



# 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>

Copyright © 2019 by Matthew Campbell

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

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](http://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](mailto:orders-ny@springer-sbm.com), or visit [www.springeronline.com](http://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](mailto:editorial@apress.com); for reprint, paperback, or audio rights, please email [bookpermissions@springernature.com](mailto: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](http://www.apress.com/9781484245101). For more detailed information, please visit <http://www.apress.com/source-code>.

Printed on acid-free paper

# Table of Contents

<b>About the Author .....</b>	<b>vii</b>
<b>About the Technical Reviewer .....</b>	<b>ix</b>
<b>Chapter 1: Installing RStudio.....</b>	<b>1</b>
Install R .....	1
Installing R on Mac .....	3
Installing R on Windows .....	7
Installing R on Linux .....	8
Install Git .....	9
Install RStudio .....	11
<b>Chapter 2: Hello World.....</b>	<b>15</b>
Interactive R Console .....	15
Vectors.....	16
Hello World .....	16
Terminal.....	18
Environment .....	19
Importing Data.....	21
Datasets .....	22
History .....	26
Connections.....	27
Conclusion .....	27

TABLE OF CONTENTS

- Chapter 3: RStudio Views ..... 29**
  - Files, Plots, Packages, Help, and Viewer Pane .....29
  - Files .....29
  - Plots.....30
  - Packages .....34
  - Help .....37
  - Viewer .....38
  - Conclusion .....38
  
- Chapter 4: RStudio Projects.....39**
  - Create a New RStudio Project.....39
  - Project Types .....41
  - RStudio Project Tour .....43
  - Conclusion .....48
  
- Chapter 5: Repeatable Analysis.....49**
  - Organizing Datasets.....49
    - Download Life Expectancy Data .....50
    - Add Dataset to R Project.....51
  - R Code Files .....54
  - Data Exploration.....57
    - Summarizing Dataframes .....57
  - Conclusion .....61
  
- Chapter 6: Essential R Packages: Tidyverse.....63**
  - R Packages .....63
  - Tidyverse.....64
  - magrittr .....64

tibble .....	66
dplyr .....	67
SQL Like Joins .....	69
stringr .....	71
Conclusion .....	72
<b>Chapter 7: Data Visualization .....</b>	<b>73</b>
ggplot2 .....	73
htmlwidgets .....	79
r2d3 .....	82
Conclusion .....	85
<b>Chapter 8: R Markdown .....</b>	<b>87</b>
R Markdown Documents .....	87
R Markdown Example .....	92
R Markdown Reproducibility .....	94
Conclusion .....	97
<b>Chapter 9: Shiny R Dashboards .....</b>	<b>99</b>
New Shiny Apps .....	99
Understanding Shiny .....	102
User Interface .....	102
Server .....	106
App .....	107
Customizing Shiny Apps .....	108
Conclusion .....	112

TABLE OF CONTENTS

- Chapter 10: Custom R Packages..... 113**
  - Custom R Packages Use Cases..... 113
  - Create New Custom R Package ..... 114
    - hello.R..... 115
    - DESCRIPTION ..... 116
    - hello.Rd..... 117
  - Build R Package ..... 117
  - R Package Documentation ..... 118
  - Adding Datasets to R Packages..... 122
  - Code Completion ..... 124
  - Conclusion ..... 124
- Chapter 11: Code Tools ..... 125**
  - Source Control Integration ..... 125
    - Setup Git..... 126
  - Debugger ..... 133
  - More Code Tools..... 135
  - Conclusion ..... 136
- Chapter 12: R Programming ..... 137**
  - Objects..... 137
    - Essential Class Types..... 138
    - Flow Control ..... 142
  - Functions..... 146
  - Importing JSON Data..... 147
  - Conclusion ..... 149
- Index..... 151**

# 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.

## CHAPTER 1 INSTALLING RSTUDIO

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



CRAN  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

*About R*  
[R Homepage](#)  
[The R Journal](#)

*Software*  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

*Documentation*  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

The Comprehensive R Archive Network

#### Download and Install R

Precompiled binary distributions of the base system and contributed packages. **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\)OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

#### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2018-07-02, Feather Spray) [R-3.5.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

**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.



R for Mac OS X

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) [here](#). Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old](#) directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

As of 2016/03/01 package binaries for R versions older than 2.12.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting accordingly.

R 3.5.1 "Feather Spray" released on 2018/07/05

**Important:** since R 3.4.0 release we are now providing binaries for OS X 10.11 (El Capitan) and higher using non-Apple toolkit to provide support for OpenMP and C++17 standard features. To compile packages you may have to download tools from the [tools](#) directory and read the corresponding note below.

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type  
`md5 R-3.5.1.pkg`  
 in the *Terminal* application to print the MD5 checksum for the R-3.5.1.pkg image. On Mac OS X 10.7 and later you can also validate the signature using  
`pkgutil --check-signature R-3.5.1.pkg`

Latest release:

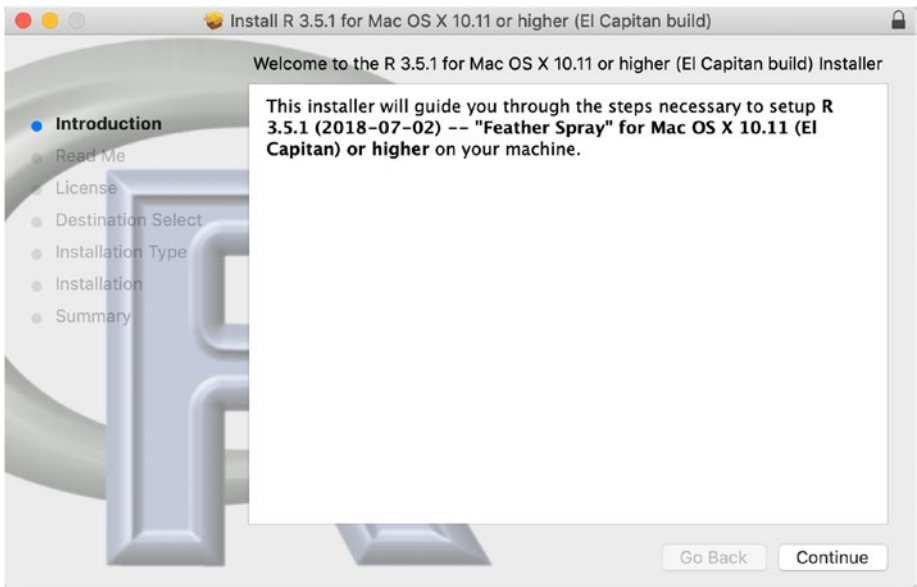
[R-3.5.1.pkg](#)  
 MD5=3a6b\_5ba0060b0020257d1e511e17e708  
 SHA1:  
 ba6b\_76d91916a2a8998256a511e25417279d1d9621  
 (ca. 74MB)

**R 3.5.1** binary for OS X 10.11 (El Capitan) and higher, signed package. Contains R 3.5.1 framework, R.app GUI 1.70 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 5.2. The latter two components are optional and can be omitted when choosing "custom install", they are only needed if you want to use the `texi2k` R package or build package documentation from sources.

- CRAN
- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)
- About R
- [R Homepage](#)
- [The R Journal](#)
- Software
- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)
- Documentation
- [Manuals](#)
- [FAQs](#)
- [Contributed](#)

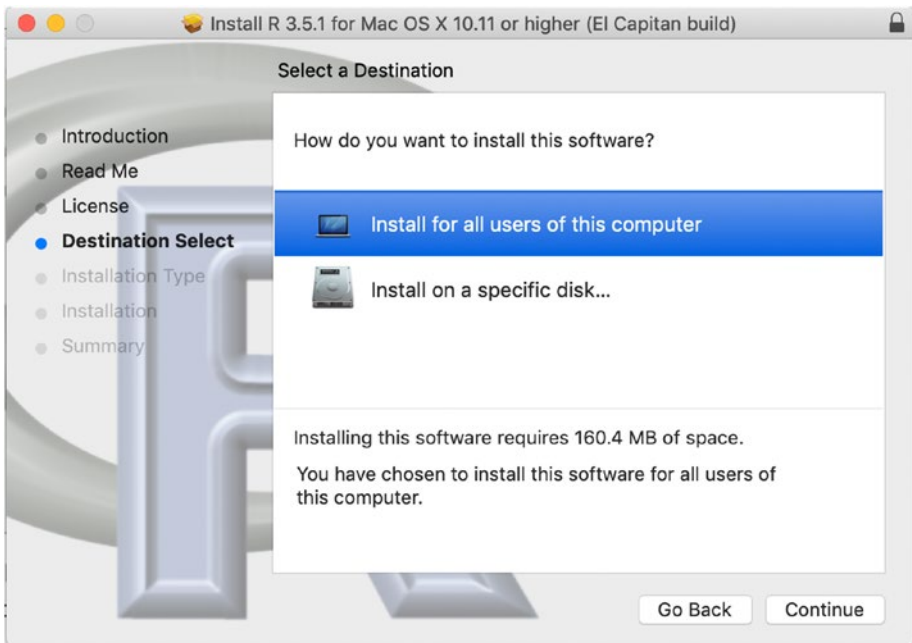
**Figure 1-3.** Mac R installer

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.



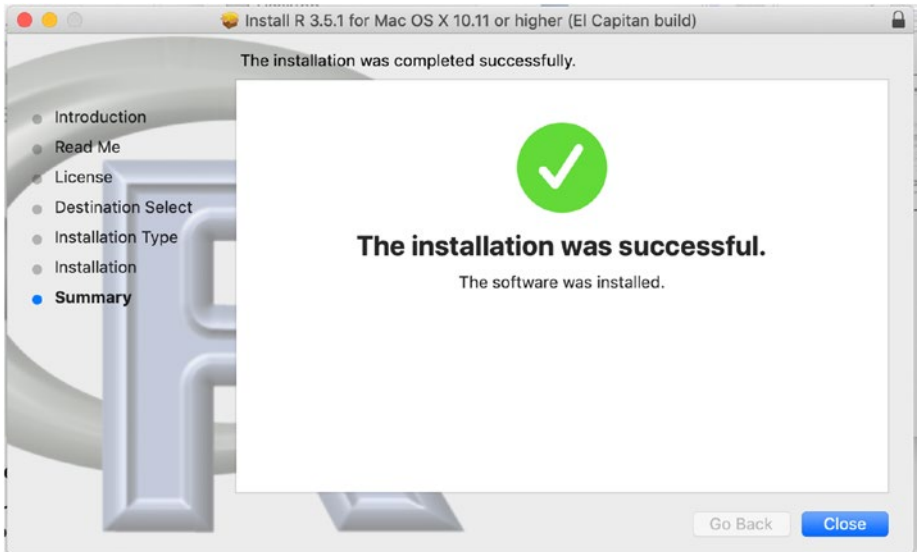
**Figure 1-4.** *Installing R on Mac*

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.