Introduction to R

Marco Chiapello

June 18, 2016

Center for Proteomics
University of Cambridge
mc983@cam.ac.uk

Overview

Introduction

Basic concepts in R

- Vector

Introduction

What R is?

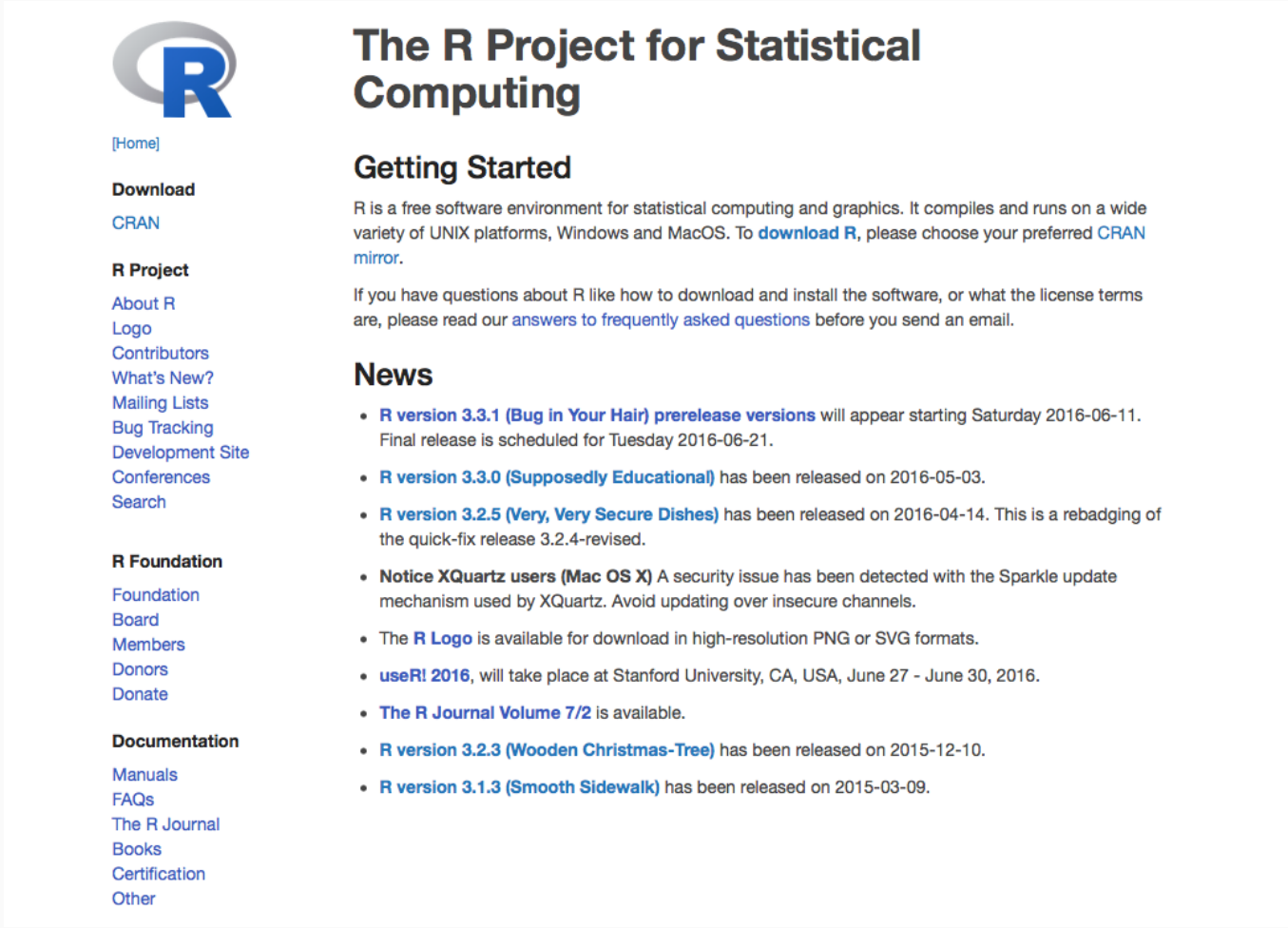
The R Project for Statistical Computing

- R is a free software environment for statistical computing and graphics
- Open source and cross platform (UNIX platforms, Windows and MacOS)
- Extensive graphics capabilities
- Diverse range of add-on packages
- Active community of developers
- Thorough documentation

What R is?

The R Project for Statistical Computing

You can find R here: `https://www.r-project.org`



[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Mailing Lists
Bug Tracking
Development Site
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

Documentation

Manuals
FAQs
The R Journal
Books
Certification
Other

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **R version 3.3.1 (Bug in Your Hair) prerelease versions** will appear starting Saturday 2016-06-11. Final release is scheduled for Tuesday 2016-06-21.
- **R version 3.3.0 (Supposedly Educational)** has been released on 2016-05-03.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, will take place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.

What R is?

The R Project for Statistical Computing

- R version 3.3.1 (released 2016-06-21)
- Currently, the CRAN (Comprehensive R Archive Network) package repository features 8609 available packages
  - `https://cran.r-project.org/web/packages/available_packages_by_name.html`
- Currently, the Bioconductor repository features 1211 available packages
  - `http://www.bioconductor.org`
- Executed using command line, or a graphical user interface (GUI)
- On this course, we use the RStudio GUI
  - `www.rstudio.com`

Getting started

- `R` is a program which, once installed on your system, can be launched and is immediately ready to take input directly from the user
- There are two ways to launch R:
  1. From the command line (particularly useful if you're quite familiar with Linux)
  2. As an application called RStudio (very good for beginners)

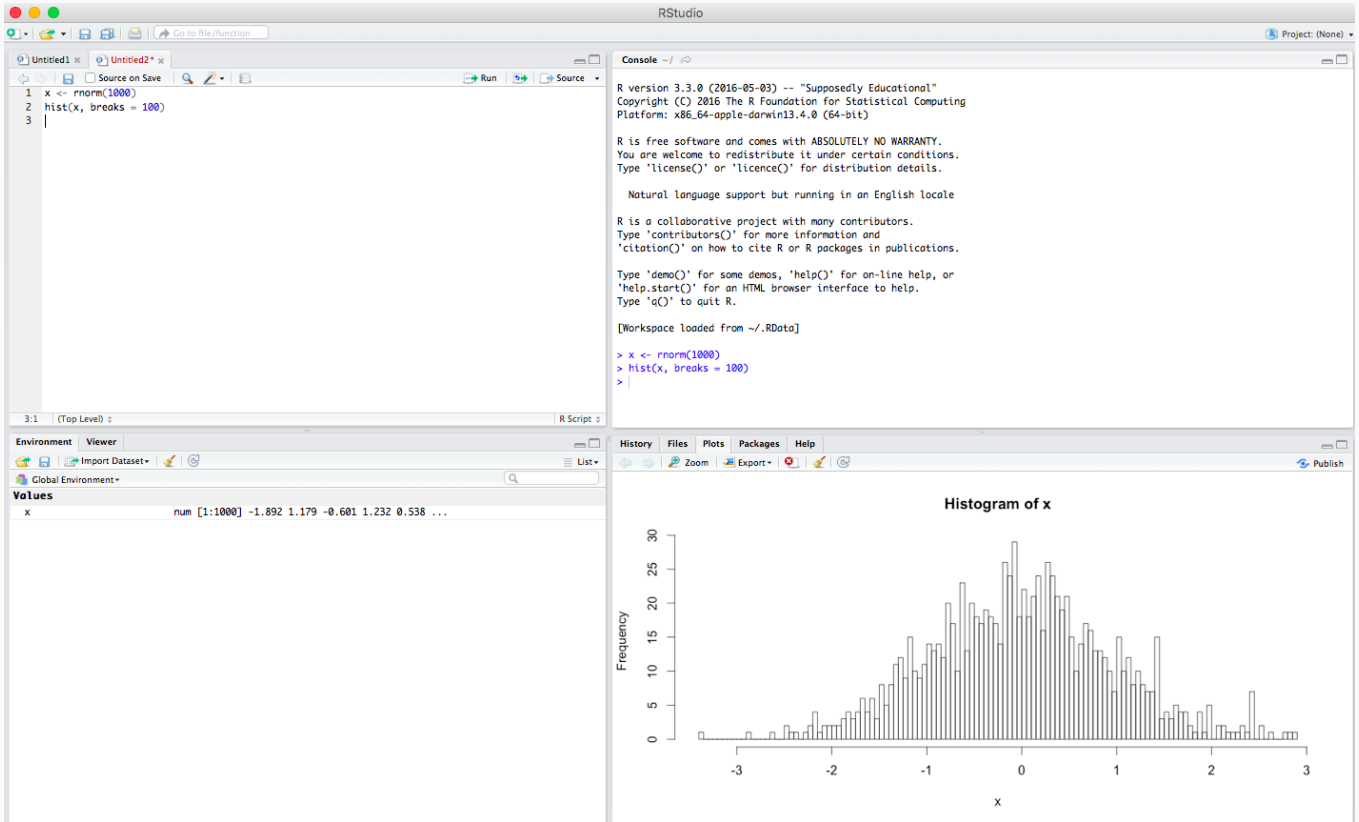
Launch R

`R` can be launched in 2 ways:

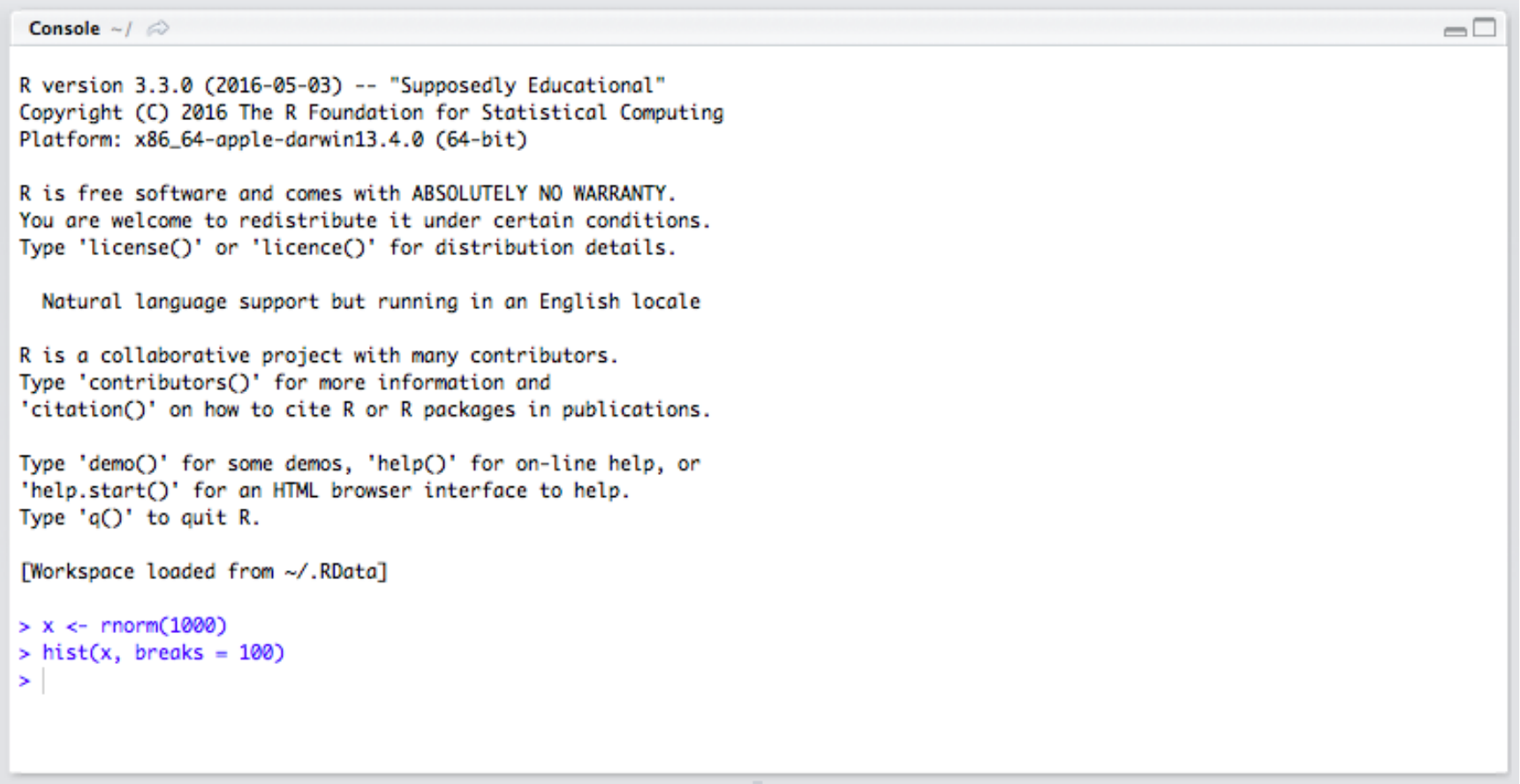
1. From command line
    - To start `R` you need to enter the console (also called terminal or shell)
    - To start `R`, at the prompt simply type: `R`
2. Using RStudio
    - To launch RStudio, find the RStudio icon and double-click

RStudio

Since we will use RStudio in this course, let's have a look of the program

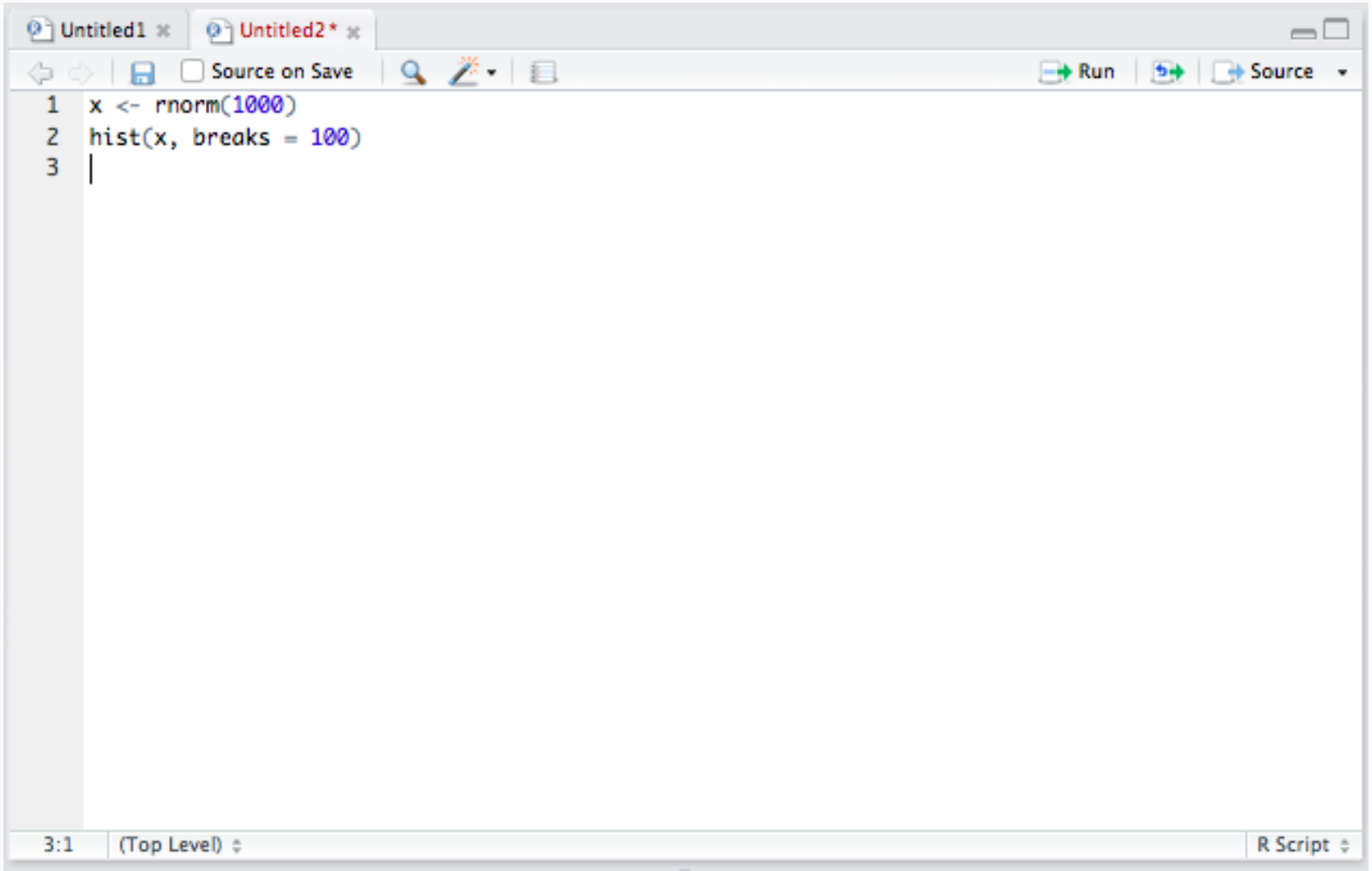


R console



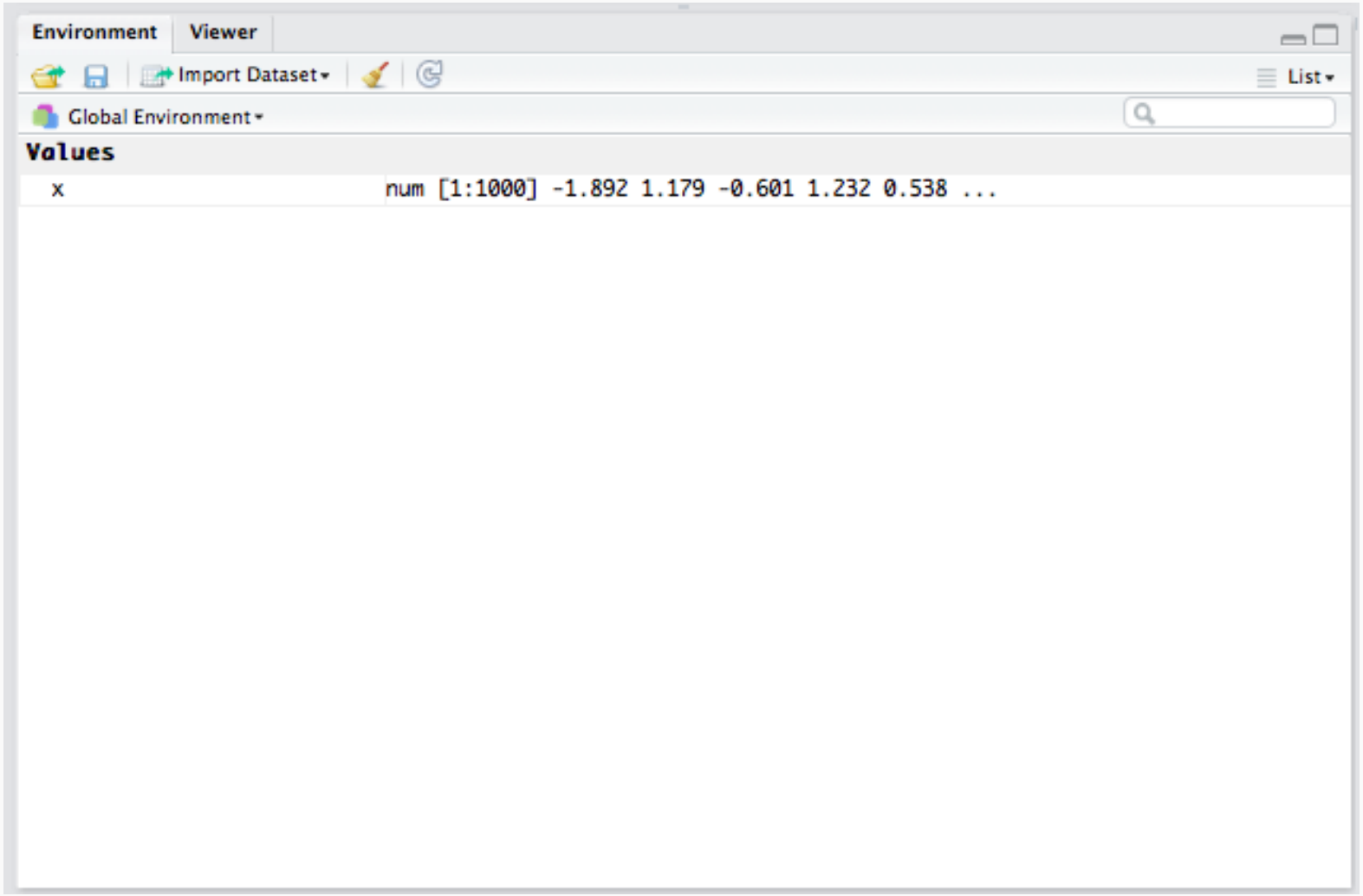
It is the place where you can interactively run R commands

Source editor for R scripts



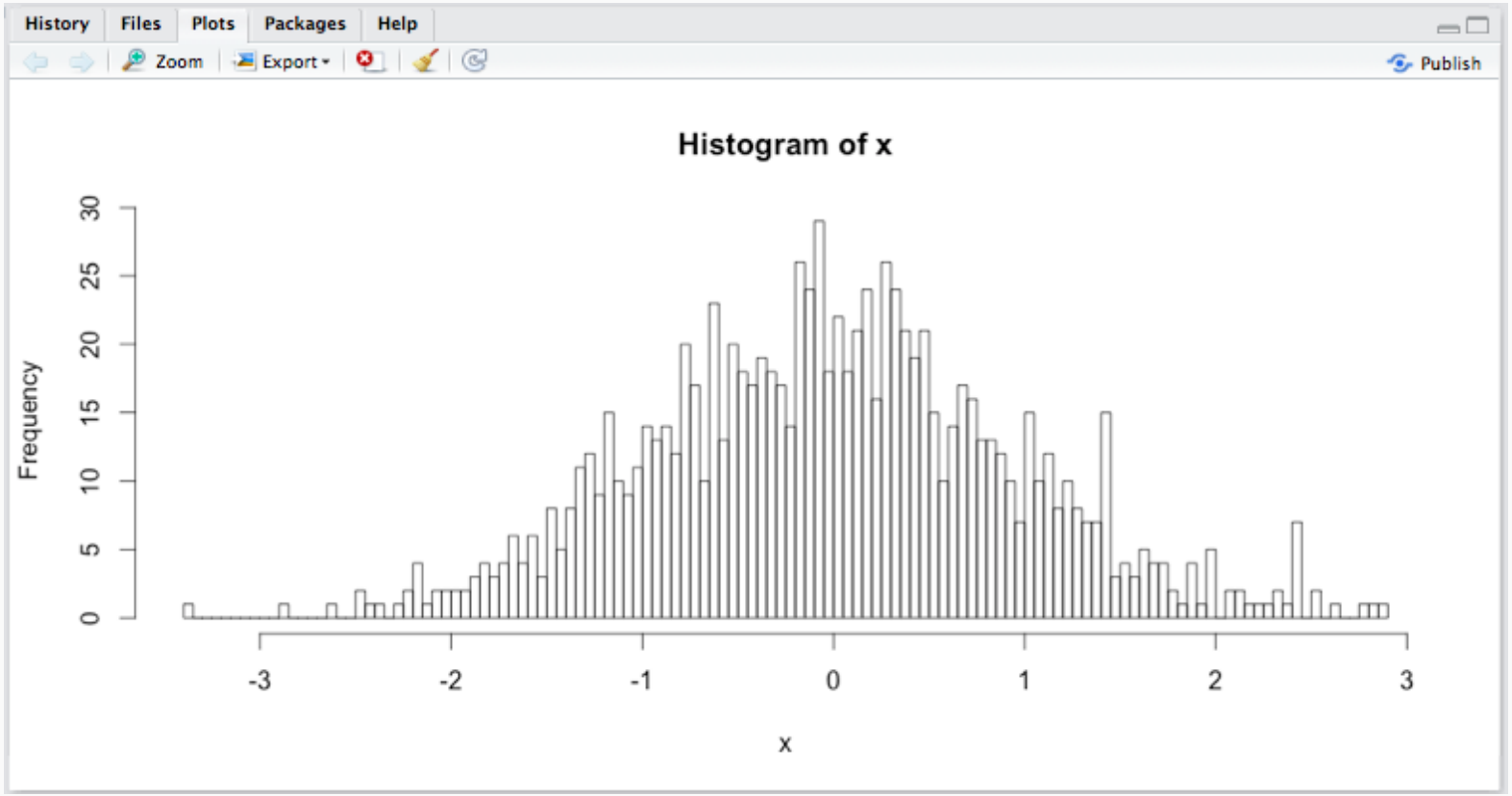
It is the place where you can write your scripts

Workspace



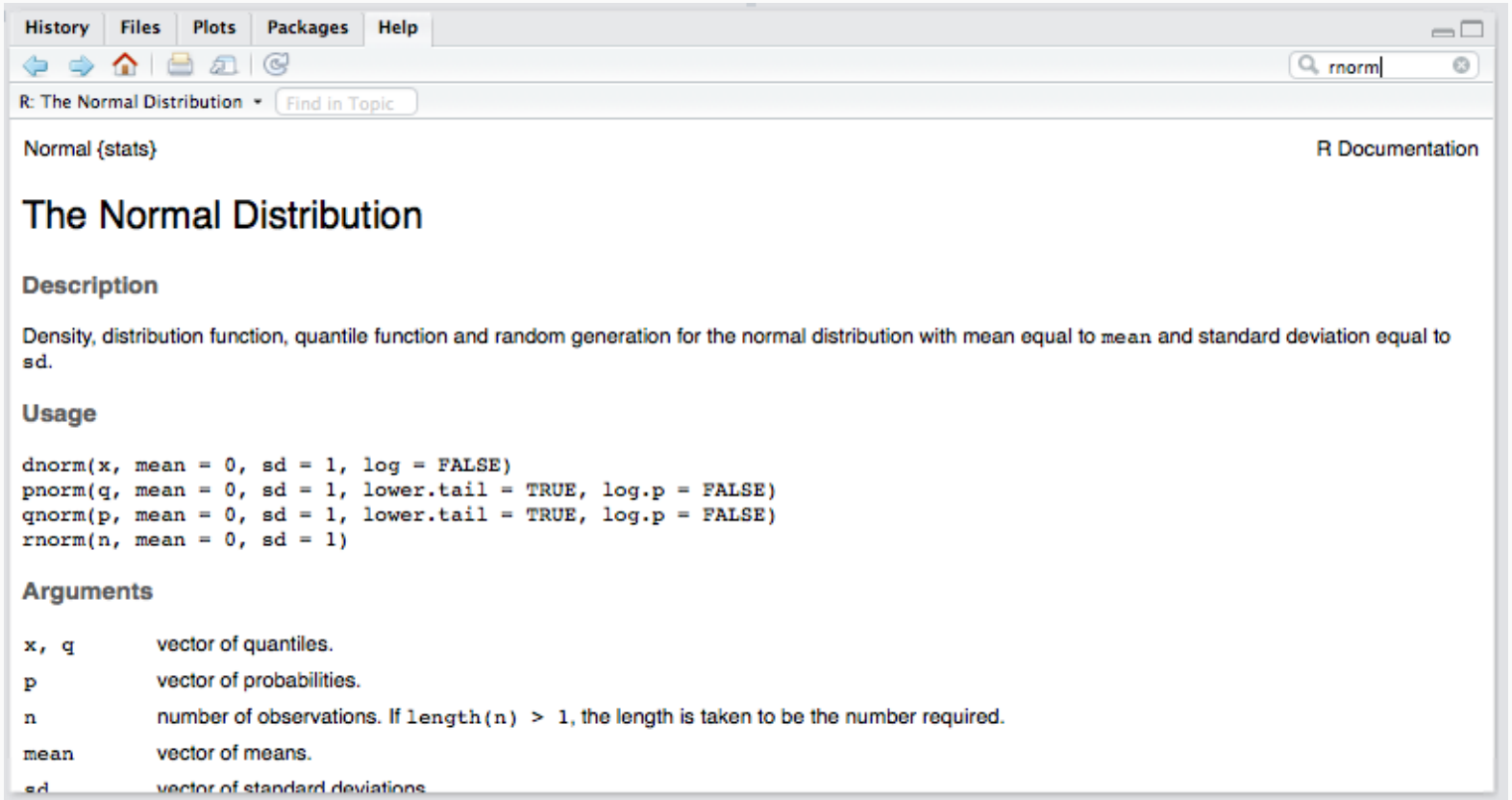
It is the place where you can view object in the global environment

Plot pannel



It is the place where you can view your plots

R help

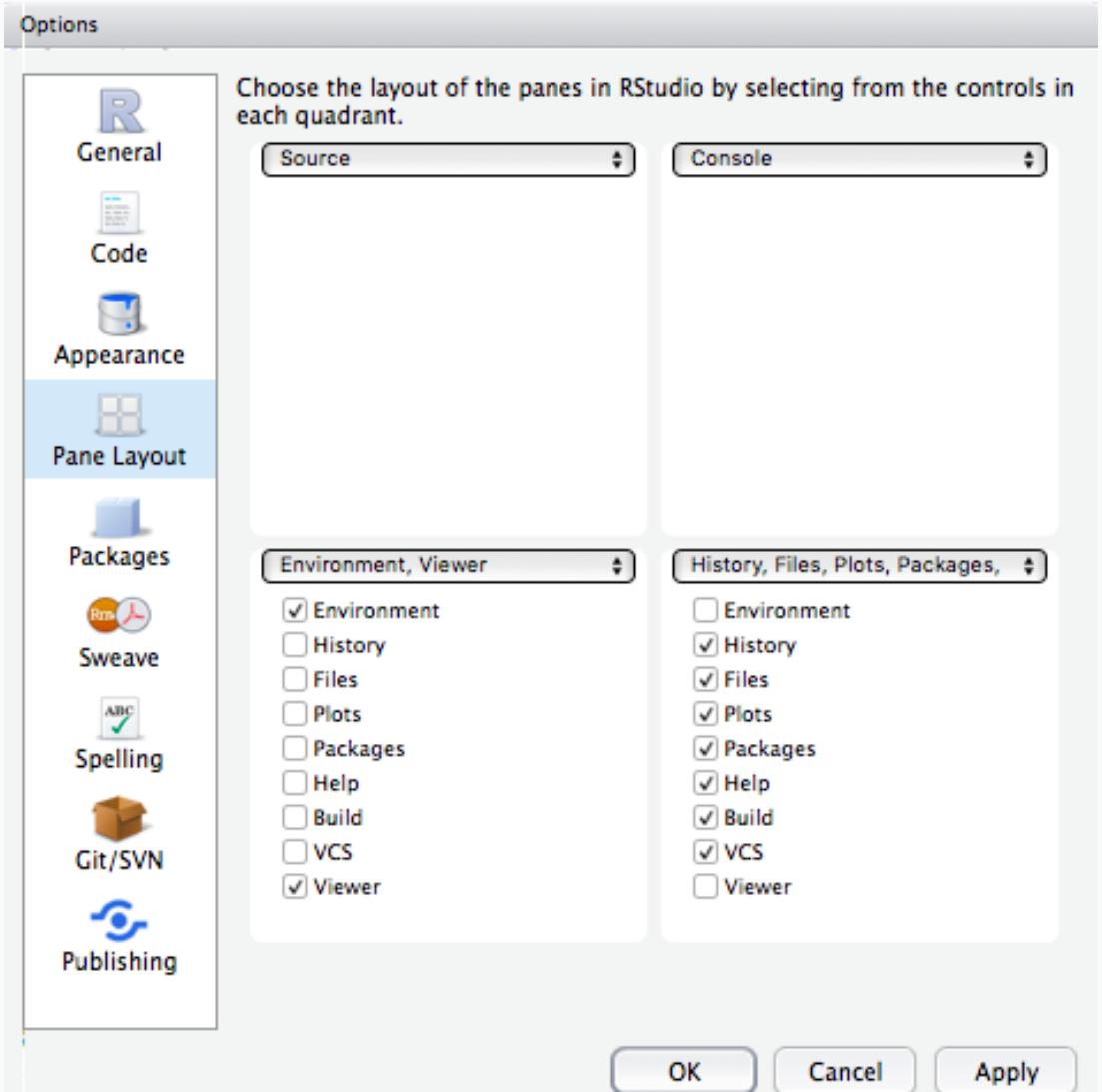


It is the place where you can find help

RStudio

The GUI is divided into 4 main sub-windows

These sub-windows are customizable



Basic concepts in R

Numbers

The command line can be used as a calculator

```
> 5 + 7
[1] 12
> 5 - 7
[1] -2
> 5 * 7
[1] 35
> 5 / 7
[1] 0.7142857
```

Note: The number in the square brackets is an indicator of the position in the output

Numbers

You can solve simple or complex calculations

```
> (((20/5)^2)-((5+1/3+4/5-10)*(2-34))-20)
[1] -127.7333
```

But, of course, R is not a calculator

Variables

A **variable** is a letter or word which takes (or contains) a value.

We use the assignment 'operator', <-

- We can assign a number to a variable

```
> x <- 5
> x
[1] 5
```

- We can assign the result of an operation to a variable

```
> y <- 5 + 7
> y
[1] 12
```

Variables

- We can assign use the variables to perform calculation

```
> x + y
[1] 17
```

- We can assign the change the content of the variable

```
> x
[1] 5
> x <- x - y
> x
[1] -7
```

Function

Functions in R perform operations on arguments (the input(s) to the function).
Arguments are always contained in parentheses, i.e. curved brackets (), separated by commas.

```
> sum(3, 4, 5, 6)
[1] 18
> max(3, 4, 5, 6)
[1] 6
> min(3, 4, 5, 6)
[1] 3
```

Function extention

R contains a lot of pre-builtin functions, but through the so called *packages* is possible extend the R functionalities enormously. Alternatevely, you can write your own function

```
> summ <- function(a,b){ a + b }
> summ(1,2)
[1] 3
```

Vector

Vector