Learn RStudio IDE

Quick, Effective, and Productive Data Science

Matthew Campbell

Apress®

Learn RStudio IDE

Quick, Effective, and Productive Data Science

Matthew Campbell

Apress®

Learn RStudio IDE: Quick, Effective, and Productive Data Science

Matthew Campbell
Yardley, PA, USA

ISBN-13 (pbk): 978-1-4842-4510-1 ISBN-13 (electronic): 978-1-4842-4511-8
https://doi.org/10.1007/978-1-4842-4511-8

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Steve Anglin
Development Editor: Matthew Moodie
Coordinating Editor: Mark Powers

Cover designed by eStudioCalamar

Cover image designed by Freepik (www.freepik.com)

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail editorial@apress.com; for reprint, paperback, or audio rights, please email bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/9781484245101. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

Table of Contents

About the Author

Matthew Campbell is a research data analyst who has worked on data problems in educational psychology, psychometrics and other research areas for over 15 years. Matt is passionate about technology which he uses to create stories with data, find insights that inform analysis and solves problems for businesses. He received his Masters in Management of Information Systems from the University of Phoenix and Bachelors in Psychology from Pennsylvania State University. Matt has authored 4 books on mobile app development, lead coding bootcamps and worked on various large-scale research projects.

About the Technical Reviewer

Dr. Prachee Chaturvedi is Data Scientist and Digital Strategy Lead in Regulatory, R&D, in Bayer and was a member of Monsanto's Leadership Development Program. She received her PhD in Biological Engineering from University of Florida, Gainesville, FL, and Bachelors in Mechanical Engineering from HBTI, Kanpur, India. Currently, Prachee is focused on establishing comprehensive digital strategy for Regulatory Sciences and developing digital tools for automated and on-demand analysis for Regulatory Safety Data Submission at Bayer Crop Science R&D. She has authored 15 scientific publications, 2 book chapters, and led various cross functional scientific projects.

CHAPTER 1

Installing RStudio

RStudio is an integrated development environment (IDE) that adds modern features like syntax highlighting and code refactoring to R. The strength of RStudio is that it brings all the features that you need together in one place. Before we install RStudio, we will need to install the latest version of R for our operating system as well as another program called git. Both of these software packages are integrated into RStudio, but they do not come with the RStudio IDE.

R is a statistical programming language and we need this language to use RStudio. Git is a popular version control system that will be integrated into RStudio. Version control systems like git are used to manage copies of the code you are working on. Git helps you see the differences between versions of a file so that you can get insight into why one version of your code behaves differently than another version. Git is also required to use Github, an online community where programmers share code and projects.

Install R

R is a free open source software used for statistical programming and graphics. You can get the latest version of R from CRAN, the Comprehensive R Network. This is a network of mirrored servers that have copies of R as well as R packages. R packages are extensions to the core R programming language contributed by R users in the community.

M. Campbell, *Learn RStudio IDE*, https://doi.org/10.1007/978-1-4842-4511-8_1

To install R, go to this URL: `https://cran.r-project.org/mirrors.html` and you will see a list like the one featured in Figure 1-1.

CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: main page, windows release, windows old release.

If you want to host a new mirror at your institution, please have a look at the CRAN Mirror HOWTO.

0-Cloud	
https://cloud.r-project.org/	Automatic redirection to servers worldwide, currently sponsored by Rstudio
http://cloud.r-project.org/	Automatic redirection to servers worldwide, currently sponsored by Rstudio
Algeria	
https://cran.usthb.dz/	University of Science and Technology Houari Boumediene
http://cran.usthb.dz/	University of Science and Technology Houari Boumediene
Argentina	
http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
Australia	
https://cran.csiro.au/	CSIRO
http://cran.csiro.au/	CSIRO
https://mirror.aarnet.edu.au/pub/CRAN/	AARNET
https://cran.ms.unimelb.edu.au/	School of Mathematics and Statistics, University of Melbourne
https://cran.curtin.edu.au/	Curtin University of Technology
Austria	
https://cran.wu.ac.at/	Wirtschaftsuniversität Wien
http://cran.wu.ac.at/	Wirtschaftsuniversität Wien

Figure 1-1. *CRAN mirrors*

Scroll down the list of servers and find the one that is geographically closest to you and then click the link. Don't worry too much about getting the absolute closest server, just picking any from your country should be fine. If you find a broken link, simply go back and try another server. You will be brought to a new screen that lists versions of R for Windows and Mac as you can see in Figure 1-2.

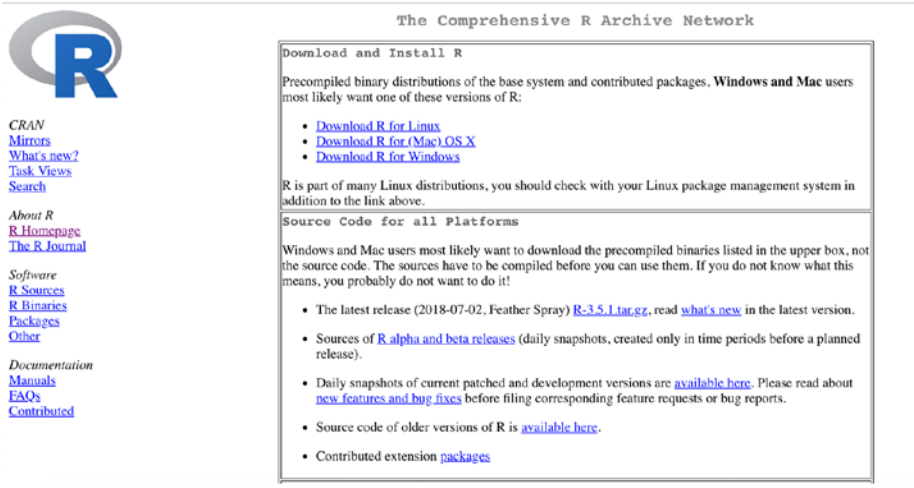

Figure 1-2. *Installing R on CRAN*

Click the link for your operating system. Each operating system has different instructions to set up R.

Installing R on Mac

If you are installing R on Mac, you will be presented with a lot of detail about the different R releases. Most people will simply go to the area toward the middle of the page under the heading Latest release as shown in Figure 1-3: and then click the file that ends in with the .pkg file extension. The exact name of the file will be different based on the most current release of the installer.

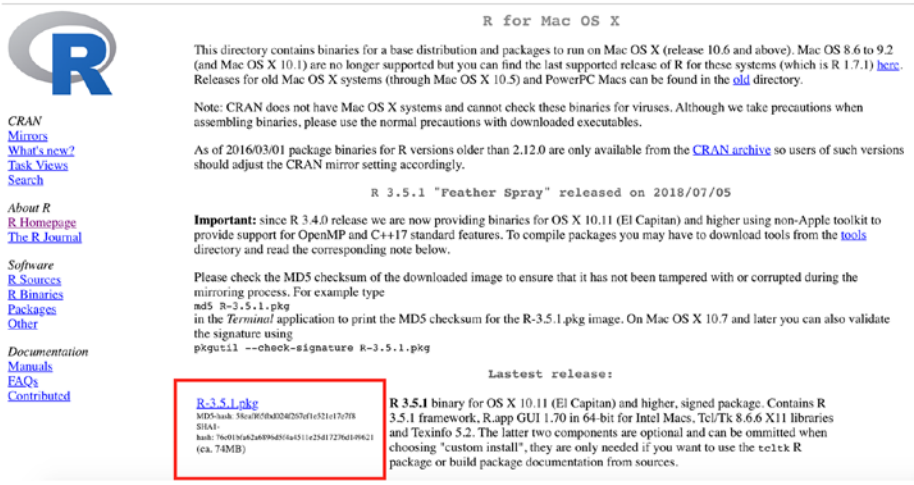

Figure 1-3. *Mac R installer*

Click the pkg link to install the most recent version of R. Note that if you have a special need for an older version of R, you can scroll down to see options on installing previous version of R. The pkg file will appear in your Downloads folder. Go to your Downloads folder and click the pkg file to install R. A screen like Figure 1-4 will appear.

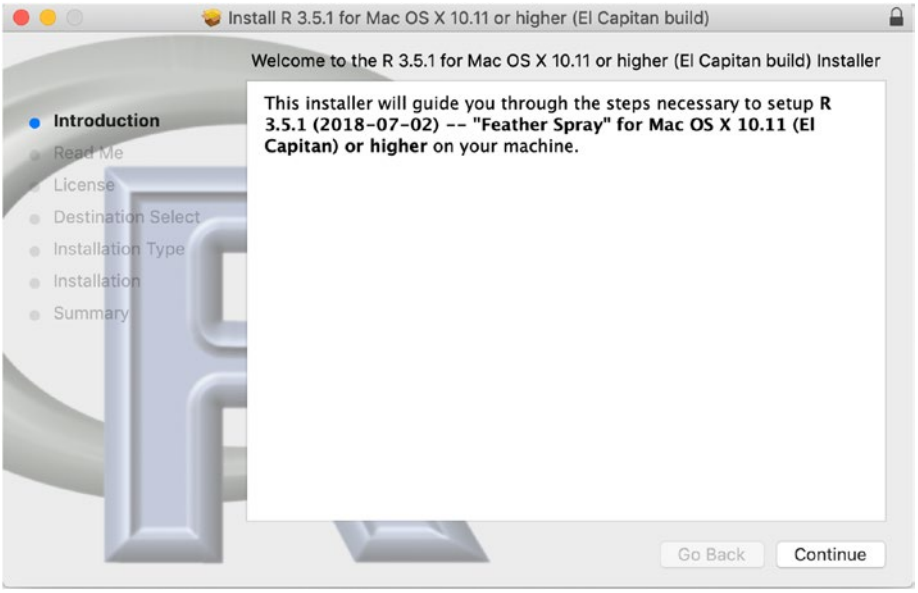

Figure 1-4. *Installing R on Mac*

You will get a series of screens like the one in Figure 1-4. Click Continue for each screen. At one point, a model dialog box will pop up asking for you to agree to the software license. Click Agree to move on. Next, you will be presented with an option to install to a specific disk as shown in Figure 1-5.

Figure 1-5. *Choosing R location*

Most people will simply choose the option to install for all users. Choose the first option and click the Continue button. At the next screen, click the Install button. You will be prompted for your Mac password since the R installer will make changes to your Mac. Type that in to continue. R will install, and when everything is done you should see a screen like in Figure 1-6.

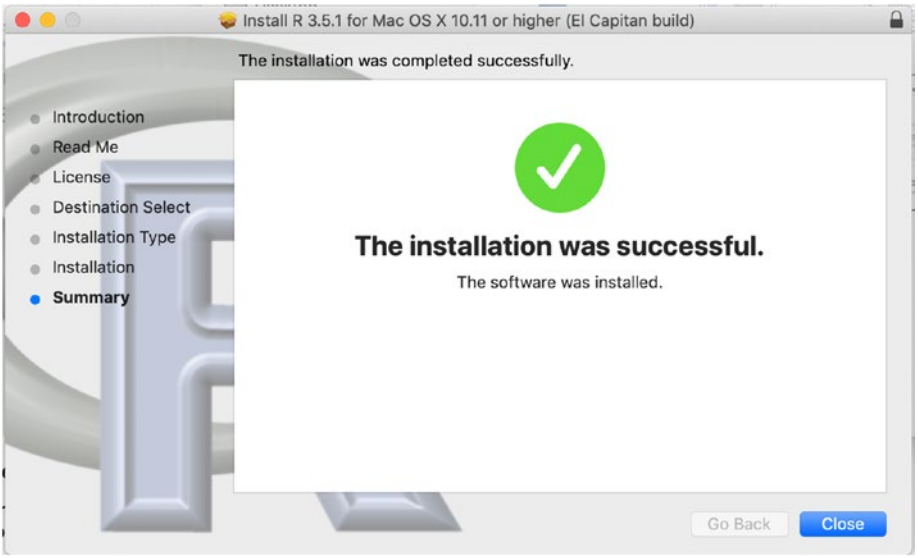

Figure 1-6. *Successful R installation*

Click the Close button and R will be installed on your Mac. You will find R in your Applications folder. Double click the R.app program in your Applications folder to verify that you have R installed and working. The native R Console screen will appear. You can play around with the system here if you already know a little R, or you can simply exit the program. We will revisit using R once RStudio is installed.

Installing R on Windows

Once you click the R for Windows link, you will see a page of instructions. Click on "install R for the first time" as shown in Figure 1-7.