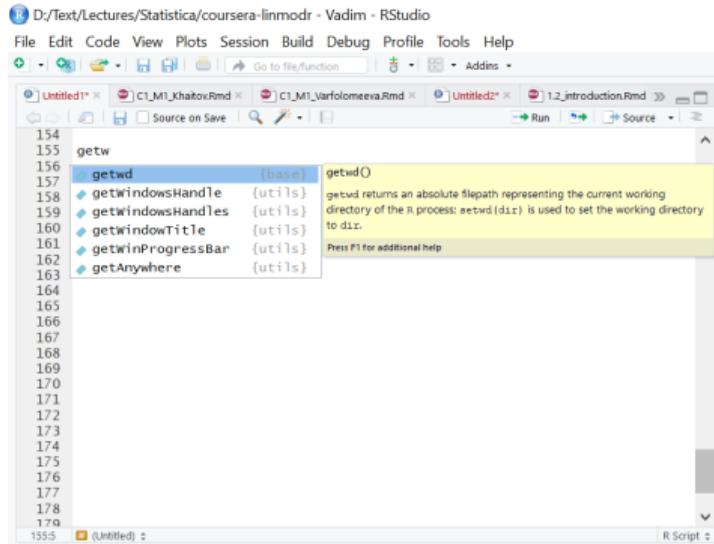
Results:

- r - Specify the base of the log function - Stack Overflow**
<https://stackoverflow.com/.../specify-the-base-of-the-log-fun...> ▾ Перевести эту страницу
16 июн. 2012 г. - Arguments. x a numeric or complex vector. base a positive or complex number: the base with respect to which logarithms are computed. Defaults to e=exp(1). So you call function log with the second parameter 10 (base that you want). For example: > log(10,10) [1] 1 ...
What is the meaning of the dollar **sign "\$"** in R ... Ответов: 3 2 мар 2017
natural logarithm - R Commander ERROR: could not ... Ответов: 5 5 апр 2016
nonlinear functions - R: Using equation with natural ... Ответов: 2 19 июн 2014
How to create an **R function** programmatically? Ответов: 2 19 окт 2012
Другие результаты с сайта stackoverflow.com

Сообщество Stack Overflow

The screenshot shows a web browser window with the URL <https://stackoverflow.com/questions/11059853/specify-the-base-of-the-log-function>. The page title is "r - Specify the base of the log function". The main content is a question titled "Specify the base of the log function". The question text reads: "I am bit confuse with R log function: $\log(10) = 1$ in base 10 but when I type this in R $\log(10) = 2.302585$ So my question is how do I implicitly tell the base in R log function." Below the question are standard Stack Overflow controls: upvote (-3), downvote (1), share, improve this question, and edit. The question was asked 5 years, 9 months ago and has been viewed 4,473 times. A sidebar on the right displays "UPCOMING EVENTS" for the "2018 Community Moderator Election" which ends in 6 days.

RStudio может подсказывать и сама



Начните набирать название функции и нажмите Tab или Ctrl + Space.

Не бойтесь сообщений об ошибках и предупреждений

Разобраться в них обычно удается даже при минимальном знании английского языка.

```
| sqr(4)
```

```
Error in sqr(4) : could not find function "sqr"
```

```
| sqrt(-1)
```

```
[1] NaN  
Warning message:  
In sqrt(-1) : NaNs produced
```

Что делать, если консоль неистово плюсует

Знак + в консоли значит, что во введенной команде чего-то не хватает, и R ожидает продолжения.

Один из самых частых случаев – потерялась закрывающая скобка.

```
> sqrt(4  
+  
+  
+  
+
```

Для выхода из этого положения нажмите Esc.

Установка пакетов

В.М.Хайтов, к.б.н.

Что такое пакеты

Пакет - набор функций, предназначенных для решения определенного класса задач.

Примеры пакетов:

- ggplot2 – набор функций для рисования графиков
- dplyr, tidyr – средства управления данными и их преобразования в удобную форму
- car, MASS, lmer – инструменты регрессионного анализа

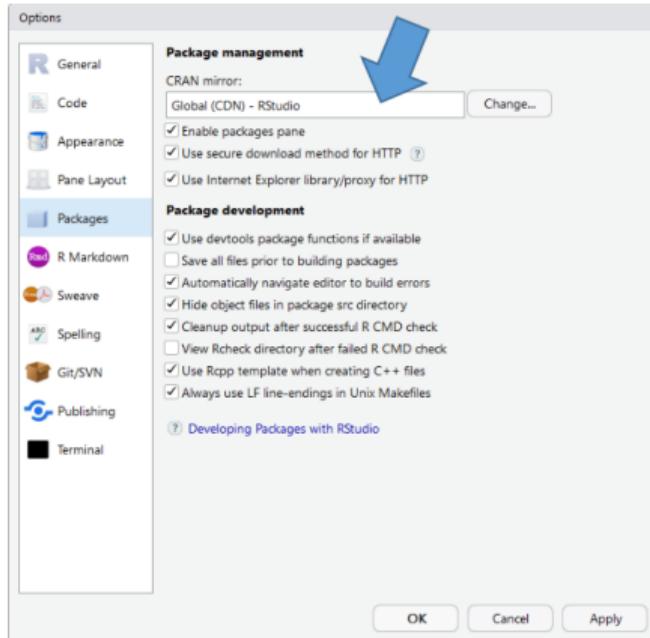
Установка пакетов

Установка пакетов происходит из Сети. Лучше работать, имея постоянное подключение.

Главный репозиторий находится на сайте CRAN (<https://cran.r-project.org/>).

Если вы работаете под Windows, то запускайте RStudio от имени администратора.

Выбор зеркала



Из множества зеркал CRAN лучше выбрать ближайшее к вам.

Tools -> Global options -> Packages -> CRAN mirror

Установка пакетов в R

```
install.packages('ggplot2') # Устанавливаем пакет ggplot2  
install.packages('tidyverse') # Устанавливаем пакет tidyverse
```

Пакеты достаточно один раз установить в локальную библиотеку.

Команда `library()` активирует нужный пакет, если он уже установлен, и выполняется один раз за сеанс работы в R.

```
library(ggplot2) # Активируем пакет ggplot2
```

Список установленных пакетов



List of installed packages			
Name	Description	Version	Action
gridBase	Base Graphics Device	3.4.3	...
grDevices	The R Graphics Devices and Support for Colours and Fonts	3.4.3	...
grid	The Grid Graphics Package	3.4.3	...
gridExtra	Miscellaneous Functions for "Grid" Graphics	2.3	...
gtable	Arrange 'Grobs' in Tables	0.2.0	...
highr	Syntax Highlighting for R Source Code	0.6	...
Hmisc	Harrell Miscellaneous	4.1-1	...
htmlTable	Advanced Table for Markdown/HTML	1.11.2	...
htmltools	Tools for HTML	0.3.6	...
htmlwidgets	HTML Widgets for R	1.2	...
httr	Tools for Working with URLs and HTTP	1.3.1	...
jpeg	Read and write JPEG images	0.1-8	...
jsonlite	A Robust, High Performance JSON Parser and Generator for R	1.5	...
kernlab	Kernel-Based Machine Learning Lab	0.9-25	...
KernSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.23-15	...
knitr	A General-Purpose Package for Dynamic Report Generation in R	1.20	...
koRpus	An R Package for Text Analysis	0.10-2	...
labeling	Axis Labeling	0.3	...
lambda.r	Modeling Data with Functional Programming	1.2.2	...
lattice	Trellis Graphics for R	0.20-35	...
latticeExtra	Extra Graphical Utilities Based on Lattice	0.6-28	...

Лаборатория экологии
морского бентоса
(гидробиологии)



Кандалакшский
государственный
заповедник



Санкт-Петербургский
государственный
университет