Some tips with reproducibility

Clement Lee

2021-03-26

Introduction

- So far, **intro.Rmd** & **intro.md** are identical, and so are **template.Rmd** & **template.md**, because there is no R code chunks in the Rmd file. This allows the outputs to be generated directly via pandoc, e.g.:

```
pandoc -t html -o intro.html intro.md
pandoc -t html -o intro.html intro.Rmd # same as above
pandoc -t beamer -o template.pdf template.md
pandoc -t beamer -o template.pdf template.Rmd # same as above
```

Introduction

- Through the use of the R package `rmarkdown`, we can actually run some R code and use the results when generating the outputs dynamically.
- This removes the need of, for example, the following process:
  1. re-run a separate R script when you change your analysis;
  2. save the necessary plots and copy the numerical answers manually;
  3. include them in the tex file;
  4. repeat the above error-prone steps.

A simple example

- The R code chunk is wrapped in the usual way by two lines of triple backticks (‘) in the source file.
- In order to actually run the code, the top line of triple backticks is followed by `{r}` instead of `r` - see the source file:

```r
a <- rnorm(5) # 5 draws from the standard normal distribution
a
```

```
## [1] -0.08000814 -1.03577903  0.06524007  1.03070045 -1.14739007
```

- The results are now printed after the code upon output generation - see **INTRO.pdf** for the actual workflow of generating the outputs.

Reproducibility

- The nature of random number generator results in different answers generated every time. You are likely to get different numbers to mine in the previous slides.
- By setting the random seed, the same numbers are always obtained.

```
set.seed(1234) # this is the extra line compared to before
b <- rnorm(5)
b
```

```
## [1] -1.2070657  0.2774292  1.0844412 -2.3456977  0.4291247
```

- A tip: If you want the students to generate data for reproducible analysis but based on different data (so they don't copy answers from each other), one way is to use their own student ID as the seed.

Numerical result & inline code

- The objects created in code chunks are retained, so you can obtain, for example, the following:

```
sum(a)
```

```
## [1] -1.167237
```

- Results can also be obtained inline e.g. the sum of b's elements is -1.7617683. This is achieved by a pair of backticks (like in verbatim mode), with `r` following the first backtick, in the source file.
- The same applies to equations:

$$\sum_{i=1}^{5} b_i = -1.7617683$$

- Also look at how the date is generated at the top of the source file.

What if I want to hide the code and/or results?

- In the top line of triple backticks, chunk options can be specified to customise the code and the results. For example, the code can be hidden:

```
## [1] "The code is suppressed by echo = FALSE in source file"
```

- The results can be hidden while still evaluated:

```
sum(b) # results = FALSE in source file
```

- Or not evaluated at all:

```
sum(b) # eval = FALSE in source file
```

- The combination of chunk options makes code evaluation and reporting highly flexible. The authoritative reference for all chunk options is https://yihui.org/knitr/options/.

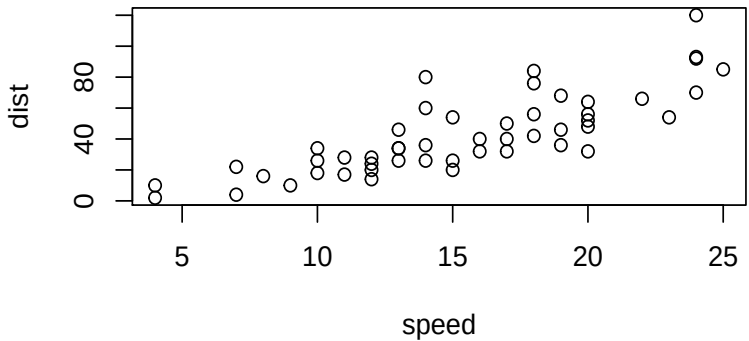
Plots

Plot from R directly

- When evaluating a standalone script, R would have plotted to a graphical device.
- By putting the relevant lines in a code chunk, the plot will be included directly in the output.
- The source code for the plot in the next slide is a simple example.
- The plot can be highly customised using the chunk options; see https://yihui.org/knitr/options/#plots.

Plot from R directly

```
data(cars)
plot(cars)
```



Existing images

- In **template.Rmd**, we use `![CAPTION](PATH/TO/IMG)` to
- Now, we use the function `knitr::include_graphics("PATH/TO/IMG")` in R code chunk to achieve the same (or better) - see next slide. The size is adjustable, while the function is still agnostic.
- In HTML output, the caption will also become the alternative text for accessibility. I haven't found a simple way of including tooltip text though.
- Again, if `PATH/TO/IMG` is an URL, generating a pdf will result in an error, while generating an HTML document will not.

Existing images alternative

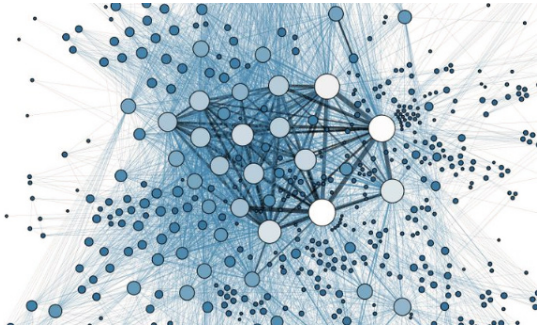


Figure 1: CAPTION HERE

- Observe the options in source file