Vol. 0: Basics in R

Sarah Schwartz & Tyson Barrett

Last updated: 2018-08-13

Contents

Chapter 1

Introduction

What is R?

R is a language and environment for statistical computing and graphics. (R Core Team, 2018)

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

What is R Markdown?

According to R Studio:

> "R Markdown is a format that enables easy authoring of reproducible web reports from R. It combines the core syntax of Markdown (an easy-to-write **plain text** format for web content) with embedded **R code chunks** that are run so their output can be included in the final document".

Dynamic Reporting

From Penn State Statistics:

The traditional way** to write a report**

1. Run your analysis in software, like SPSS or R and manually save our output
   - *i.e. saving the ANOVA table or using pdf() to save the graphs*
2. Type your your description and interpretation in a text editor like *Word*
   - *either drag/drop tables and figures, or worse copy-paste and retype all the numbers*

A report written in this way can be problematic. For instance, imagine your *Mentor/collaborator/journal reviewer* telling you that they want to use a sub-sample instead of the entire sample. Or to include a nother variable. You would have to redo all of your work!!

Therefore, in this way **dynamic also means reproducible**, in the sense that people who get the file from you can reproduce the entire work in the report.

How does R Markdown work out to be a .pdf or .html file?

`R Markdown` is a file with the file extension **.Rmd**, the `knitr` package will then transform the file into a **Markdown** file with the extension **.md.** Then Rstudio can (Xie, 2015):

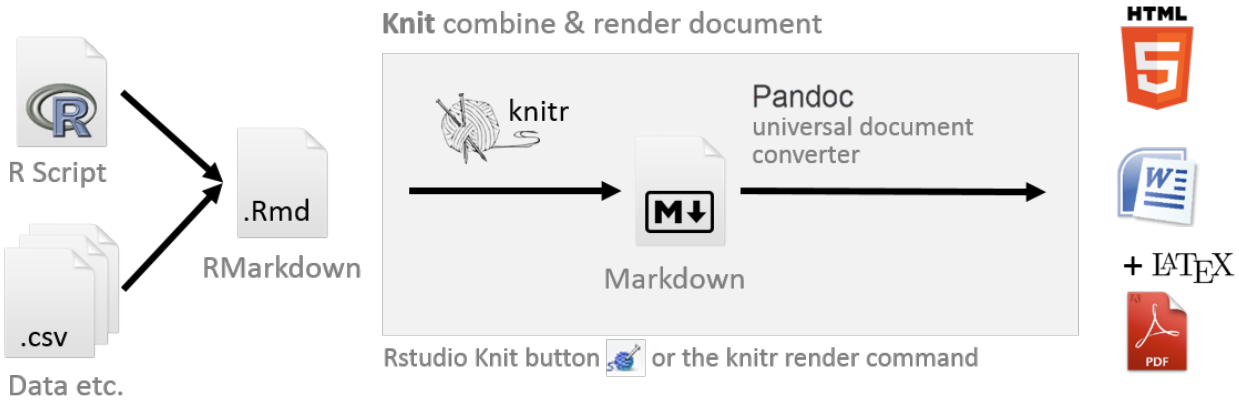

Figure 1.1:

- Use `LaTeX` to transform the file into a **.pdf**
- Load another package called `markdown` to transform the file into **.html**
- Use Pandoc to even convert to file to a **Word** document (ugly)

Is this a popular** method for creating reports?**

Check out Rpubs. This website shares lots of documents written in the way we will introduce below.

`R Markdown` documents are fully reproducible. Use a productive **notebook** interface to weave together narrative text and code to produce elegantly formatted output. Use multiple languages including R, Python, and SQL (Allaire et al., 2018).

`knitr` is an engine for dynamic report generation with R. It is a package in the statistical programming language R that enables integration of **R code** into LaTeX, LyX, HTML, Markdown, AsciiDoc, and **text** documents (Xie, 2018).

Figure 1.2:

Figure 1.3:

Chapter 2

Software Installation

Here is where we talk about installing software.

2.1 R

2.1.1 First Time Installation

2.1.2 Periotic Updating

2.2 R Studio

2.2.1 First Time Installation

2.2.2 Periotic Updating

2.2.3 Panel Layout

2.3 TeX (optional)

2.3.1 TinyTeX

2.3.2 MAc

2.3.3 Windows

Chapter 3

Packages Management

We describe packages and their management

3.1 What are packages

3.2 How to install packages

3.3 Updating packages

3.4 Suggested packages

Chapter 4

Data Management

How do you get data into R, view and work with in, and then save it for later use.

4.1 Importing Data From Various Formats

4.1.1 Text Format (.csv, tab-delimited, ect.)

Use `read.csv()`

4.1.2 Excel Format (.xls, .xlsx)

Use `haven::read.spss()`

4.1.3 SPSS Format (.sav)

Use `readxl::read.excel()`

4.1.4 REDCap (API directly)

4.2 Viewing Data Within R Studio

4.2.1 The Environment Tab

4.2.2 Notebook Display

4.3 Saving Data in R Format

Use `save(..., file = "name.RData")`

Chapter 5

Data Wrangling

5.1 Subseting Data

5.1.1 Select Variables (columns)

Use `dplyr::select()`

5.1.2 Select Observations (rows)

Use `dplyr::filter()`

5.2 Symbol Opporators

5.2.1 Logical Opporators

5.2.1.1 Ineqalities (<, >, <=, <=)

5.2.1.2 AND (&)

5.2.1.3 OR (|)

5.2.1.4 Within a List (%in%)

5.2.2 The Assignemnt Opporator (<-)

5.2.3 The Pipe Opporator (%>%)

Bibliography

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., and Chang, W. (2018). *rmarkdown: Dynamic Documents for R*. R package version 1.10.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2018). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.20.