Dominik Jung

The Modern Business Data Analyst

A Case Study Introduction into Business Data Analytics with CRISP-DM and R

BONUS CHAPTER

Springer

8 Appendix

8.1 Installing R and RStudio

If you are not familiar with installing and setting up software like R and RStudio you can use the following tutorial. In the tutorial I describe how to install R and RStudio software on your local machine and highlight common pitfalls during the installation process.

8.1.1 Setup Base R

RStudio is an integrated development environment (IDE) for the statistical computing language R. The IDE is free to use and can be installed as desktop application. Furthermore, there exists a browser-accessible version for servers or the cloud. The software is provided by the RStudio (now Posit) company, which has no connection to the R foundation. However, it is popular for pushing forward the non-profit and commercial usage of R.

Let us now install R and then RStudio on your local computer:

Go to the R project website on: *www.cran.r-project.org* and click on the “Download R” button for the system of your choice.

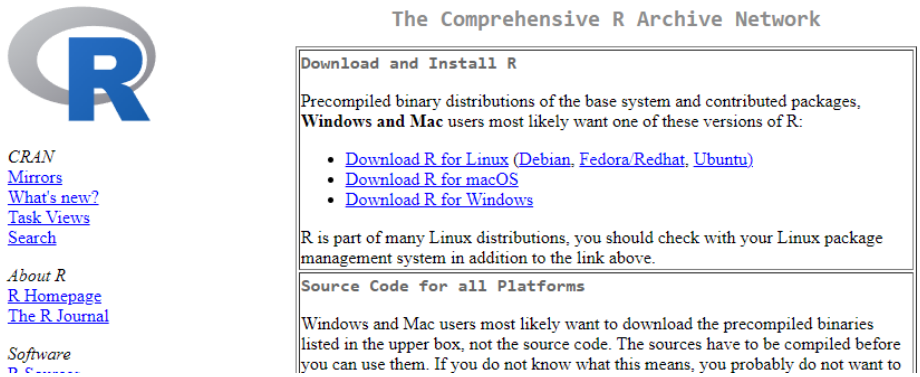

Figure 8-1 The official R website looks a bit outdated but has everything you need

After downloading, please start the installer on your local machine. R is available as free to use under the terms of the free software foundation's GNU General Public License. Accept the license conditions by pressing “Next >”.

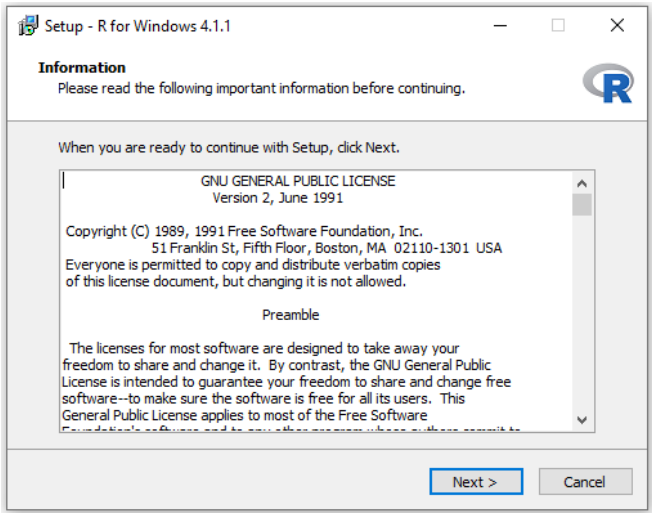

Figure 8-2 Accept the terms to use R

Then you can choose a folder where your current version of R should be installed. Best practice is to install them in “Program Files\R\R-Version”. You can download other versions of R and install them in this folder. You can then switch easily between the different versions.

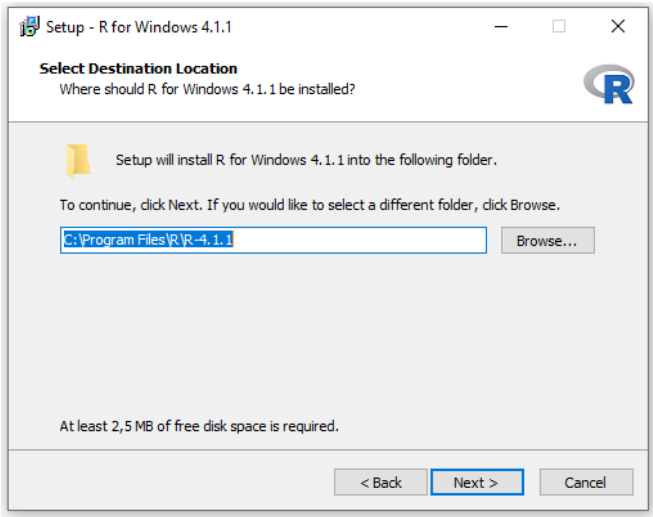

Figure 8-3 I recommend to select the default location for installation

In the next step, you have to specify which components of R should be installed. Install all existing components to use the full capabilities of R.

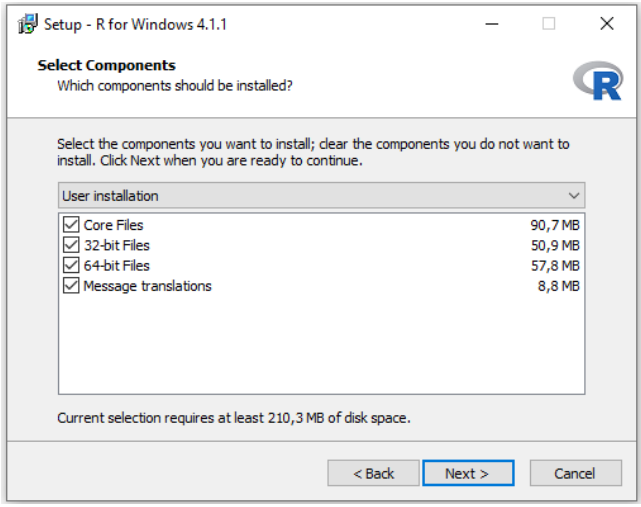

Figure 8-4 Install the specific R version for your system (32 or 64 bit)

In the next step, you are asked if you want to customize the system startup. On modern windows and iOS machines this is not necessary, hence I recommend that you do not customize the R startup options to use the default configuration.

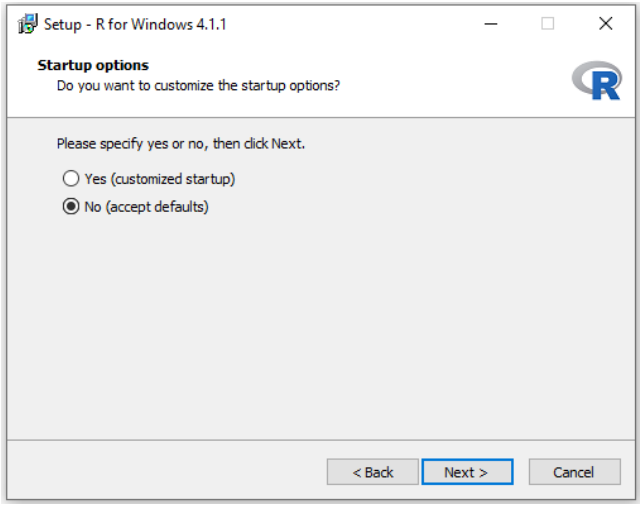

Figure 8-5 In general you do not have to specify the startup options, hence select "No" to accept the defaults

If this is your first R installation you can associate R with `.RData` files on your computer. Otherwise deselect the option. I also recommend not to save the version number in registry if you plan to install multiple versions of R for maintenance reasons on your system. Many programs set the version of R by examining environment variables like `R_HOME`, `R_MIKTEX`, `R_TOOLS` or if it is not set, they look in the registry. Which can result in problems if you have another R version running.

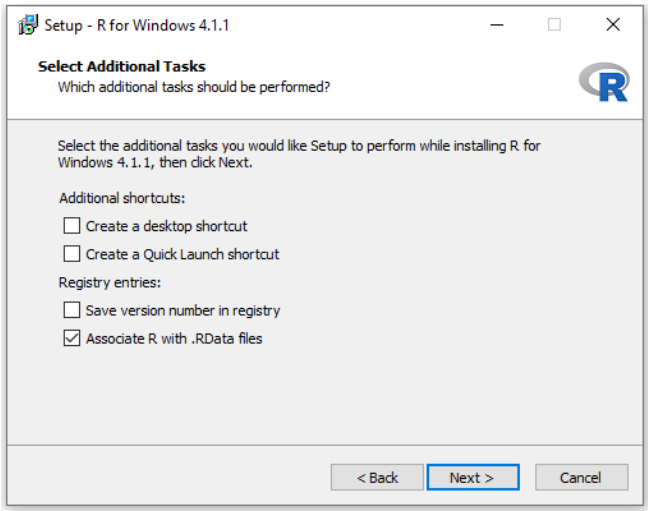

Figure 8-6 Do not forget to click the checkbox to associate R with `.RData` files if not already done for another R version

Congratulation you installed R! You can now run and start R scripts on your computer. However, we will continue and install some more programs to setup a professional R development environment for you.

8.1.2 Setup RStudio

In the next step, we are going to install a professional integrated development environment for R – the RStudio Code IDE. The RStudio IDE is the de-facto standard IDE for R development. It's designed for business data analytics and statistical research. And it is so well designed and supports analysts so excellently that it has inspired several other IDEs for other analytics-related languages and has often been copied. For example, the well-known Python IDE Spyder offers an "RStudio mode", or the Rodeo IDE of the AI company "yhat" was initial developed because there was no usable analytics IDE like RStudio for Python. For a few years, RStudio also supports Python and many other languages in addition to R, so that it is increasingly used outside of R. To emphasize that the company offers solutions for other relevant data science tools and programming languages besides R, the company changed its name from RStudio to Posit in 2022/2023. However, the editor kept its name during the transformation.

Today, there exist different versions of the IDE you can use if you want to program in R. There is a portable version, that can be used without installation which you will probably use when you have no admin rights on your local machine. The server version can be installed in the cloud and is often used if you run your R code on computing clusters processing big data or building data- or computing-intensive models. In this book we will install the default version, the local RStudio Desktop. For that purpose, go to the RStudio website and download the "RStudio Desktop (free)": *www.rstudio.com/products/rstudio*

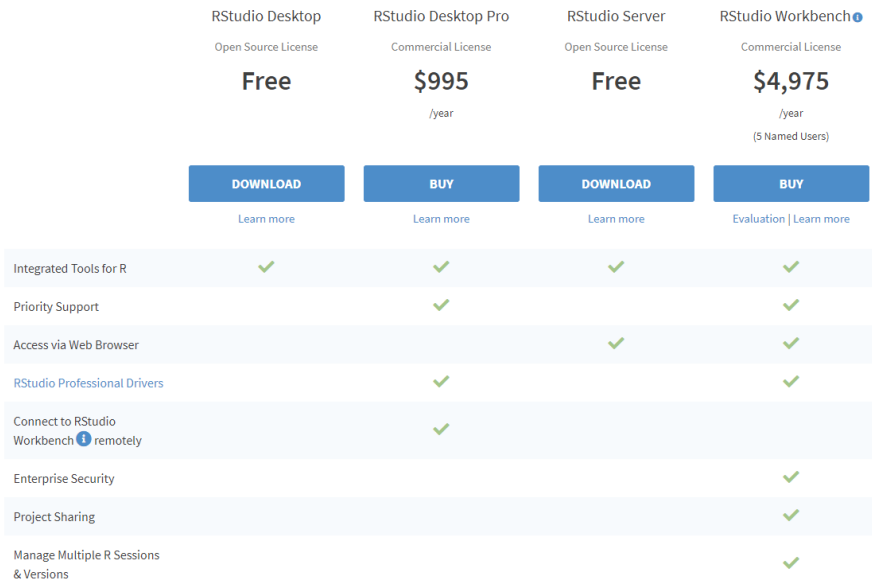

Figure 8-7 Install the free version of RStudio Desktop

If you choose "RStudio Desktop" you can choose between the portable version and the default version that needs admin rights. In this book we download the installer and run it.

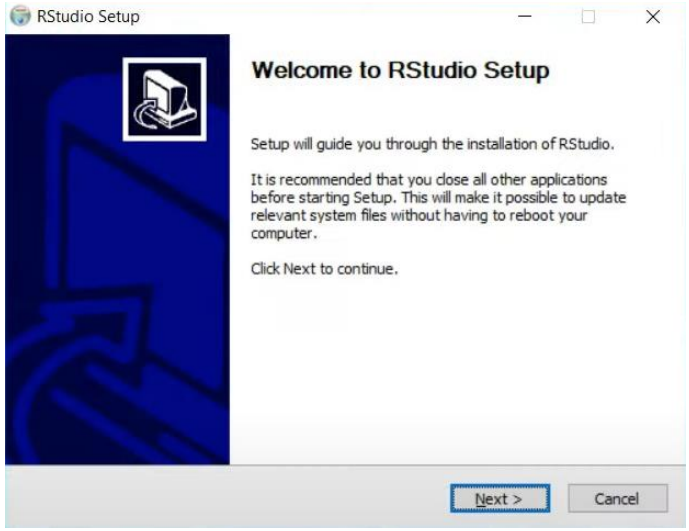

Figure 8-8 The following wizard will guide you through the installation process of RStudio

If the installer opens, press "Next >" and choose the path of where you want to install your RStudio IDE.

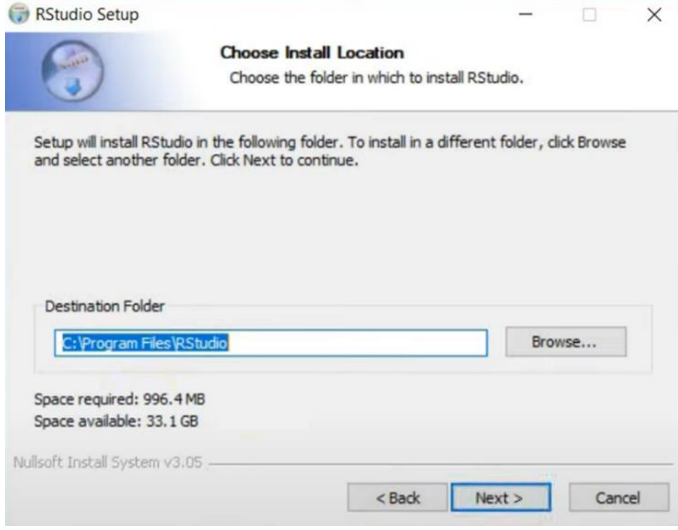

Figure 8-9 I recommend that you install RStudio in the default location

After you have successfully configurated your R installation you can finish the process by pressing "Next >" in the last dialog of the wizard. This will start the installation and the following window will appear:

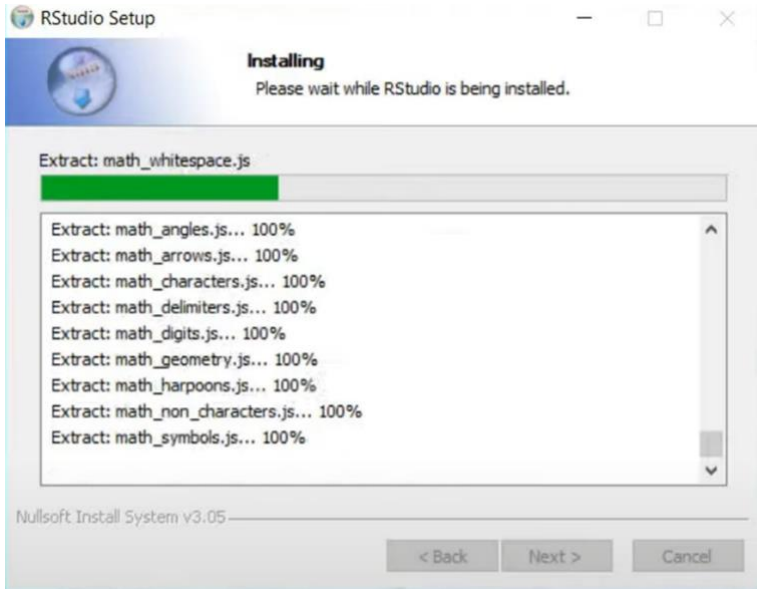

Figure 8-10 Please wait until the installation is finished.

After your installation finished successfully you can open the RStudio IDE to use R in a more modern way. It will look like in Figure 8-11.

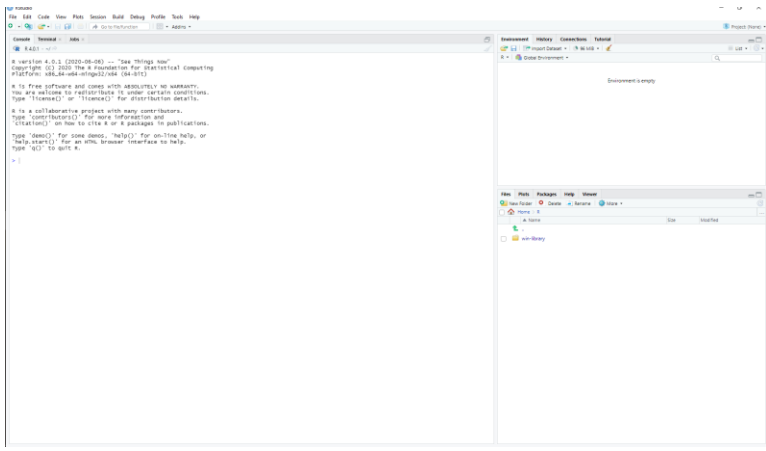

Figure 8-11 Your fully setup RStudio with console, workspace and project files

	I further recommend that you install RTools on your local machine. RTools is a bundle of programming tools you will need if you want to compile R packages. However, it is not always necessary if you do not build R packages with C/C++/Fortran code from source. If you have not Rtools on your machine your R version uses precompiled packages from CRAN. You can find it online at: *www.cran.r-project.org/bin/windows/Rtools/* It might be that your IDE does not find your R version if it has been installed at unknown folders. Then you have to setup your R-Version in the menu. Just click on the “Tools” section in the menu, then click on “Global Options” which allows you to specify your R-Version in the “General” section.