R-Markdown

Wei Zhong

R Markdown Introduction

R Markdown provides an unified authoring framework for data science, combining your code, its results, and your prose commentary. R Markdown documents are fully reproducible and support dozens of output formats, like PDFs, Word files, slideshows, and more.

R Markdown files are designed to be used in three ways:

- For communicating to decision makers, who want to focus on the conclusions, not the code behind the analysis.
- For collaborating with other data scientists, who are interested in both your conclusions, and how you reached them (i.e. the code).
- As an environment in which to do data science, as a modern day lab notebook where you can capture not only what you did, but also what you were thinking.

R Markdown basics

This is an R Markdown file, a plain text file that has the extension .Rmd:

It contains three important types of content:

- An (optional) YAML header surrounded by `---`s.
- Chunks of R code surrounded by triple backward quotes
- Text mixed with simple text formatting like # heading and `_italics_`.

R-Markdown Execute

When you open an .Rmd, you get a notebook interface where code and output are interleaved. You can run each code chunk by clicking the Run icon (it looks like a play button at the top of the chunk), or by pressing Cmd/Ctrl + Shift + Enter. RStudio executes the code and displays the results inline with the code.

To produce a complete report containing all text, code, and results, click "Knit" or press Cmd/Ctrl + Shift + K.

When you knit the document, R Markdown sends the .Rmd file to knitr, which executes all of the code chunks and creates a new markdown (.md) document which includes the code and its output.

Text formating with Markdown

```
1. *italic*  or _italic_
2. **bold** or  __bold__
3. `code`
4. superscript^2^ and subscript~2~
```

Headings

```
# 1st Level Header

## 2nd level Header

### 3rd level Header
```

Lists

```
*   Bulleted list item 1

*   Item 2

    * Item 2a

    * Item 2b

1.  Numbered list item 1

1.  Item 2. The numbers are incremented automatically
            in the output.
```

Links and images

<https://www.google.com>

[linked phrase](https://www.google.com)

![Virtual Screening Data](http://www.molsoft.com/gui/plot.png)

Code chunks

To run code inside an R Markdown document, you need to insert a chunk. There are three ways to do so:

- The keyboard shortcut Cmd/Ctrl + Alt + I
- The "Insert" button icon in the editor toolbar.
- By manually typing the chunk delimiters `{r}` `and`.

Chunk name

Chunks can be given an optional name: {r by-name}. You can more easily navigate to specific chunks using the drop-down code navigator in the bottom-left of the script editor.

There is one chunk name that imbues special behaviour: setup. When you're in a notebook mode, the chunk named setup will be run automatically once, before any other code is run (enable caching globally for a document).

Chunk options

Chunk output can be customised with options, arguments supplied to chunk header. Knitr provides almost 60 options that you can use to customize your code chunks. Here we'll cover the most important chunk options that you'll use frequently. You can see the full list at `http://yihui.name/knitr/options/`.

The most important set of options controls if your code block is executed and what results are inserted in the finished report:

Chunk options

- eval = FALSE prevents code from being evaluated. (And obviously if the code is not run, no results will be generated). This is useful for displaying example code, or for disabling a large block of code without commenting each line.
- include = FALSE runs the code, but doesn't show the code or results in the final document. Use this for setup code that you don't want cluttering your report.
- echo = FALSE prevents code, but not the results from appearing in the finished file. Use this when writing reports aimed at people who don't want to see the underlying R code.
- message = FALSE or warning = FALSE prevents messages or warnings from appearing in the finished file.
- results = 'hide' hides printed output; fig.show = 'hide' hides plots.
- error = TRUE causes the render to continue even if code returns an error. This is rarely something you'll want to include in the final version of your report, but can be very useful if you need to debug exactly what is going on inside your .Rmd.

Table

By default, R Markdown prints data frames and matrices as you'd see them in the console:

For example, in the chunk you can type: mtcars[1:5,].

If you prefer that data be displayed with additional formatting you can use the knitr::kable function or stargazer.

YAML header

You can control many other "whole document" settings by tweaking the parameters of the YAML header. For example:

```
---
title: "STAT 429 Homework # 2 "
author: "Wei Zhong (NetID: wzhong8)"
date: "October 8, 2016"
header-includes:
   - \usepackage{bbm}
   - \usepackage{physics}
output:
  pdf_document:
    fig_width: 7
    fig_height: 3.2
    fig_caption: true
geometry: margin = 1in
fontsize: 11pt
---
```

Bibliographies and Citations

You can specify a bibliography file using the bibliography field in your file's header. The field should contain a path from the directory that contains your .Rmd file to the file that contains the bibliography file: e.g. bibliography: ~/Dropbox/prospectus_ref.bib

You can use many common bibliography formats including BibLaTeX, BibTeX, endnote, medline.

To create a citation within your .Rmd file, use a key composed of '@' + the citation identifier from the bibliography file. Then place the citation in square brackets.

Bibliographies and Citations Examples

Here are some examples:

Separate multiple citations with a ;

- Blah blah [@smith04; @doe99].

You can add arbitrary comments inside the square brackets:

- Blah blah [see @doe99, pp. 33-35; also @smith04, ch. 1].

Remove the square brackets to create an in-text citation: @smith04

- says blah, or @smith04 [p. 33] says blah.

Add a '-' before the citation to suppress the author's name:

- Smith says blah [-@smith04].

Bibliographies and Citations

When R Markdown renders your file, it will build and append a bibliography to the end of your document. The bibliography will contain each of the cited references from your bibliography file, but it will not contain a section heading. As a result it is common practice to end your file with a section header for the bibliography, such as # References or # Bibliography.