Software guidance

Throughout the course, we will be performing small exercises to apply various text analysis techniques for automated classification. Though there are no prerequisites to registering on this module, a minimal experience with R or Python is strongly recommended, and some basic statistical knowledge would be advantageous.

1. R and RStudio

Before the first session of the module, it is recommended that students have R and RStudio installed on their laptops, as the software will be used for demonstrations and exercises.

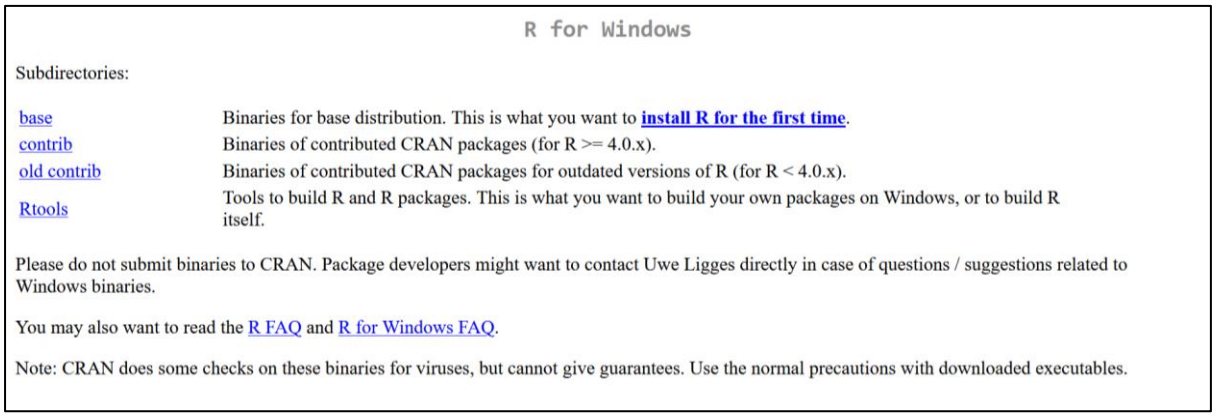
R is the statistical software environment that is used to manipulate and work with data. The R interface is not terribly user friendly, so we use RStudio as a viewer on top of R. What this means is that, in practice, once you have R downloaded onto your machine you will not need to open R directly. Instead, all your scripts and analysis will be conducted via RStudio.

Installing R

R is and open-source software, so it is free to install and use. You can download R for Windows, Mac or Linux by going to the following site:

http://cran.r-project.org/

Select the option appropriate for your operating system, and then choose the "base" subdirectory (see below an example for Windows):



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 >= 4.0.x).
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 4.0.x).
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.

Once the download is complete, you can double-click on the file to install R.

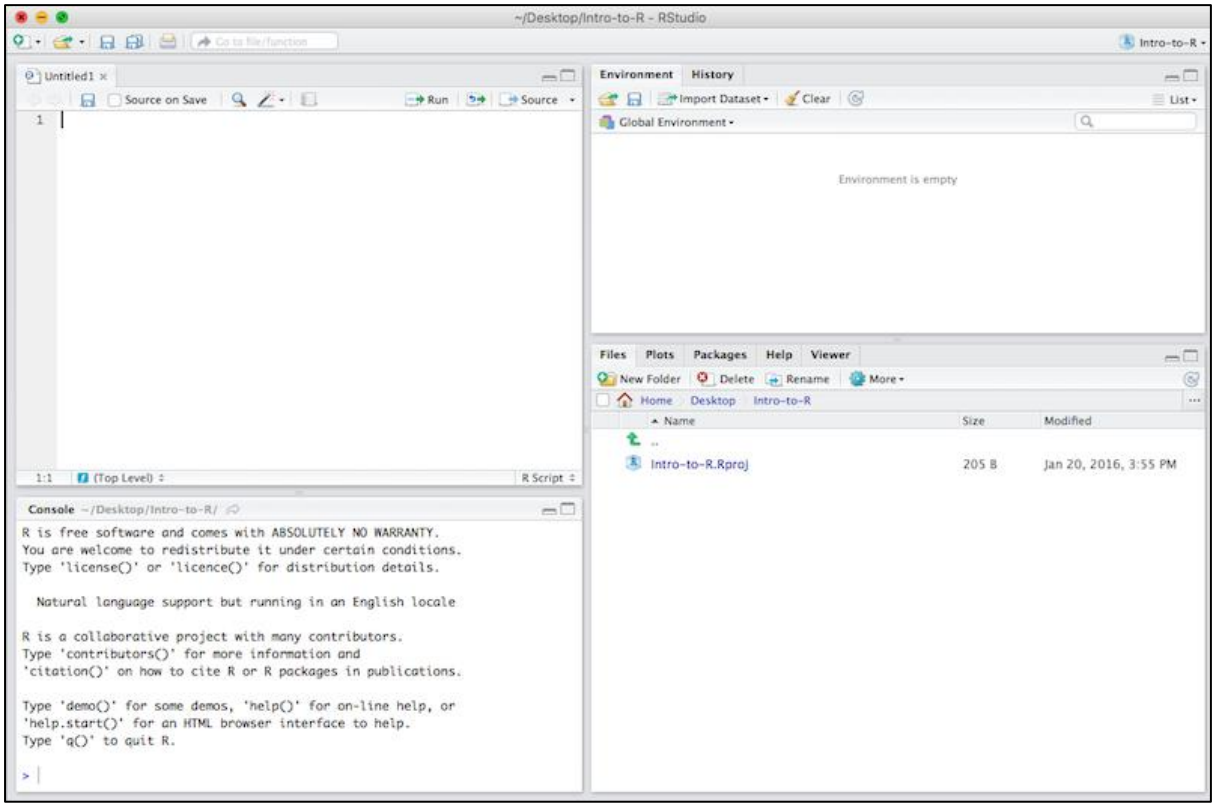
Installing RStudio

To make R a bit more manageable, we will also use RStudio in this course. You can download RStudio by visiting the following site:

http://www.rstudio.com/products/rstudio/download/

Click the installer that corresponds to your operating system to download RStudio Desktop. Wait for the installer to download, then click on it to proceed with the installation, clicking "Next"/"Keep" and "Install" along the way if prompted. Wait for the installer to finish.

Once installed, you can double-click on the RStudio icon to start the programme and obtain an interface that should look more or less like the following:



In case you encounter any issue in installing either R or RStudio, there are numerous troubleshooting resources online that can support you for the installation.

For example, you could have a look at these videos on YouTube for Windows or MacOS.

2. Google Colab

In the last session, we will be using Python coding language. For the purpose of this course it is not necessary to have Python installed on your laptops, as we will be using Google Colab, a free, cloud-based platform that allows users to write and execute Python code.

To use Google Colab, you need a Google account. Please set up one if haven't any. If you wish not to register a Google account, you can still use the scripts presented in Python and run them on your local machine, though most models we are going to cover require substantial CPU/GPU capacity.