Consequences of changing *this* for *that*

Jo Friend (Matriculation No. 53637383)
MSc Environmental Science
University of Freiburg, Germany

May 7, 2020

Contents

Abstract

A template to get started with writing a report weaving text and R-code.

1 Introduction

Reproducible statistical analysis means to make each step transparent. One approach weaves R-code into a LaTeX document, using the functionality of the **knitr** package by Yihui Xie. For help and documentation and examples see `https://yihui.org/knitr/`.

When you develop code, you trial a lot of code lines. Only some of them will make it into the final project document. It is thus common to work with two different R-scripts, one ending in .R, the pure R-code, and one ending in .Rnw, the project documentation. While you can do the .R in any R-editor, .Rnw are best processed in RStudio, which has a dedicated button to press (or Ctrl/Cmd-Shift-k) for compiling, which is a treat! I assume here that we use RStudio for both.

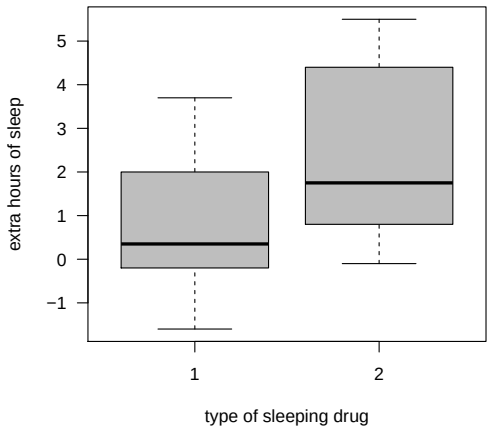
The R-code is evaluated in the console, while the project file is processed *independently* in the background. That can be confusing, as the objects in the console are *not* available to the document. It takes a bit of getting used to this parallel working. One option is to work only in R until, say, 2 hours before the end of the working day, then switch to the .Rnw-document and step-by-step copy-paste the relevant bits into it, and write the accompanying text.

2 Task 1

When using R-code in knitr-mode, one can provide a lot of options. For example, the options 'fig.height' and 'fig.width' refer to the size of a figure in R, while 'out.height' and 'out.width' are for the eventual size in the LaTeX document.

To use R within the document, you have to start and end with a special formatting. Start the chunk with <<>>= and end with @. The first entry is the name of the code chunk, if you want to name it. Like this:

```r
data(sleep)
attach(sleep)
boxplot(extra ~ group, col="grey", las=1, ylab="extra hours of sleep",
          xlab="type of sleeping drug")
```



There are many options when defining an R-chunk. Important options are:

- `eval=T/F` specifies whether the code is actually evaluate, i.e. computed and returned. If you only want to show but not evaluate the code, set it to false.
- `echo=T/F` specifies whether this code chunk is actually shown in the document. We will next see a case where you want to hide the code.
- `message=T/F` suppresses messages, such as when loading a package or attaching (and thereby overwriting) a table repeatedly.
- `cache=T/F` tells R to store the results of a code chunk (or not). This is rather important! If you were to re-evaluate each chunk every time you add code, you would have to wait endlessly for the results. Hence, R stores the results of each chunk. But now imagine you change something in chunk 12 but chunk 14 would need to be re-evaluated, as it depends on chunk 12. This would not happen, by default, as each chunk is stored without reference to which chunk it depends on. By setting for chunk 14 `cache=F`, we can force R to evaluate a chunk every time. Use prudently!

Normally, the results of an R-chunk are stored. Thus, when you write new text, and new chunks, there should be no need to think of anything, just type. Somehow, this is not always the case.

Imagine running some code takes 5 hours. We developed the code in the .R-file and don't want to re-run the entire 5 hours again after copy-pasting it into the .Rnw-file. Instead, we would rather store the results from the console, and, when compiling the document, **not** evaluate this chunk, and instead load the object saved.

This is how to do it:

1. During your code development phase, save the file into the same folder as your document (or preferably into a subfolder, e.g. "storedObjects"): `save(myobject,file="storedObjects/myobject.Rdata")`
2. In your document, copy-paste the R-code used to generate the object into an R-chunk, but set the option `eval=F`. This chunk is now ignored in R, but set appropriately in LaTeX.
3. Add a new chunk to load the stored object, setting the option `echo=F` to suppress it being rendered in the document: `load("storedObjects/myobject.Rdata")`. This will make `myobject` available to this and further chunks.

So, in the .R-file, we created some object `fm` which we then stored in the working directory as `sleepfm.Rdata`. This is the code how to proceed in the .Rnw-document (obviously, you cannot see this in the PDF, only in the .Rnw file!!):

```
fm <- lm(extra ~ group, data=sleep)
summary(fm)

##
## Call:
## lm(formula = extra ~ group, data = sleep)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.430 -1.305 -0.580  1.455  3.170
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.7500     0.6004   1.249   0.2276
## group2        1.5800     0.8491   1.861   0.0792 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.899 on 18 degrees of freedom
## Multiple R-squared:  0.1613,Adjusted R-squared:  0.1147
## F-statistic: 3.463 on 1 and 18 DF,  p-value: 0.07919
```

None of these complications is visible to the reader of the final document, apart from an empty white line in the text. I have no idea how to remove that, except with this dirty negative-vertical-space hack.

3 Conclusions

4 References

Normally, we would have a .bib file with the references, and we would compile that using BibTeX. In this case, we would have only very few references and it is probably simpler to copy-paste them into here. For this purpose, we can use the `hangparas`-environment:

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davis, R., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of atlas data: a review. *Ecography* 30: 609–628

Dormann, C.F. (2020) Calibrating probability predictions from machine-learning and statistical models. *Global Ecology & Biogeography* 27, 760–765

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Severin Hauenstein, Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. & Dormann, C.F. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 913–929