R Notebook Tutorial

Tufts BUGS workshop

Eric R. Scott

2018-01-16

Contents

What is an R Notebook? **1**

Why Notebooks? **1**

Workshop Outline **1**

Make a Notebook! 1
Basics of Markdown 2
Code chunks 3
YAML header 6
Reports 6
Advanced options for reports 7

What is an R Notebook?

- A file format (.rmd) that can be opened by R Studio
- Allow you to mix prose, code, and code output
- Uses a simple system called Markdown for formatting prose
- Markdown headings allow easy navigation in R Studio
- Notebooks can be "knit" to produce beautiful reports as html, pdf, or Word documents
- Allows configuration of what to show in these outputs (e.g. code, output, warnings)

Why Notebooks?

- It's a good practice to annotate your code anyway, and .rmd lets you do this in a richer, easier to read way
- As an analysis notebook. Organizing your code into "chunks" interspersed with prose helps you think through problems better and keep track of your thought process.
- Portability/collaboration. No need to use `setwd()`
- Beautiful, easy to navigate reports. You can hide or show code depending on who the report is for (e.g. does this person know R?)
- Creating examples, tutorials, assignments, etc. using R
- Plain text (despite richly formatted output), so very compatible with Git and other version control.

Workshop Outline

Make a Notebook!

Make a new notebook, run the chunks, preview the output

A note on filenames: When generating a report as a pdf, part of the process doesn't like filenames with spaces in them. This isn't *always* a problem, but it's best to avoid spaces in .rmd filenames to avoid problems.

Basics of Markdown

Basic Formatting

bold, *itallics*, subscript^{x}, superscript$_{y}$, formatting as `code` with backticks

List Header:

- item 1
- item 2

Note: for lists to show up correctly in reports, you must have a blank line between your list header (or any previous text) and the first bullet point of your list. You must also have a space between the "-" and the list item. For example, this won't look right: - no preceding empty line - effects the entire list, not just the first item

-no space after the hyphen and it won't format right

- this one looks right

Headings and subheadings

Nested headings can be added using different numbers of "#" at the beginning of a line. In addition to being formated differently in report output, these are integrated with R Studio and show up in the table of contents menu in the lower lefthand corner of the editor window.

Equations

Inserting equations can be done by surrounding the equation with the dollar sign character. For example, compare these two equations in the output: n = 5, $n = 5$. Two dollar signs makes a "display equation".

$$n = 5$$

More complicated equations can be made using LaTeX:

$$s = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}}$$

A cheat-sheet for making equations like this can be found here

When Markdown doesn't work

Markdown is simple, but not exhaustive—there are many things that just plain can't be done using markdown. If you know you'll be outputting reports as html, you can use html where markdown fails, but it will be ignored in pdf documents. Likewise, LaTeX will sometimes work for pdf output when markdown fails, but it will be ignored in html output. There are three places to get more info on R Markdown in R Studio: *Help > Markdown Quick Reference*, *Help > Cheatsheets > R Markdown Cheatsheet*, and *Help > Cheatsheets > R Markdown Reference Guide*. If I have a specific problem I want to solve, I often find it more helpful to google the problem to start.

Code chunks

Inserting and naming code chunks

Code chunks can be inserted with the insert button, with a keyboard shorcut (*cmd + option + i* by default) or just by typing the characters that enclose a code chunk (starting with three backticks)

Code chunks can be named by clicking on the small gear icon in the upper right corner of a code chunk and typing a name into the name field. This name will show up in the table of contents menu nested under whatever subehading it's in.

```
testtext <- "This code chunk has a name"
testtext
```

```
## [1] "This code chunk has a name"
```

In addition to using code chunks, you can also insert inline code. For example, random number:-0.6706567. Previous output: This code chunk has a name

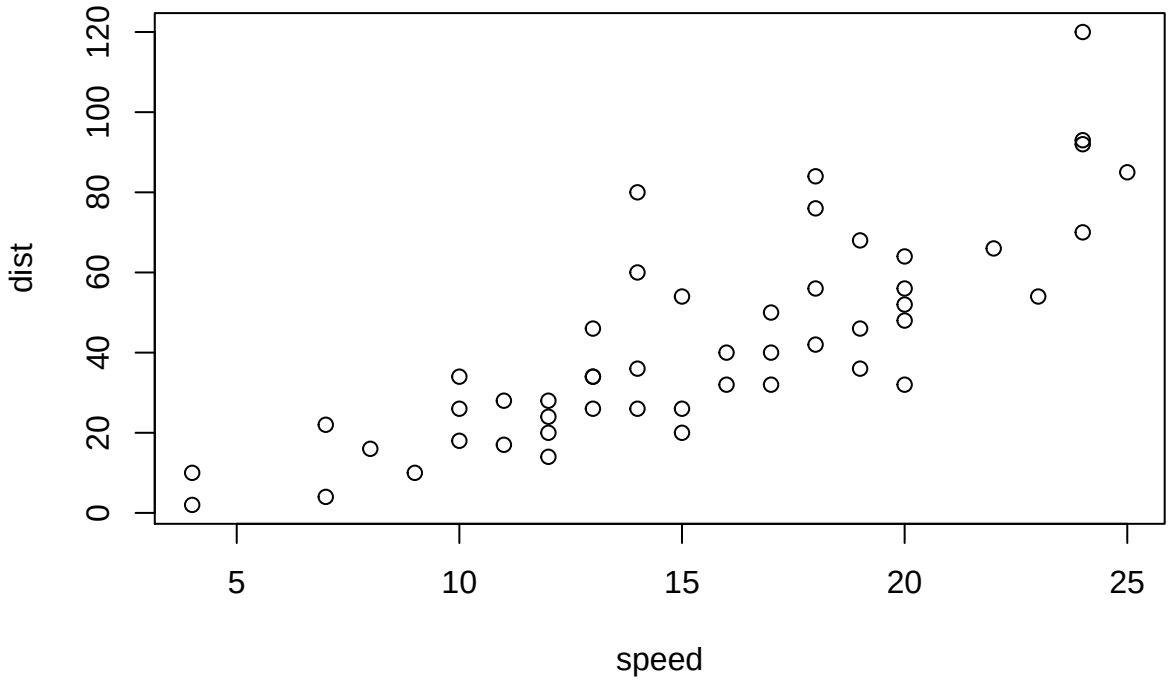
Running code in code chunks

To run code in a code chunk you can do a few things. You can run the whole chunk with the "play" button in the corner of the chunk. You can run a single line with the usual keyboard shorcut (*cmd + enter* by default). You can run all previous chunks with the icon between the gear and the "play" button. Finally, you can find several more options for running code in the Run button. *Restart R and Run All Chunks* is particuarly useful. To generate a report (e.g. as a pdf) the R Notebook must be totally self-contained. That is, all the chunks must run without error and can't rely on objects created by other R files. By restarting R and running all chunks, you can easily troubleshoot issues related to lack of self-containedness.

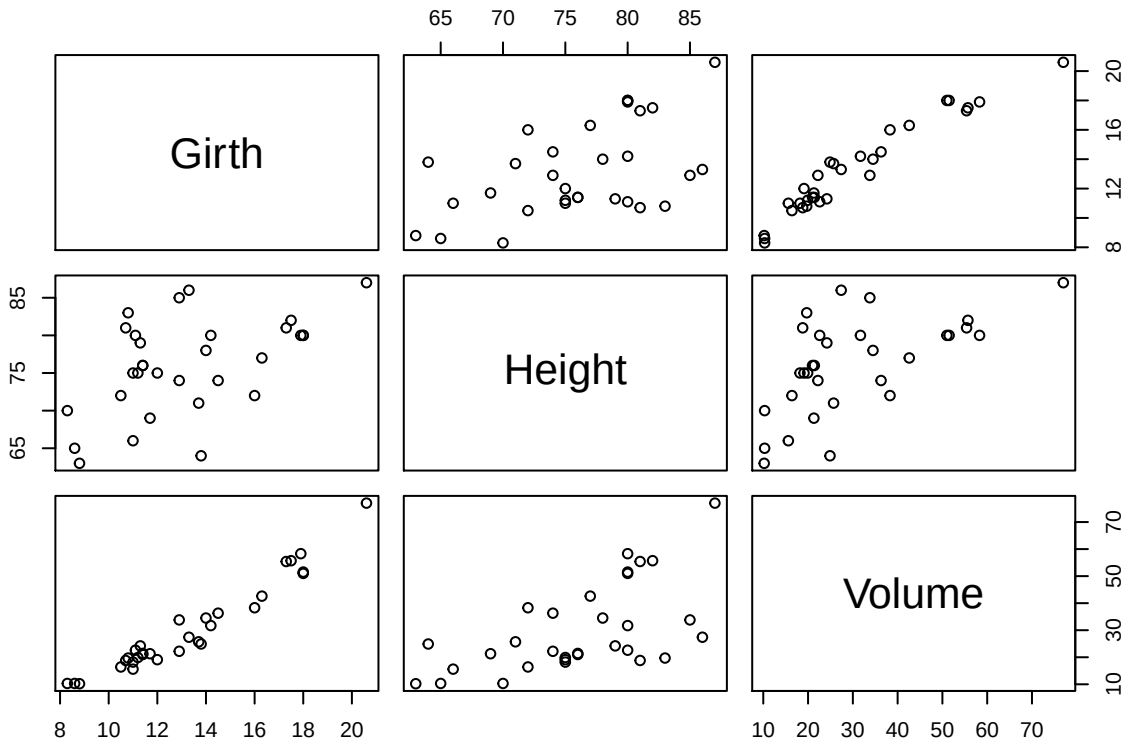
Speaking of self-containedness, R Notebooks don't require you to use `setwd()`. Your working directory is automatically whatever folder the R Notebook file is in. If you want to point to another folder, I recommend using the path explicitly rather than trying to `setwd()`. `setwd()` doesn't always work like you'd expect in R Notebooks and seems to only affect code in the same code chunk as the `setwd()` call.

If your code chunk produces multiple outputs, they will show up in separate "tabs."

```
plot(cars)
```




```
plot(trees)
```




```
cor(trees)
```

```
##            Girth    Height    Volume
## Girth  1.0000000 0.5192801 0.9671194
## Height 0.5192801 1.0000000 0.5982497
## Volume 0.9671194 0.5982497 1.0000000
```

This style of output works with html file reports, but obviously not with Word or pdfs. In the case of

non-interactive formats, the three outputs in this example will be shown after each line of code that produces them.

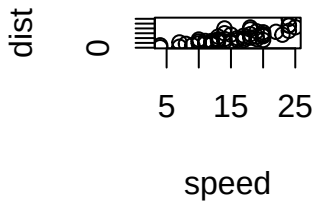
Code chunk options

In the upper right corner of code chunks, there is a gear icon that can be used to access common chunk options. You can specify whether you want the code, output, or both to be shown in the report you generate. You can also specify whether you want the code run or not. This is useful for keeping example code or code that doesn't work yet in your notebook while avoiding errors when generating reports.

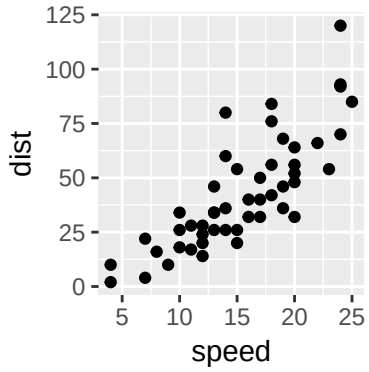
You can specify figure size here as well. *Note*: In my experience, base R plots tend to have very large margins and the end up a lot smaller than you'd expect based on setting these options compared to ggplot versions.

```
#base R plot resized to 2" x 2"
plot(cars)

#ggplot resized to 2" x 2"
library(ggplot2)
```




```
ggplot(cars) +
  geom_point(aes(x = speed, y = dist))
```



Using code chunks for non-code-like output

```
library(knitr)
```

Use `kable()` to produce pretty tables in pdf and Word output. They are, however, less pretty in html output (including in-line output and Preview output)

```
head(trees)
```

```
##   Girth Height Volume
## 1   8.3     70   10.3
## 2   8.6     65   10.3
## 3   8.8     63   10.2
## 4  10.5     72   16.4
```

```
## 5  10.7     81   18.8
## 6  10.8     83   19.7
```

```
kable(head(trees))
```

Girth	Height	Volume
8.3	70	10.3
8.6	65	10.3
8.8	63	10.2
10.5	72	16.4
10.7	81	18.8
10.8	83	19.7

- `include_graphics()`—similar to using markdown syntax, but allows use of code chunk options

```
include_graphics("kitten.jpg")
```

YAML header

The default YAML header in a new R Notebook file only has the title and output method (what filetype you want to use to generate a report—it's best not to change this manually). However, many other things can go in the YAML header. See the header in this document for some examples.

Reports

Generating pdf, word, or html documents is done using the "knit" button (which, by default says "preview"). This invokes functions from the R package `knitr` to produce the desired output. For pdf, that involves LaTeX, so you must have this installed on your computer to generate pdf output.

The major difference between "Preview" and "knit..." is that preview does not run all of your chunks explicitly. That is, if you create a new chunk with some code in it, but do not run it, no output from that chunk will show up in the preview. However, anytime you "knit" a R Notebook, it will start by running all of the chunks and generating output.

Filetypes available by default:

- PDF (requires LaTeX)
- Word (not as pretty, but easily editable)
- html (.nb.html files are great to share! Only subtle difference with .html output.)

Advanced options for reports

- In gear icon, output options...
- Floating TOC with html. Tutorial here
- Code-folding
- themes and highlighting
- Setting global chunk options (`opts_chunk$set <- ...`)