

Introduction to R and RStudio

1 How to get started?

1.1 What is R?

R is a programming language for statistical analysis, data science, data management, and visualisation. R has multiple advantages over its competitors, of which the following two seem to be the most important.

- **Open Source:** R is an open-source software. No user or subscription fees are required. Anyone can download the basic R packages on their computer and start using it right away (more on that later).
- **Vast array of packages:** R has a vast array of available packages (~10K), predominantly written and published by users. Thanks to that virtually all operations regarding data management, data wrangling, statistical analysis, machine learning, and many more can be performed using R.

Unfortunately, R is not a perfect solution for all problems. Again, two main issues regarding data management with R can be named:

- **Speed:** R and its packages are simply slower than their counterparts in Python or MATLAB.
- **Data Handling:** R stores the objects in the physical memory and, in general, needs more memory to operate than, for example, Python. Thus, R is not an ideal choice when dealing with Big Data or problems of high computational complexity.

Nevertheless, thanks to its open source license and broad area of applications, R is widely used within the data science and research community.

1.2 What is RStudio?

RStudio is a user interface for the statistical programming software R. While R can be used without any user interface (it has a built-in command window interface), it is much easier to learn R programming and use R in general with the help of some user interface tools, out of which RStudio is by far the most popular.

1.3 Installation

While in labs, both R and RStudio are already installed it is highly recommended that you install R on your PCs as well. R is supported by UNIX, Windows, and Mac.

To start working with R and RStudio, first you should download R. Go to <http://cran.us.r-project.org/> and follow the instructions for your operating system. Using default settings is recommended.

Afterwards go to <https://rstudio.com/> and download the RStudio Desktop.

Install both R and RStudio. Run the latter. In most cases RStudio will automatically find the most recent R version you have installed and you can start working right away. If RStudio does not find the R folder automatically you will be prompted to paste the appropriate path.

2 RStudio interface

RStudio interface is divided into four main sections: Script, Workspace, Console, and Help.

Script

Script is a document, where one stores R commands. This allows to reference the commands later or repeat the analyses.

Console

Output of all your operations will appear in the console window. Here you can write R commands as well; however, it is highly recommended to use scripts rather than the console only. Still, the console is useful for ad-hoc operations, for example to verify whether the script runs smoothly.

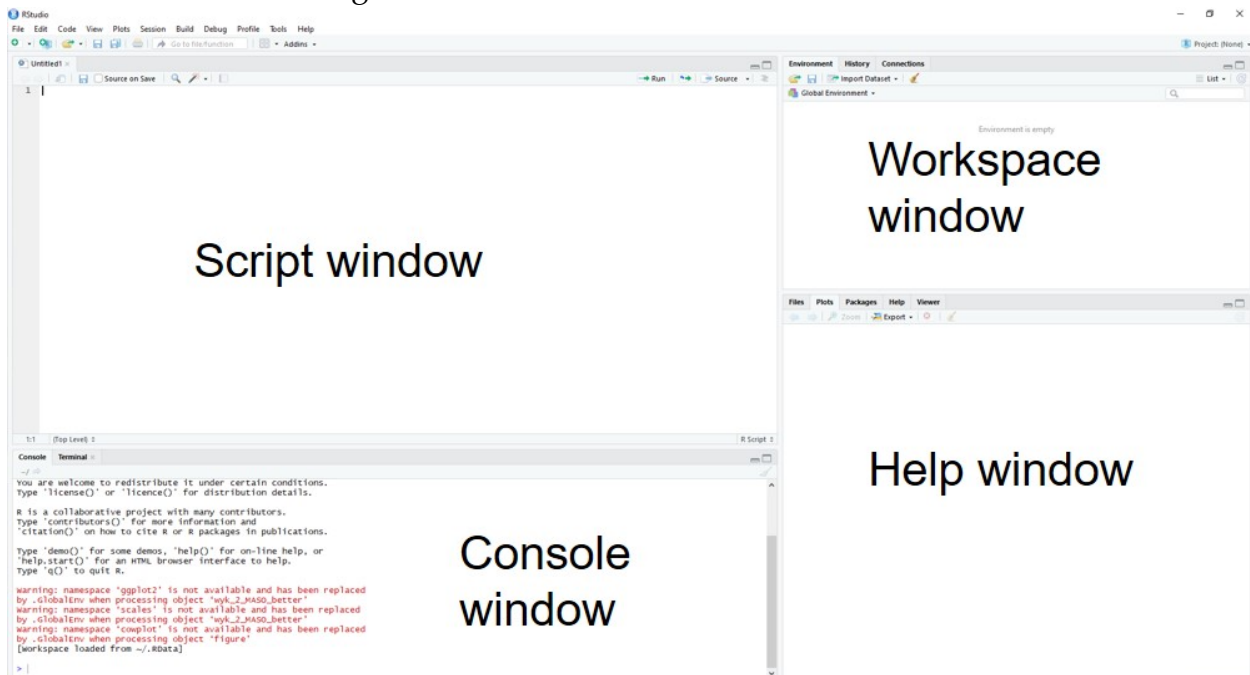
Workspace

The workspace window lists all objects that you have created (uploaded) that are ready to use, including variables, datasets, vectors, lists, and user-created functions.

Help

The help window has several useful tabs. The first is the FILES tab, which allows you to quickly navigate between your R scripts and data files. The PLOTS tab is where the ad-hoc plots you make will appear. This also allows for plot export (to .png, .jpg files for example). The PACKAGES tab allows you to browse the R packages installed on your computer as well as to load those packages. Finally, the HELP window allows you to search for in-built function and package help manuals.

Figure 1: RStudio interface - basic windows



2.1 Script window tips

To open a new script you can:

- go to File, New File, New Script
- press the white sheet with a green plus positioned directly under the File tab
- Press `ctrl+shift+N`

To save script you can:

- press `ctrl+S`
- use one of the save icons above the script body

To load script you can:

- press `ctrl+O`
- press the file with a green arrow, directly under the Code tab

To run a script you can:

- Press "Run" in the top right corner of the script window
- Press `ctrl+A` and `ctrl+ENTER`

To run a part of the script you can:

- Select the part you want to run in your script and press ctrl+Enter
- Pressing ctrl+Enter while having nothing selected will run the current line of the script

To find/replace a phrase in a script you can:

- press ctrl+F
- press the magnifying glass icon on top of the script widow

Commenting:

- use the # sign to comment everything that appears in this line after the # sign
- if you have a part of the script selected (e.g. several lines) you can use ctrl+shift+c to comment the whole section

Creating reference points in a script:

- use ##### Title ##### to divide your script into more manageable parts. You can quickly jump between parts using the drop-down list below the script window.

Jumping between console and script

- use ctrl+1 to jump to the script window
- use ctrl+2 to jump to the console window

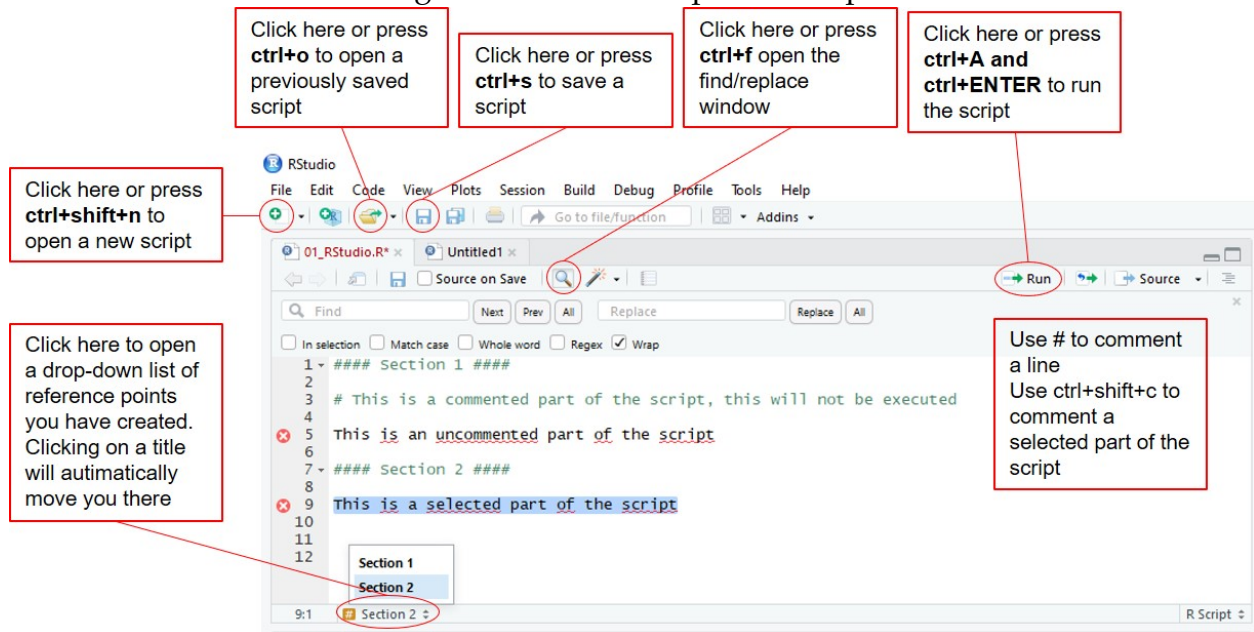
2.2 Console window tips

- If you see the ! sign at the bottom of the console window, it means that R is ready to Run
- Press Up and Down arrow keys to copy the most recently run lines into the console window
- Press ENTER to run the command you wrote

2.3 Workspace window tips

- Use the file icon to load a workspace (all data appearing in the environment tab)
- Use the save icon to save your workspace

Figure 2: RStudio script - basic tips



- Press a dataset to view it in the Viewer mode in the console or running it from the script
- Use the broom icon to clear your workspace
- Press the blue arrow next to a dataset to see a summary of it in the workspace window

2.4 Help window tips

- There are two search windows in the help tab. The one in the upper-right corner allows you to search for functions. The one in the upper-left corner allows you to look for specific phrases in a help file
- The plots tab allows you to export an image. Press the Export button

Figure 3: RStudio workspace - basic tips

The image shows the RStudio workspace interface with several callout boxes providing basic tips:

- Click here to load a workspace**: Points to the folder icon in the Environment pane.
- Click here to save a workspace**: Points to the floppy disk icon in the Environment pane.
- Click here to clear the workspace**: Points to the broom icon in the Environment pane.
- Click here to view the dataset in the Viewer**: Points to the 'A' dataset entry in the Data section.
- Click here to see a summary of the dataset**: Points to the 'X' dataset entry in the Data section.

The Environment pane shows the Global Environment with the following data and functions:

| Data | |
|-------------|-----------------------|
| A | 2 obs. of 2 variables |
| X | 2 obs. of 2 variables |
| column_1: | int 1 2 |
| column_2: | int 3 4 |
| values | |
| b | 19 |
| name | "Nicolas Cage" |
| Y | num [1:3] 1 2 3 |
| Functions | |
| hello_world | function (text) |

On the right side, there are three callout boxes with brackets pointing to the corresponding sections in the Environment pane:

- datasets**: Points to the 'Data' section.
- variables and vectors**: Points to the 'values' section.
- user functions**: Points to the 'Functions' section.