You Can Learn R

Sarah Manski

2023-04-02

Contents

Preface

0.1 Why did I create this resource?

In my experience teaching beginning R learners, the initial reaction of nearly every person trying to learn R is "R is hard!" The steep learning curve is enough to turn away even the most advantaged R learners. For example, I started computer programming in high school, was exposed to R in my undergraduate studies, and went to graduate school for statistics where I taught myself R (with the help of my more experienced R user brother). I had all the advantages when learning R and it was still a daunting task. In Fall 2022, I was part of a weekly professional development group on making math and statistics education accessible and inclusive. We were asked to think about how we could improve accessibility in our own field. My ambition got the better of me and I decided to try tackling one of the least accessible learning experiences I know: learning R. I created this resource to make learning R, which is often exclusive and inaccessible to most audiences, a more accessible and inclusive experience for anyone hoping to learn R.

Note: I created this resource, not to teach people R, but to teach people how to learn R and give them a set of tools to get started.

0.2 Who is this for?

- As the title implies, this is for YOU! My goal is to create a resource that enables anyone to start learning R and caters to as many audiences as possible.

- This resource was originally designed and presented to early-career researchers including a diverse group of graduate students and post-doctoral scholars at Michigan State University, as well as early-career agronomists in Africa.

0.3 About the Resources

- I will continue to add to the recommended packages, learning resources, and R communities. However, I am only one person and am not familiar with all the packages, resources, and communities you might use. If you find something you think should be added or would be beneficial to improve this book, please let me know!
- Resources will lean toward quick tutorials and learning to get up and going with certain tasks and packages. This is because there are many extensive online courses you can take in R, but this resource is aimed at researchers that have limited time and want to teach themselves R as they go, in order to complete data-driven tasks related to their research.

Chapter 1

An introductory seminar

This part is intended to be given as an in-person, cooperative learning seminar. To follow along, visit https://posit.cloud/content/5486172

1.1 R: What, why, and how?

1.1.1 What is R? What is RStudio/ Posit Cloud?

R is a statistical programming language that can be used for any data-driven task.

RStudio is a desktop environment for working with R while Posit Cloud is an online version of RStudio.

You can think of R like the underlying computer language on a smartphone and RStudio like the operating system on a smartphone. While R contains the commands and functions you need to solve data-driven tasks, RStudio is a human interface that allows you to run and save R code, view data and plots, access help documents, and much more.

1.1.2 Why use R?

- R is free and open-source
- R is the most extensive set of tools for data science and statistics
- Using R promotes reproducibility
- R is not just for statistics
  - Data visualization
  - Data manipulation

  - Generating reports
  - Creating data-driven apps

- Anyone and everyone can benefit from working with R
- R illustrates, communicates, and automates
- R has an expansive and welcoming learning community

1.1.3 Why use RStudio or Posit Cloud?

- RStudio and Posit Cloud are free and can be run on any system
- RStudio is a desktop application and supports offline use
- Posit Cloud allows online access to your projects and offers convenient sharing

1.1.4 How R works

Base R, packages, and functions: A smartphone analogy

- Base R is similar to the factory-installed apps that come standard on a smartphone. This functionality is already built-in when you download R.
- Packages are like additional apps you can download to add functionality to do various tasks (such as an email app adds the ability to receive and send emails). Packages need to be installed once and then loaded each time you need to use it. R has over 10,000 packages and more are being created every day!
- Functions are commands in R that perform a specific function, similar to the way phone apps perform specific tasks.

1.1.5 How to get started with R

- Option 1: Posit Cloud
  - If you have regular internet access and want a quick start-up to R that you can use while following this book, I recommend Posit Cloud.
  - https://posit.cloud
- Option 2: R and RStudio
  - If you want to use R but may not always have internet access, download R and RStudio.
  - https://posit.co/download/rstudio-desktop/

(Note: RStudio and Posit are the same, but RStudio rebranded in early 2023 as Posit in order to also encompass the Python community)

1.2 How to learn R

- Motivate yourself: What do you want to do with R?
- Projects related to your own research
- Sample projects using sample data
- R Community Projects like #TidyTuesday
- Write out the steps explicitly
- Strategic Searching
  - R Help menus and package documentation
  - Search online using package names, specific sites, and "R"
  - Copy and paste examples - just know what they do!
- Learn with others and find your R community
- Manage your expectations
  - R can do anything, but you don't need to know it all

For a great related article, see [this link] (https://www.dataquest.io/blog/top-tips-for-learning-r-from-africa-rs-shelmith-kariuki/)

1.3 When things go wrong

First things first: EVERYONE makes mistakes and tings WILL go wrong.

- Common Errors:
  - Capitalization
  - Misspelling
  - Closing punctuation
  - Continuing punctuation
  - Conflicting code
  - Copy/ paste errors
  - Libraries not loaded
  - Unsaved objects
- Troubleshooting
  - Parse your error message to find helpful information
  - Run code one line at a time
  - Strategic searching: copy/paste error message into Google

- When you're really stuck
  - Turn it off and on again
  - Take a break
  - Ask someone for help

For more details, see the section in R for Graduate Students discussing troubleshooting error messages

Chapter 2

Resources for learning R

2.1 Overview of resources

This part is intended to provide external resources that will help you in learning R. This part is divided into three sections: recommended packages, learning resources, and learning communities. The first section gives a list of commonly used packages that you may find useful in your research. These encompass a small subset of the thousands of packages available and will not include all the packages you will eventually need, but offers some recommended packages to get started.

The second section includes recommended learning resources including written materials, interactive tutorials, data sources, and sites to search. These resources only encompass a small number of the expanse of learning materials related to R. However, they offer a few places to get started to help learners avoid the overwhelm of too many learning options and nowhere to start.

The third section gives recommended learning communities. The worldwide R learning community is expansive and welcoming to learners at all levels. We hope you can find a community among those listed to interact and learn alongside other people interested in R.

Finally, an appendix is included offering numerous R package cheatsheets. Many packages have a cheatsheet designed for aiding quick understanding and use of functions. While I don't recommend using these for learning R, they can be useful as a quick reference when working in R.

2.2 Recommended packages

2.2.1 Tidyverse

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures. Installing the tidyverse will install all tidyverse packages. To see a full list of tidyverse packages, see https://www.tidyverse.org/packages/

Core tidyverse - Loading the tidyverse using library(tidyverse) will load 8 core packages from the tidyverse that you are likely to use in everyday data analyses. These packages are:

- ggplot2 - ggplot2 is a system for creating graphics, based on The Grammar of Graphics (gg). ggplot2 is a powerful tool for visualizing data.
- dplyr - dplyr provides a grammar of data manipulation, providing a consistent set of verbs that solve the most common data manipulation challenges. These include select, filter, mutate, summarize, etc.
- tidyr - tidyr provides a set of functions that help you get to tidy data. Tidy data is data with a consistent form where every variable goes in a column and every observation is a row.
- readr - readr provides a fast and friendly way to read rectangular data (like csv, tsv, and fwf). It is designed to flexibly parse many types of data.
- tibble - tibble provides opinionated data frames that make working in the tidyverse a little easier.
- stringr - stringr provides a cohesive set of functions designed to make working with strings as easy as possible.
- forcats - forcats provides a suite of useful tools that solve common problems with factors, which R uses to handle categorical variables.
- purrr - purrr enhances R's functional programming toolkit by providing a complete and consistent set of tools for working with functions and vectors. Once you master the basic concepts, purrr allows you to replace many for loops with code that is ieasier to write and more expressive

2.2.2 Reporting Results

- R Markdown - R Markdown allows for reproducible reports that contain R code. R Markdown is integrated into RStudio.
- shiny - shiny is used to make interactive web apps with R in order to explore data and share findings.

2.3 Recommended learning resources

2.3.1 Written Materials

- R documentation, help files, and vignettes
- R for Data Science
- R for Graduate Students

2.3.2 Interactive Tutorials

- Posit Cloud Primers

2.3.3 Data Sources

- Built-in R datasets
  - For a quick tutorial on accessing datasets in R, see here
- TidyTuesday using the tidytuesdayR package

2.3.4 Sites to search

- StackOverflow keyword R
- R bloggers
  - R news and tutorials contributed by hundreds of R bloggers
- rdrr.io
  - A comprehensive and searchable R documentation site
- The Comprehensive R Archive Network (CRAN)

2.4 Find your R learning community

2.4.1 Global Communities

- R for Data Science Community
- RStudio and Posit communities
- R-ladies

2.4.2 Communities for under-represented R user groups

- AfricaR
- R-ladies
- Minorities in R

2.5 Cheat sheets for quick reference

2.5.1 Recommended Cheat Sheets

All cheat sheets are downloadable at the associated links.

Base R - https://github.com/rstudio/cheatsheets/blob/main/base-r.pdf

RStudio IDE - https://github.com/rstudio/cheatsheets/blob/main/rstudio-ide.pdf

Data Import with Tidyverse (readr, readxl, googlesheets4) - https://github.com/rstudio/cheatsheets/blob/main/data-import.pdf

Data Visualization with ggplot2 - https://github.com/rstudio/cheatsheets/blob/main/data-visualization.pdf

2.5.2 Full list of cheat sheets:

- arrow
- base-r
- bayesplot
- bcea
- caret
- cartography
- collapse
- data-import
- data-transformation
- data-visualization
- datatable
- declaredesign
- distr6
- estimatr
- eurostat
- factors
- gganimate
- git-github
- golem
- gtsummary

- gwasrapidd
- h2o
- imputeTS
- jfa
- keras
- labelled
- leaflet
- lubridate
- mapsf
- mlr
- mosaic
- nardl
- nimble
- oSCR
- overviewR
- package-development
- packagefinder
- parallel_computation
- plumber
- purrr
- quanteda
- quincunx
- randomizr
- regex
- reticulate
- rmarkdown
- rphylopic
- rstudio-ide
- sas-r
- sf
- shiny
- sjmisc
- slackr
- sparklyr
- stat2r
- strings
- survminer
- syntax
- teachR
- tidyeval
- tidyr
- time-series
- torch
- tsbox
- vegan
- vtree

- xplain

Chapter 3

Accessibility Recommendations

3.1 Tips for limited internet access

- R help files, documentation, and vignettes are downloaded with installed packages and can be used offline after initial package installation
- Download R Cheatsheets as a quick reference
- Bookdown books such as R for Graduate Students can often be downloaded as a pdf, epub, or HTML file

3.2 Translated materials

- R for Data Science
  - Spanish
  - Italian
  - Turkish
- Translated Cheatsheets
  - R User translated cheatsheets

3.3 Blind R users

- Go to https://r-resources.massey.ac.nz/StatSoftware/ for comprehensive information