Brief introduction to R

Martín Brun

Public Sector Economics
Faculty of Economics and Business Studies
April 2021

Universitat Autònoma
de Barcelona

Why learn R?

Figure 1: Office in 1960s

Figure 2: Office in 1990s

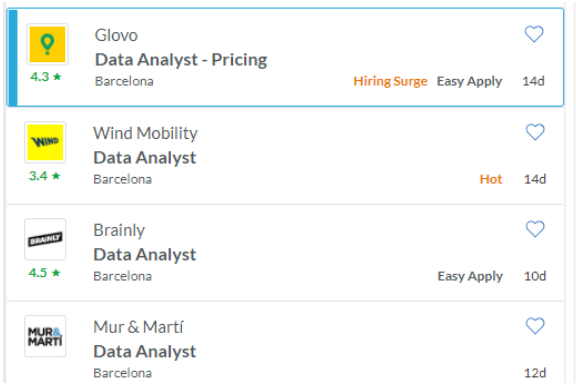
Why learn R?

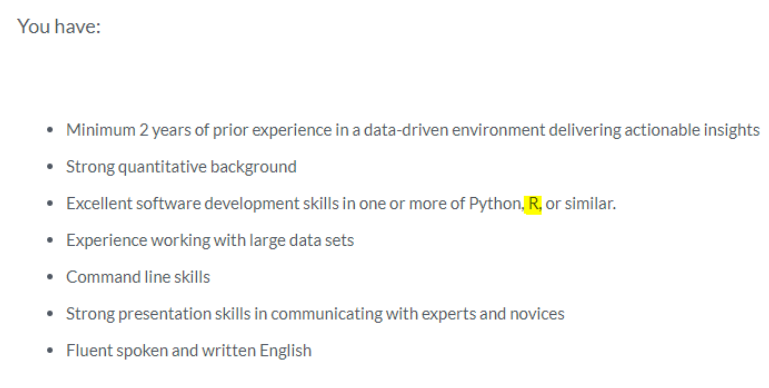
Figure 3: Office in 2020s

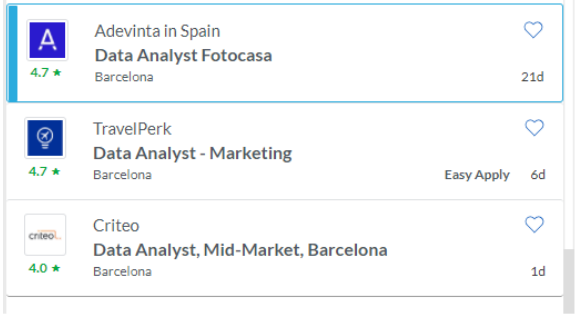
You will work with a computer.. almost exclusively

Why learn R?

Data Analyst job postings in Barcelona



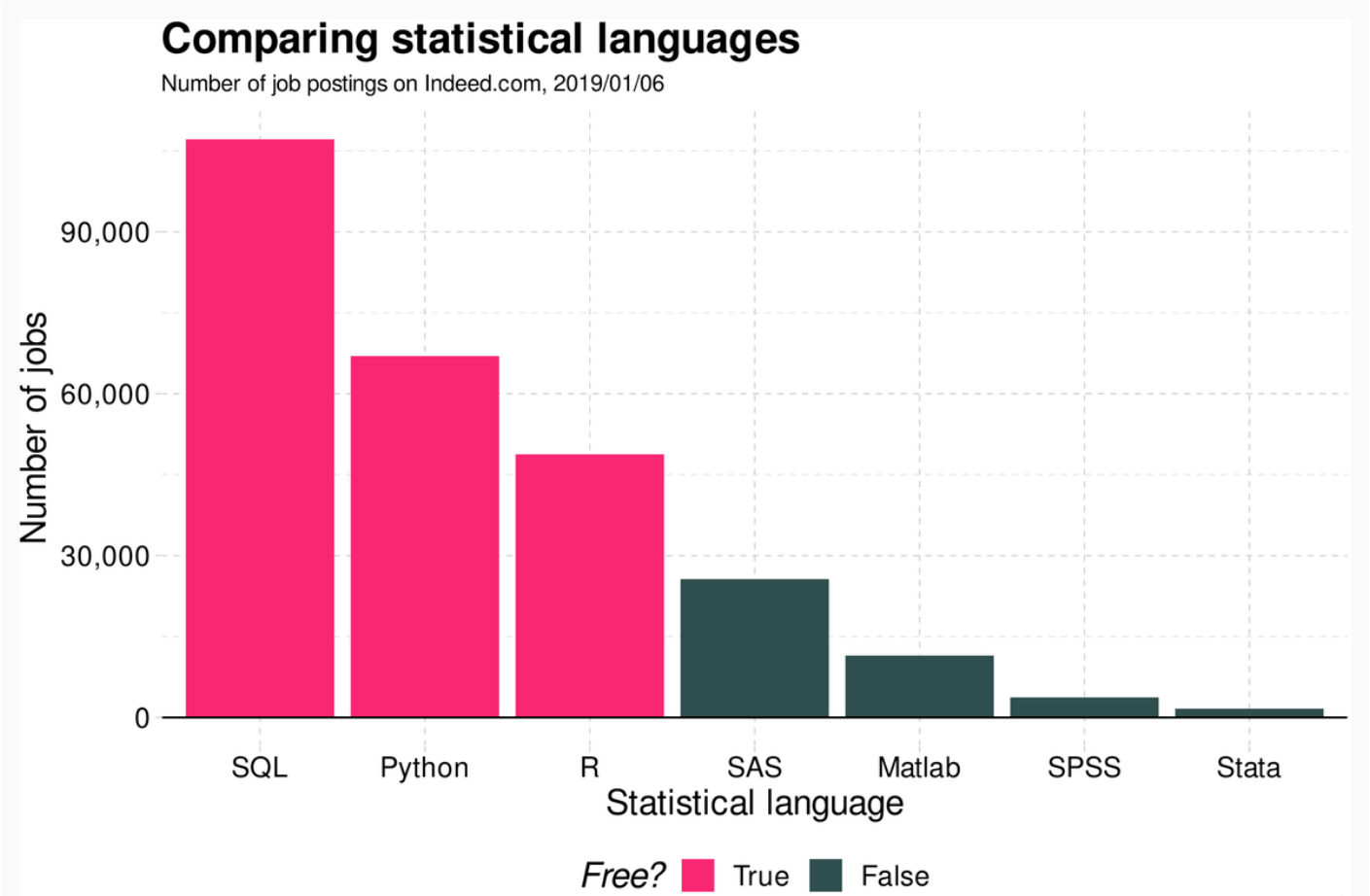




Source: Glassdoor

You will probably code

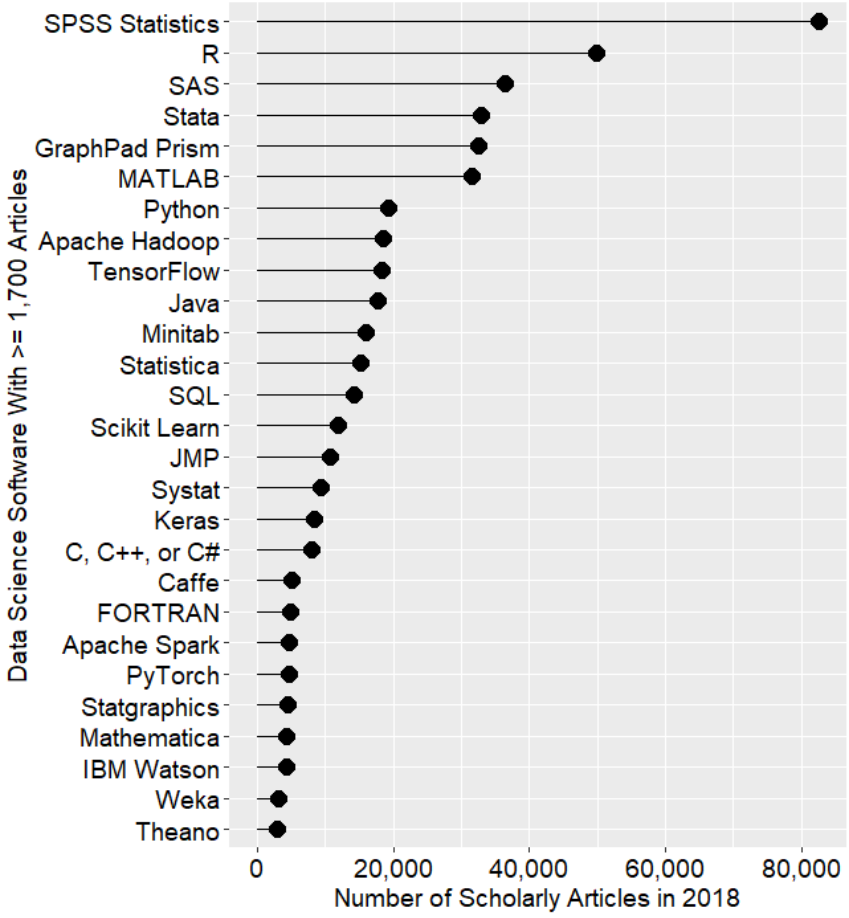
Why learn R?



Source: Indeed.com

R is not the only coding language, but is now a popular requirement for Data Science/Analysis

Why learn R?



Source: R4Stats

R is also popularly used for Research

Introduction

What is R?

R is an integrated suite of software facilities for

- data manipulation
- calculation
- graphical display

R is a **versatile** programming language, which can used within diverse environments

Introduction

First steps

Two options

- Use R online: RStudio cloud
  - Good for learning
  - Not recommended for working
- Install R into your computer: RStudio
  - Highly recommended
  - Allows to gradually adjust to your preferences

Introduction

RStudio

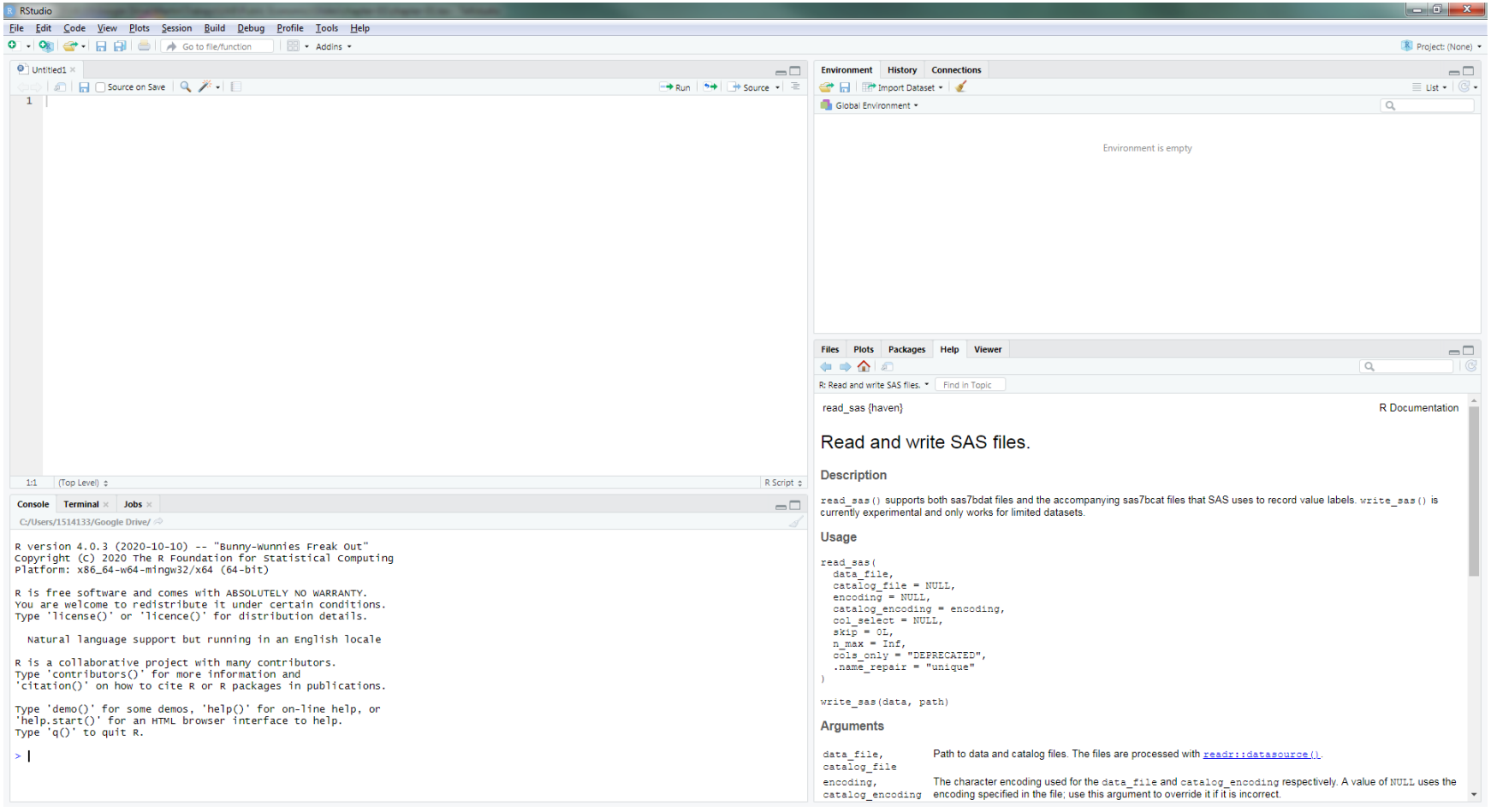
RStudio is an IDE (Integrated Development Environment)

It provides all the tools to use R language:

- code editor
- console
- help guides

Introduction

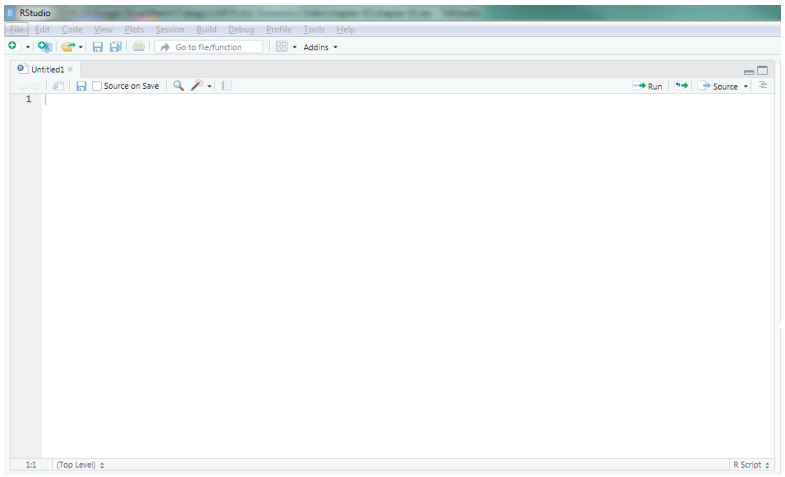
RStudio



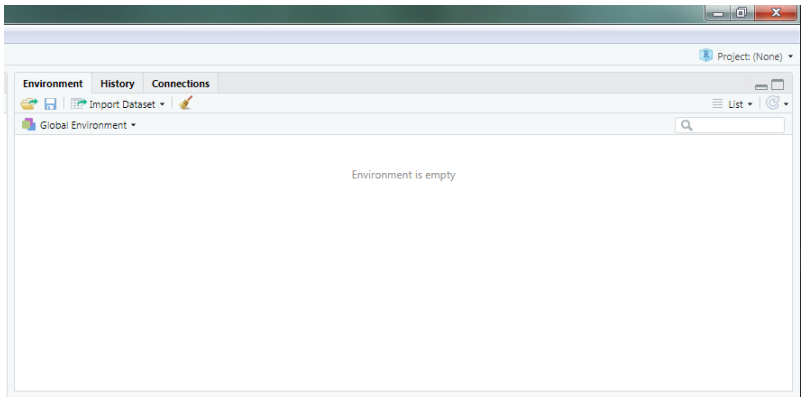
Initial view

Introduction

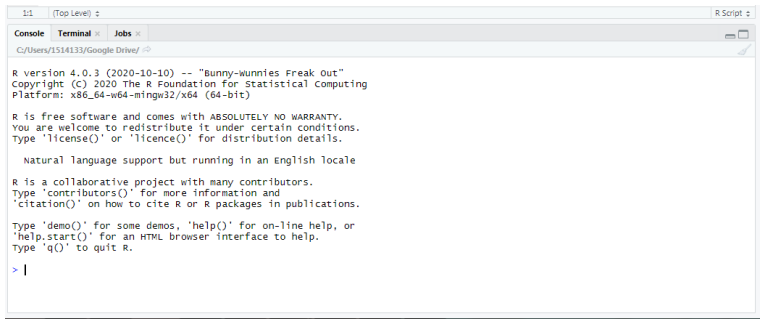
RStudio



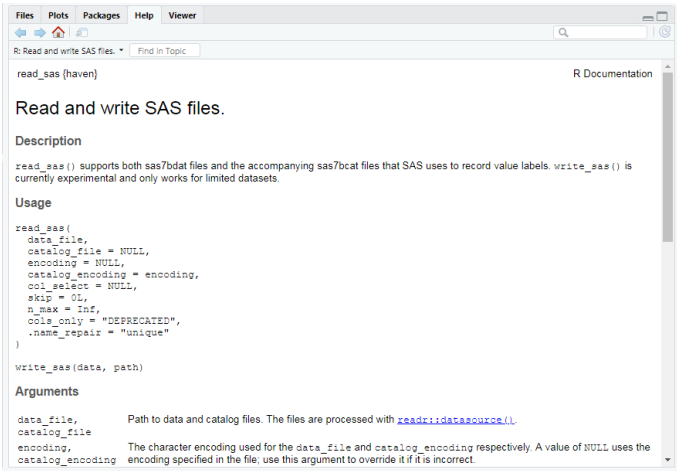
Code editor



Environment



Console



File explorer, Packages, Help, Viewer

Introduction

Scripts

Ideally, we want to be able to easily reproduce past analysis

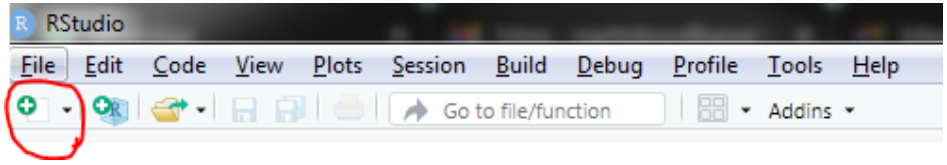
- yields accountability
- allows modifications
- facilitates cooperation

To do so, we will work with **Scripts**

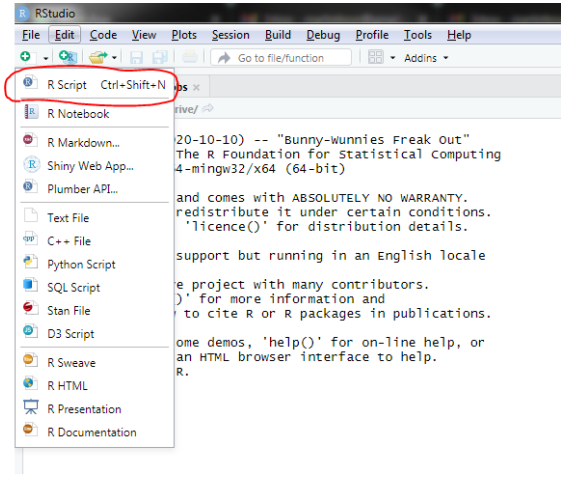
- text file containing a set of commands and comments
- can be used later to re-execute the saved commands
- can be edited to execute a modified version of the commands

Introduction

Scripts



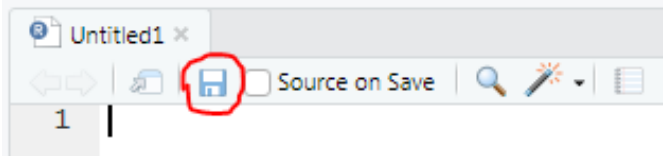
Step 1



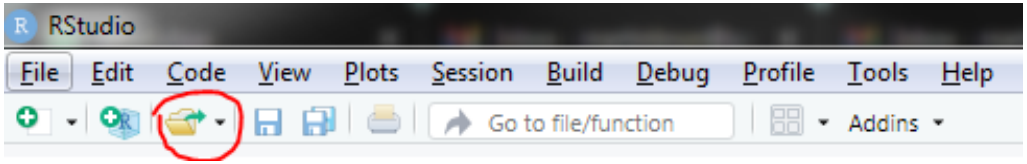
Step 2

Introduction

Scripts



Saving a Script

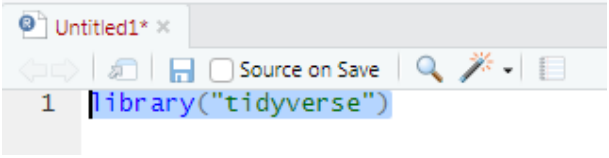


Opening a Script

Introduction

Scripts

Running a command:



Highlight text and press Ctrl+Enter

R is really punctilious

- you must **highlight all letters** in command to run it
- R is **case-sensitive** (i.e. $a \neq A$)

Introduction

Scripts

Comment your scripts!

- **code is re-used** several times
- you may know now **what you intend** to do, but **you will probably forget** that in the future
- all text following a # is understand as a comment
- comments are **not read as code**

```
1 library(tidyverse) #Loads tidyverse library
```

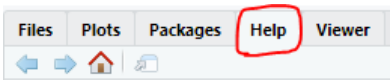
Comment on a script

Introduction

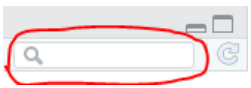
Help

R has built-in help documents

- provide a complete **summary of each function**
- details of **arguments necessary to run** function



Help menu



Search a function

?

or

`help()`

Introduction

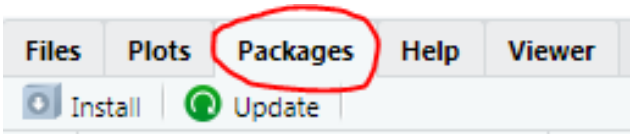
Packages

We will work with functions

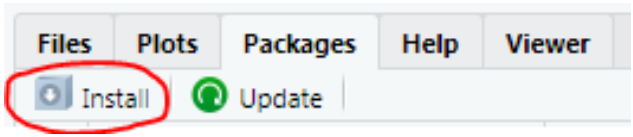
- most are pre-loaded into R
- but, there are lots of useful user-written functions, built-into **packages**

Introduction

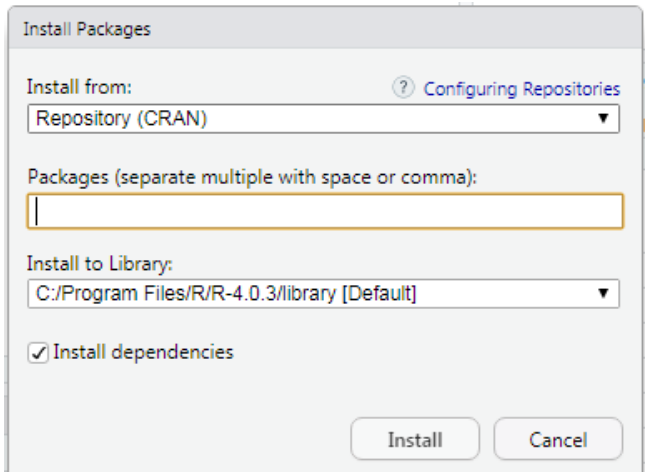
Packages



Packages menu



Installing a Package -Step 1

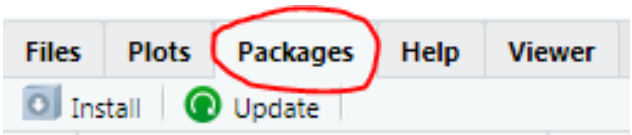


Installing a Package - Step 2

```
install.packages()
```

Introduction

Packages



Packages menu

Loading a library

`library()`

Data Manipulation

Objects:

- Vectors: group of elements, all from the same type
- Array/Matrices: vectors with more than one dimension
- Data Frames: to store datasets
- Lists: contain different objects

Data types:

- Numeric: numbers
- Character: string variables
- Logical: `TRUE` or `FALSE`
- Factor: string variables for categorization

Data Manipulation

Imputing data:

- Vectors: `my_vector1 <- 1`
  1. Name the object (i.e. `my_vector`)
  2. Use an assignment operator (i.e. <-)
  3. Assign a value (i.e. 1)

We can also input vectors:

- Vectors: `my_vector2 <- c(1,2,3)`
  - function `c` is used to specify a combination of values
  - function `seq` is used to specify a sequence of values

Reading data from files

Many times you will want to analyze data already compiled

```
read_excel(path,sheet=NULL,range=NULL)
```

- we need to specify the path for the dataset
- we can select a specific sheet and/or range to import

Basic summary statistics

Some basic summary statistics

- Sum: `sum, rowSums, colSums`
- Mean: `mean, rowMeans, colMeans`
- Standard Deviation: `sd`
- Range: `range`
- Quantiles: `quantile`

Plotting data

- R has a basic built-in package for plotting data
- `ggplot2` is a powerful and commonly used package

```
plot(x,y,type="p")
```

- we need to specify the variable depicted at the x-axis and the y-axis
- we can select a specific type of graph (i.e. "p" for points, "l" for lines)
- we can then add several features for the graph
  - title
  - axis labels
  - axis ranges
  - lines/points aesthetics
  - legend

Bibliography

There are tons of excellent guides for learning R. Feel free to explore for yourself and find the one that suits you.

I include (with link!) a few that have been useful for me:

- An Introduction to R

  Short official introduction to R. Complete, but an arid read.

- R for Data Science

  Manual covering everything from getting data to communicating results.

- Introduction to Econometrics with R

  Empirical companion for 'Introduction to Econometrics' by Stock & Watson (2015). Provides interactive scripts to reproduce results.

- RStudio cheatsheets

  Cheatsheets are concise set of notes for quick reference.

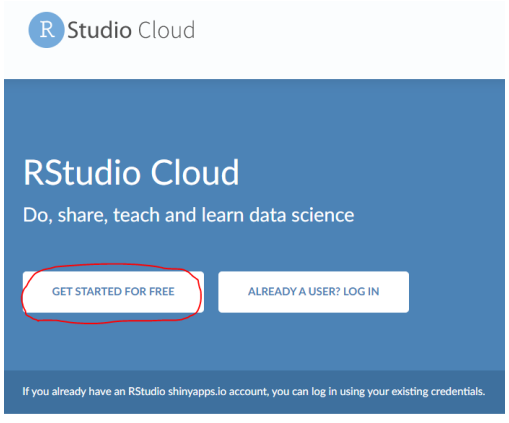
- Datacamp

  Datacamp is a non-free interactive learning plataform. Lessons include exercises with instant correction. Mostly useful for initial learning.

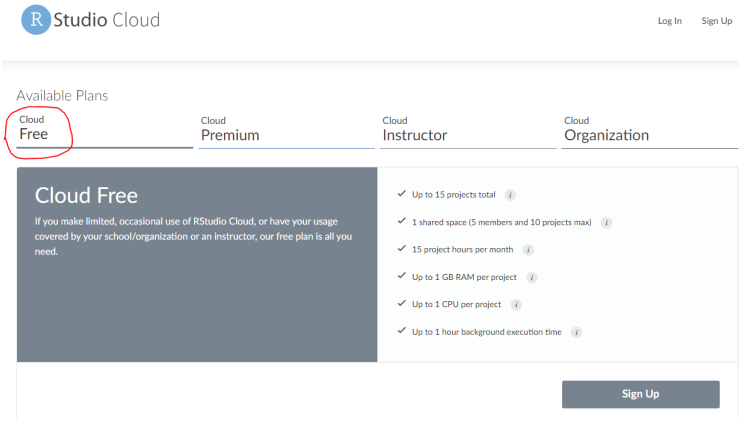
Annex

Introduction

First steps - Online



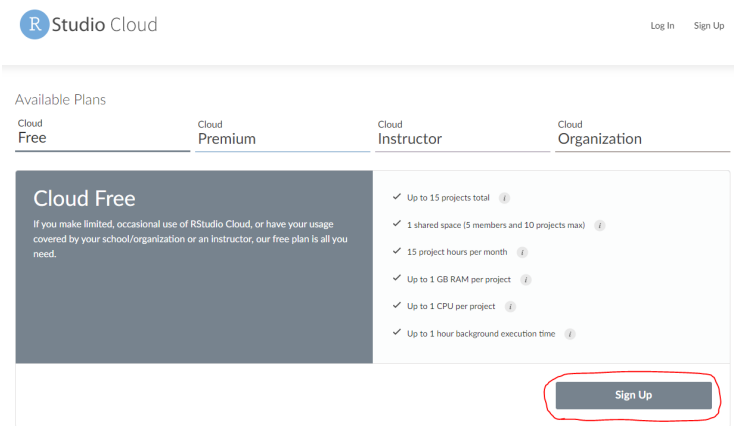
Step 1



Step 2

Introduction

First steps - Online



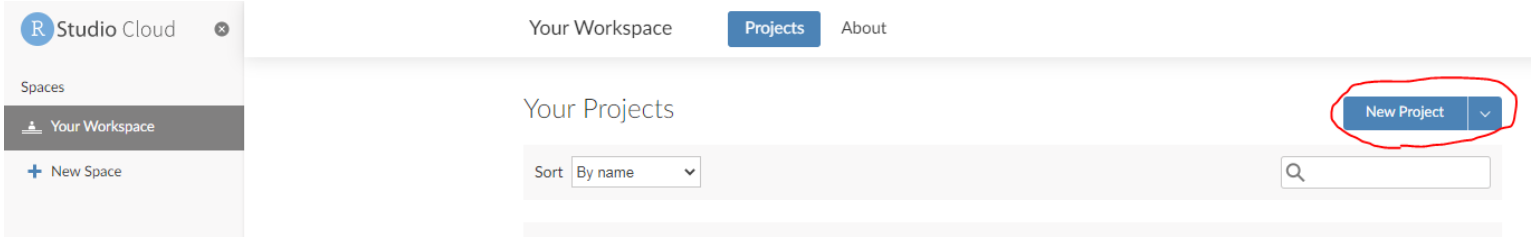
Step 3

You can register as a new user, use your Google user, or use your Github user

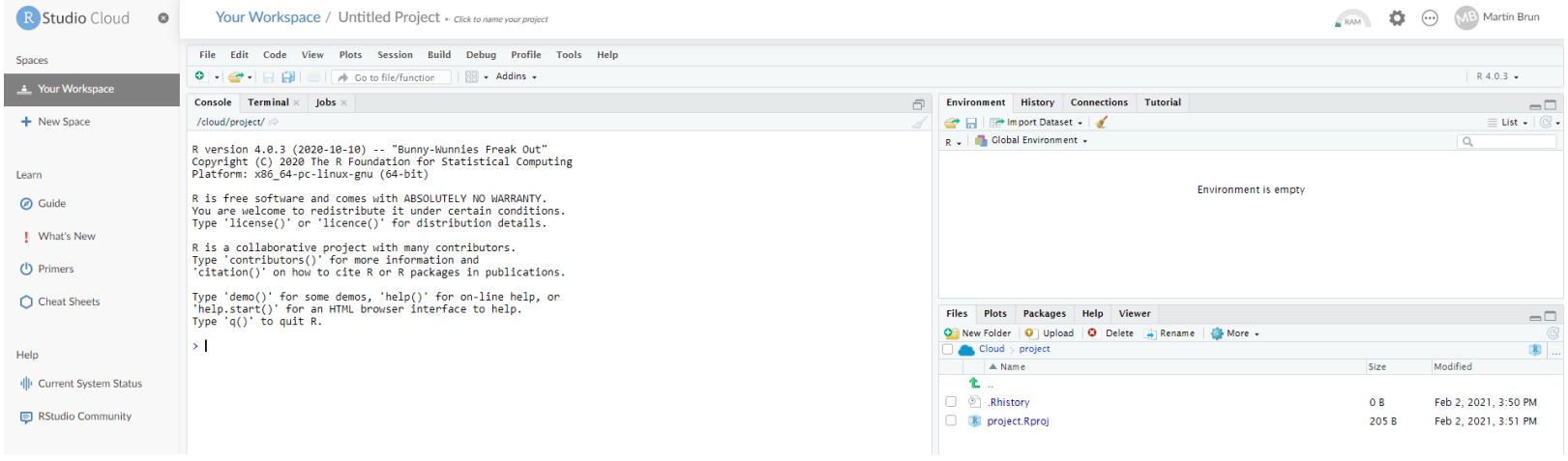
Introduction

First steps - Online

After registration, you enter the platform:



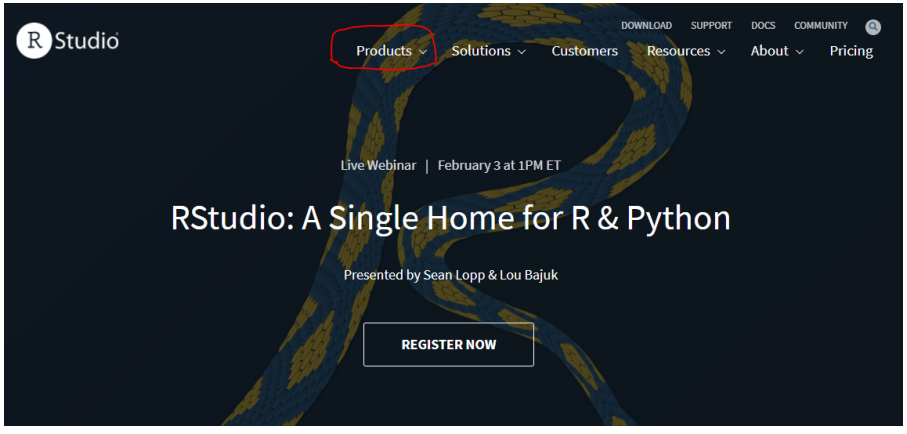
Step 4



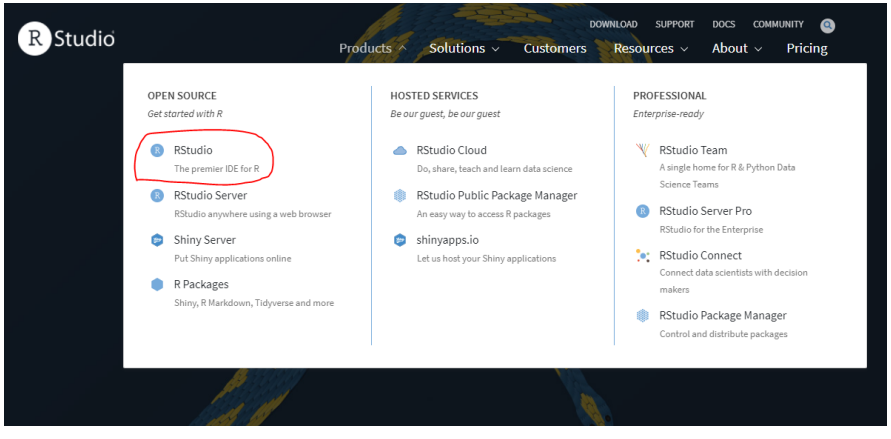
Ready to Start!

Introduction

First steps - Desktop



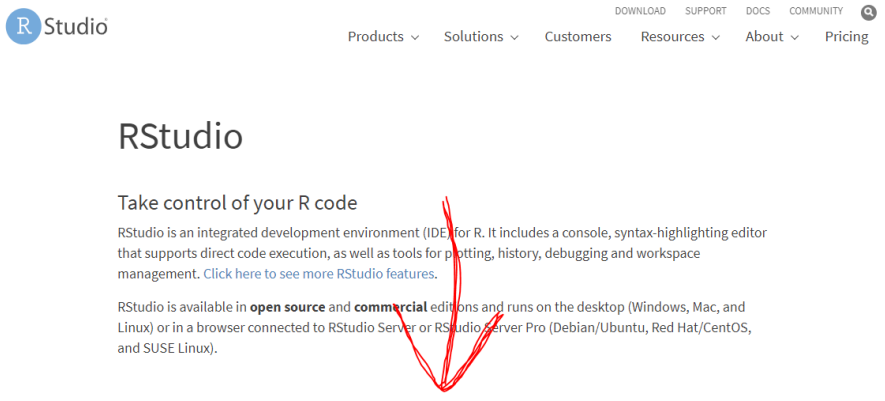
Step 1



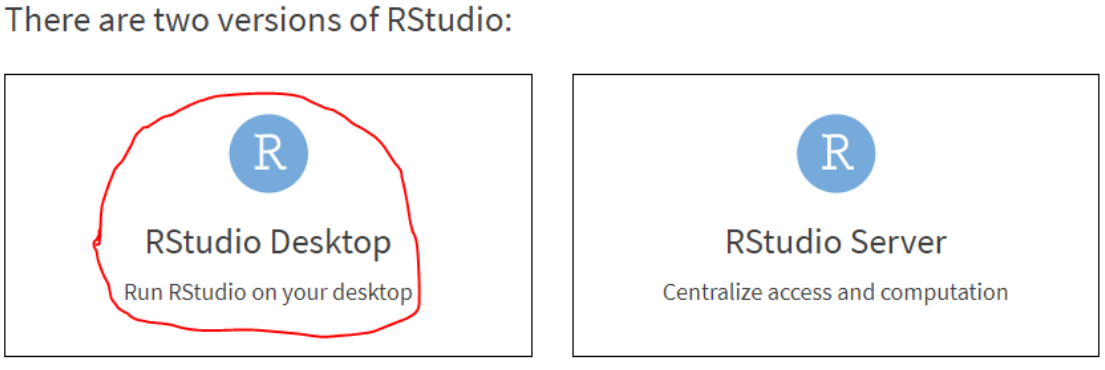
Step 2

Introduction

First steps - Desktop



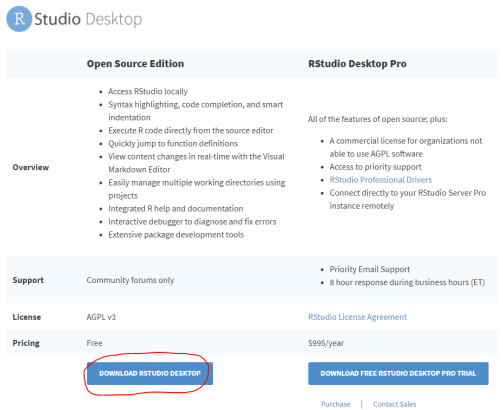
Step 3



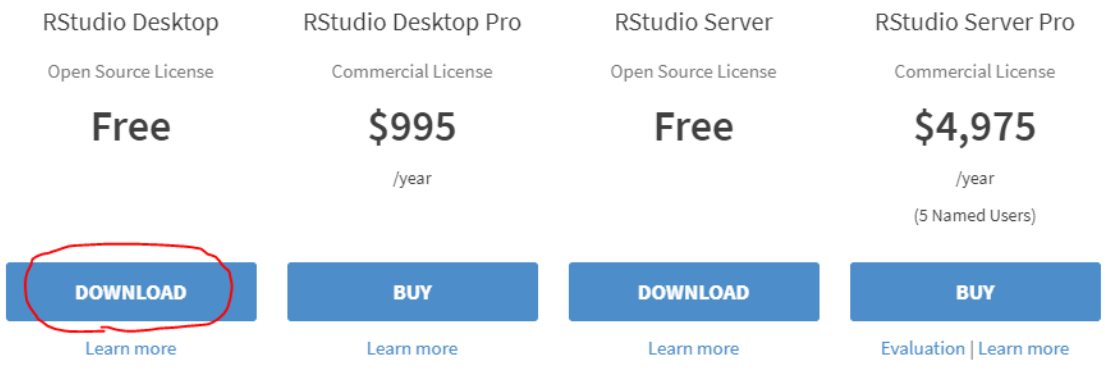
Step 4

Introduction

First steps - Desktop



Step 5

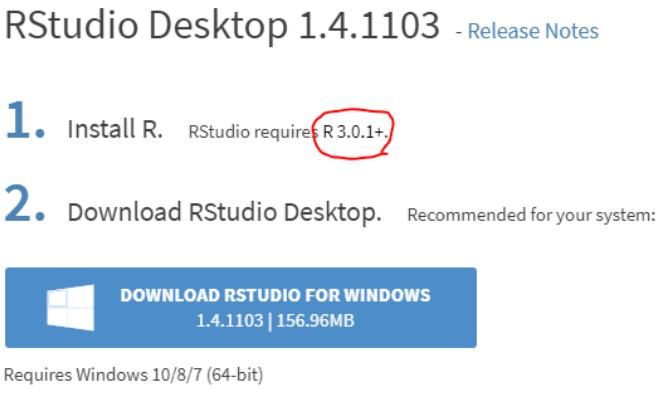


Step 6

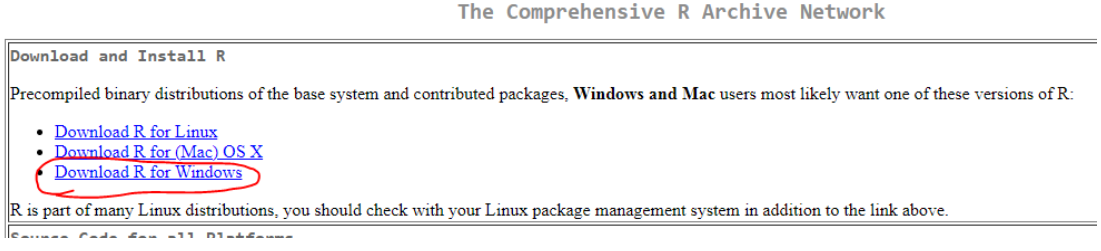
Introduction

First steps - Desktop

First, we will install the R **language**



Step 7



Step 8

Introduction

First steps - Desktop

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to **install R for the first time**.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

Step 9

R-4.0.3 for Windows (32/64 bit)

Download R 4.0.3 for Windows (85 megabytes, 32/64 bit)

Installation and other instructions

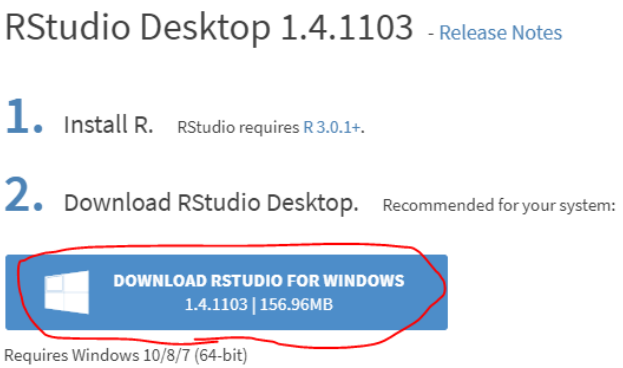
New features in this version

Step 10

Introduction

First steps - Desktop

Second, we will install the R **environment**

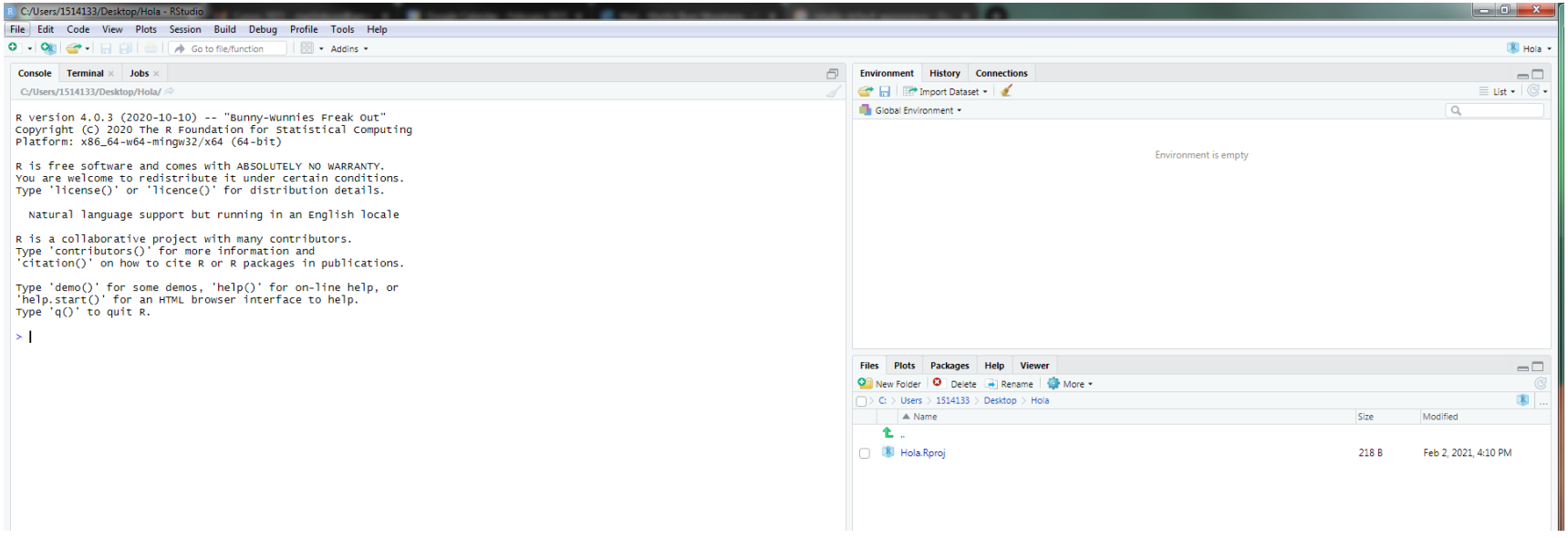


Step 11

Introduction

First steps - Desktop

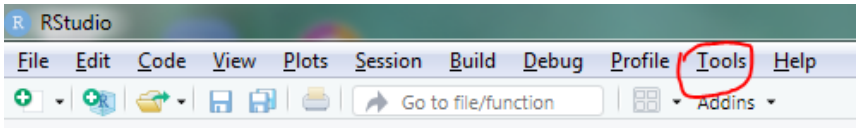
After all installations, open RStudio



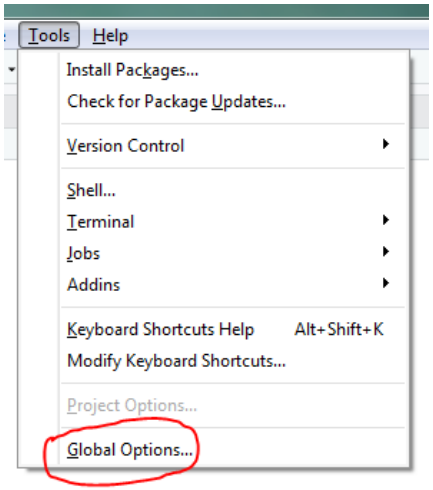
Ready to Start!

Do your eyes a favor!

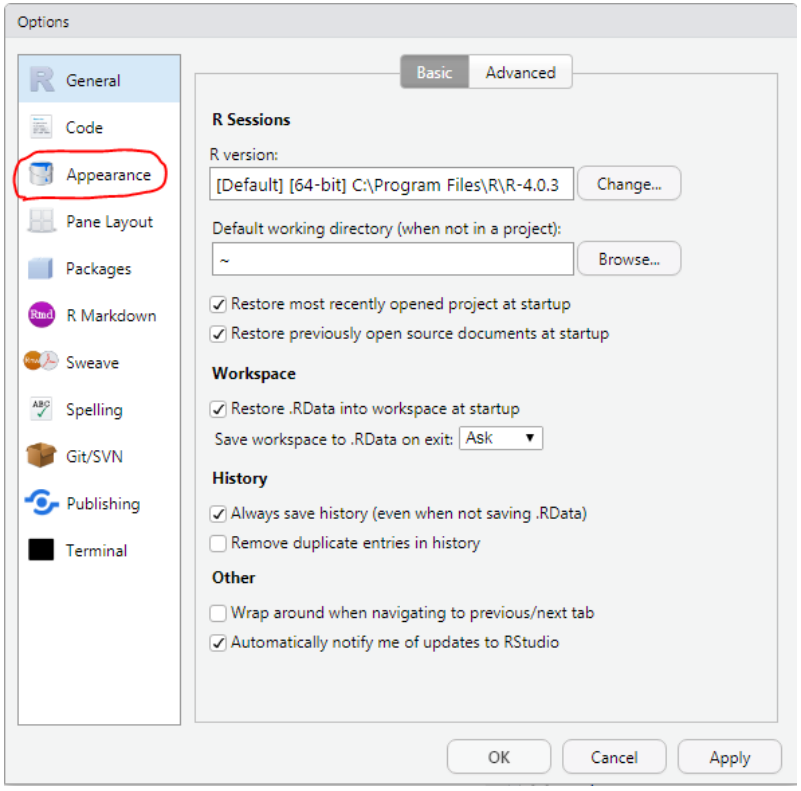
Use a Dark Mode



Step 1



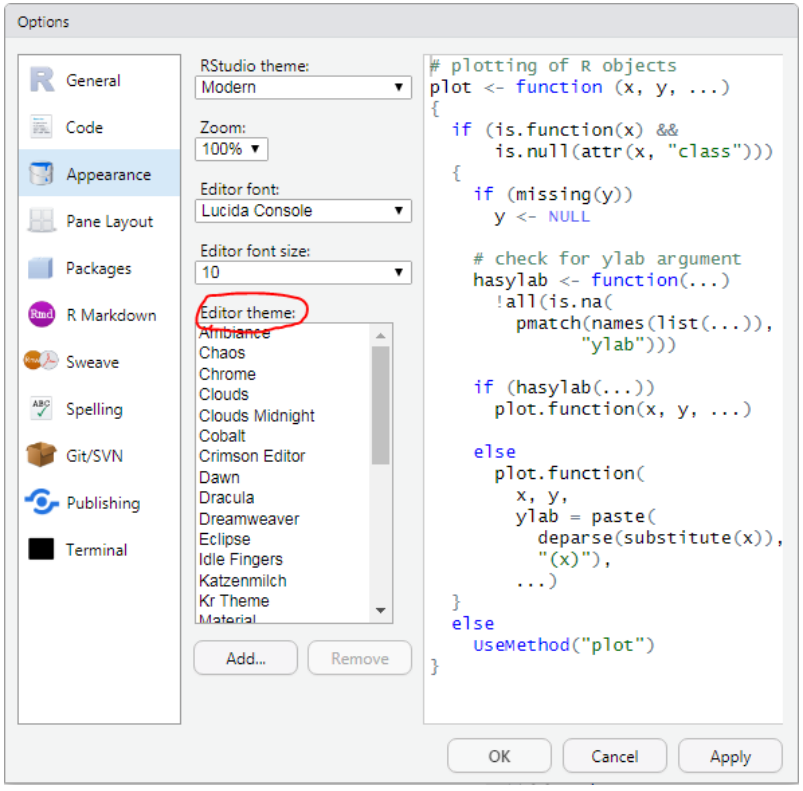
Step 2



Step 3

Introduction

First steps - Online



You can choose among many dark mode themes.

I recommend: "Idle Fingers" ":)