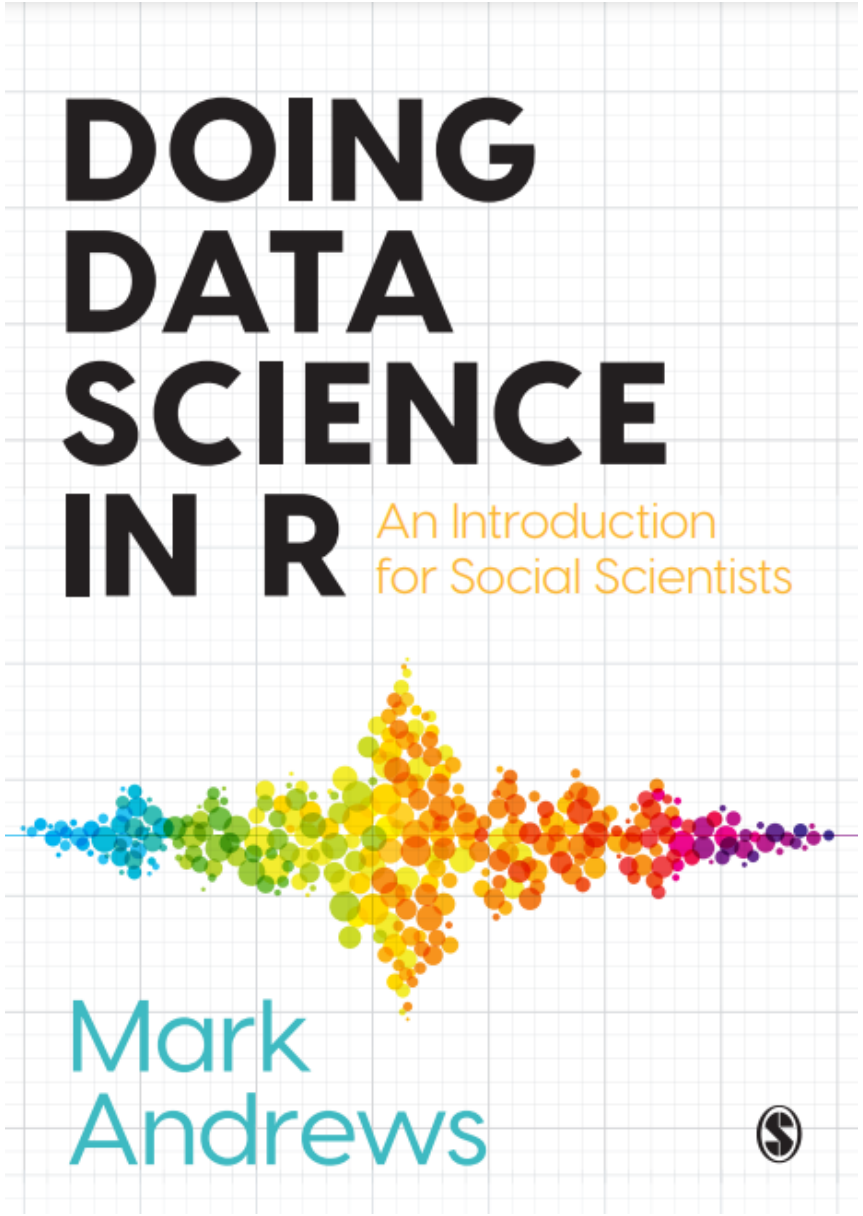

Doing Data Science in R: An Introduction for Social Scientists

Chapter 2

Introduction to R

What is R, and why should we use it?

A program for doing statistics and data analysis

They include SPSS, SAS, Stata, Minitab, Python, Matlab, Maple, Mathematica, Tableau, Excel, SQL, and many others.

A power tool for data analysis

Built into R's standard set of packages is virtually the entire repertoire of widely known and used statistical methods.

Also built into R is an extensive graphics library

A power tool for data analysis

In addition to its built-in tools, R has a vast set of add-on or contributed packages

The R programming language itself can be extended by interfacing with other programming languages like C, C++, Fortran and Python.

Open source software

R is free and open source software, distributed according to the GNU public licence.

- The freedom to run the program in any manner and for any purpose
- The freedom to study and modify the source code
- The freedom to distribute copies of the original code
- The freedom to distribute modified versions of the code.

Popularity

The Journal of Statistical Software is the most widely used academic journal describing advances and developments in software for statistics

R is currently very highly ranked according to many rankings of widely used programming languages of any kind

Installing R and RStudio

It usually involves two, or maybe three, separate steps

R has over 16,000 add-on packages

When first installing R, it's often a good idea to also install a minimal set of must-have packages. After that, additional R packages can be installed as and when they are needed.

Installing R

To install the latest version of R on Windows, go to:

https://cran.r-project.org/bin/windows/base/

For the installer for the latest version of R for Macs, go to:

https://cran.r-project.org/bin/macosx/

Installing RStudio Desktop

The RStudio Desktop is one of the software products created by the company RStudio.

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

Guided Tour of RStudio Desktop

RStudio is the company that created and maintains the RStudio Desktop software, among other pieces of software

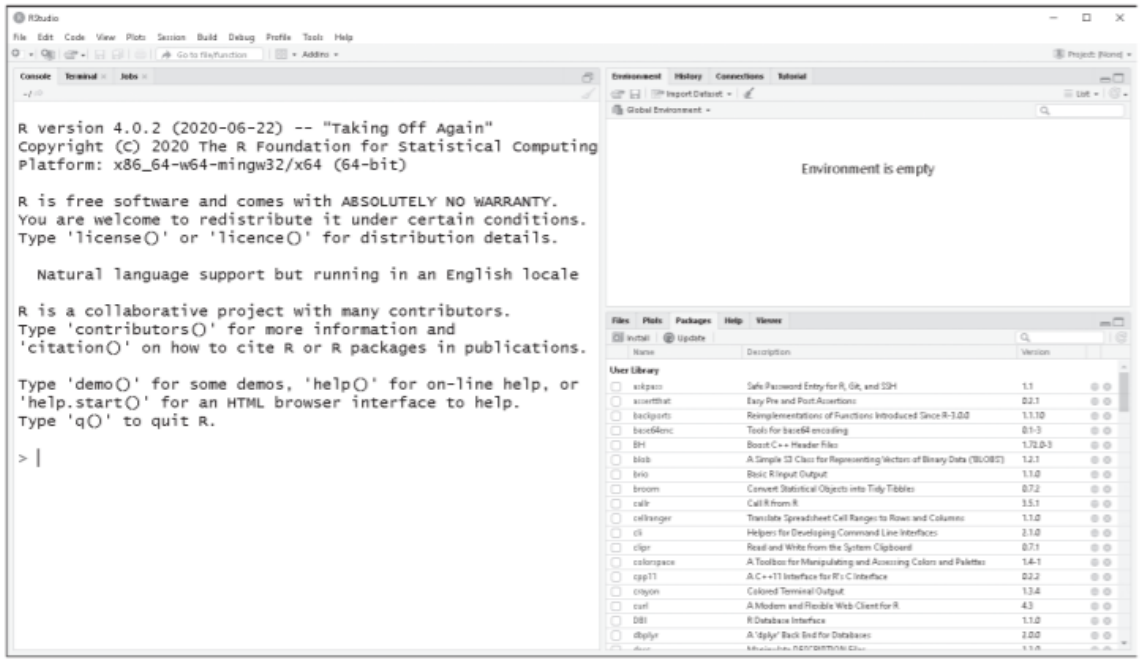

Figure 2.1 The typical layout of the RStudio Desktop when it is first opened

First steps in R

People who wish to learn R may simply not know where to begin

To learn R, it is best to learn the fundamentals first.

Step 0: Using the R console

R is a command-based system. We type commands, R translates them into machine instructions, which our computer then executes

When learning R, it is usually best to start with typing commands in the R console.

Step 1: Using R as a calculator

Just as we would with a calculator, we can start using R by doing arithmetic – adding, subtracting, multiplying, dividing, and so on.

Step 2: Variables and assignment

A major step forward in using R is the use of variables and the assignment of values to variables.

Step 3: Vectors

Vectors are one-dimensional sequences of values.

While they will often be created for us by the R functions that we use, such as by some data analysis functions, we can also create vectors ourselves using the c() function.

Step 4: Data frames

The data frame is probably the most important data structure in R.

Dframes are how we almost always represent real-world data sets in R, and most statistical analysis commands, especially modern ones, assume data is provided in the form of data frames

Step 5: Other data structures

Lists in R allow for the storage of multiple heterogeneous data structures

Matrices are equivalent to two-dimensional vectors

Arrays are n-dimensional generalizations of matrices,

Step 6: Functions

While data structures hold data in R, functions are used to do things to or with the data

In almost all functions, we put data structures in, calculations are done to or using this data, and new data structures, perhaps just a single value, are then returned.

Step 7: Scripts

Scripts are files where we write R commands, which can be then saved for later use.

Step 8: Installing and loading packages

There are over 16,000 add-on R packages as of August 2020.

Step 9: Reading in and viewing data

R allows us to import data from a very large variety of data file types, including from other statistics programs such as SPSS, Stata, SAS, Minitab, and common file formats like .xlsx and .csv. H

Step 10: Working directory, RStudio projects, and clean workspaces

Every R session has a working directory, which we can think of as the directory (or folder) on the computer's file system in which the R session is running