R Notebook Tutorial

Tufts BUGS workshop

Eric R. Scott

2018-01-16

Contents

1 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)

2 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.

3 Workshop Outline

3.1 Make a Notebook!

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

- avoid spaces in filename

3.2 Reports

- "knit" button calls functions from `knitr` package
- pdf generation requires LaTeX installation
- difference between "Preview" and "knit..." is preview does not run all chunks explicitly. Knit fails if code doesn't run.

3.3 Basics of Markdown

3.3.1 Basic Formatting

- **bold**
- *itallics*
- subscript^{x}
- superscript$_{y}$
- formatting as `code` with backticks
- lists (must have space between list header and first list item)

3.3.2 Headings and subheadings

- using different nubmers of "#"
- integration with RStudio interface

3.3.3 Equations

- Inline with "$". E.g. n = 5 vs. $n = 5$.

- Display equation with "$$"

$$n = 5$$

- 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

3.3.4 When Markdown doesn't work

- R Markdown is "portable". It works for html, pdf, Word.
- However, if your output is html, you can use html.
- If your output is pdf, you can use LaTeX
- If your output is Word, just edit it in Word
- R Markdown is limited, but you probably won't need super advanced formatting
- Getting help:
    - *Help > Markdown Quick Reference*
    - *Help > Cheatsheets > R Markdown Cheatsheet*
    - *Help > Cheatsheets > R Markdown Reference Guide*

3.4   Code chunks

3.4.1   Inserting and naming code chunks

- Insert with menu, keyboard shortcut, or just typing
- naming code chunks and integration with R Studio (**Duplicate names not allowed**)
- Inline code with '#̊code '

3.4.2   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)
```
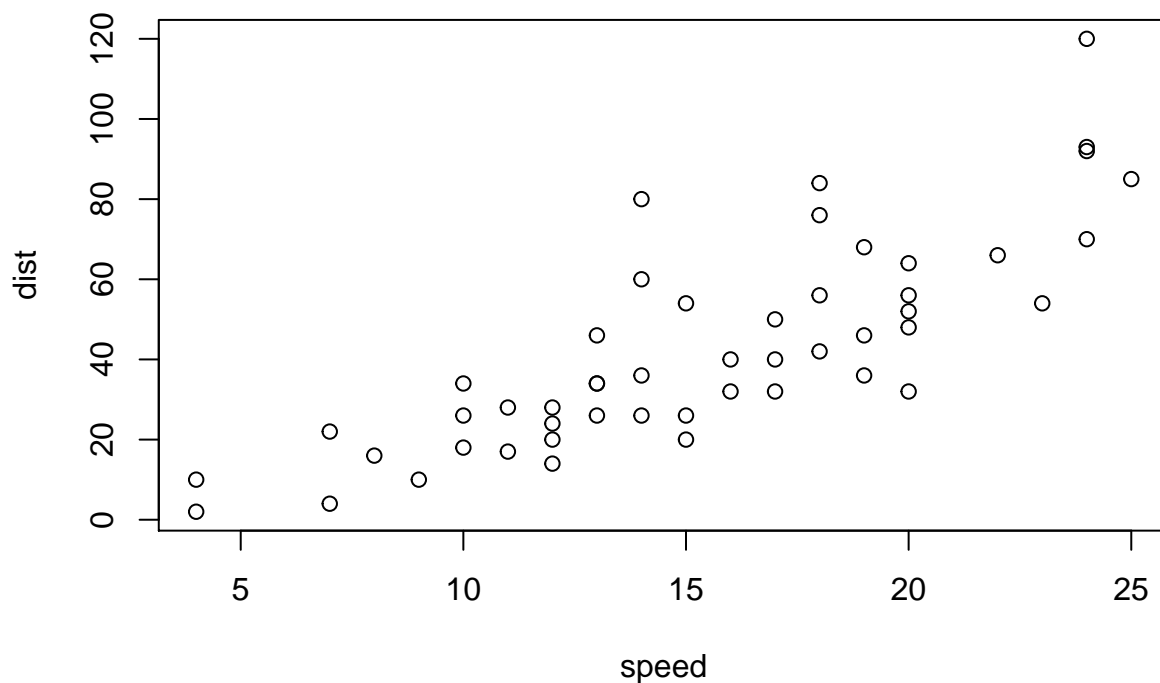


```
plot(trees)
```
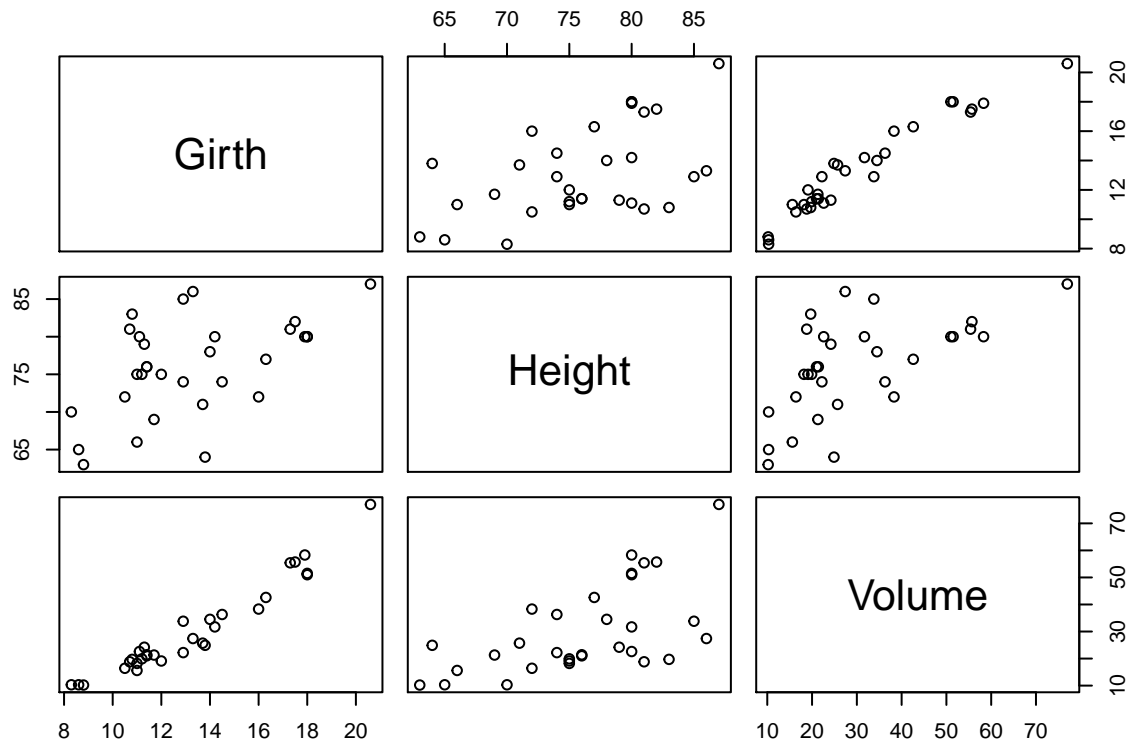
```
cor(trees)
```

```
##              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.

3.4.3 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.
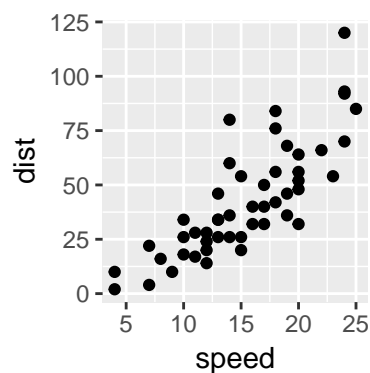
You can specify figure size here as well. *Note*: In my experience, base R plots tend to have very large margins and the end up a lot smaller than you'd expect based on setting these options compared to ggplot versions.

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

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

dist

o

0

5   15   25
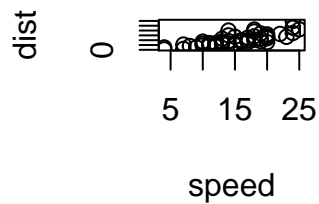
speed

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

125
100
75
dist
50
25
0

5   10   15   20   25
speed

3.4.4   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

Use `include_graphics()` to add already existing pictures. **NOTE:** figure sizes don't affect these. You have to use the `out.width` chunk option which can accept quoted values like `"2in"` or `"50%"`. Unfortunately, `out.width` doesn't affect in-line output or Preview output, only knit output.

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


3.5   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.

3.6   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 <- ...`)