Making research reproducible using literate programming

On programming a computer, we should concentrate on explaining to humans, what we want the computer to do – instead on focusing on instructing the computer (Knuth, 1984). This idea, termed literal programming, has not only gained widespread currency in software development communities, but harbors great benefit for science as well. In particular, literate programming can be thought of as an necessary condition for reproducible science. Reproducibility is an umbrella term referring to the degree that independent researchers are able to confirm the conclusions of some piece of research (Peng, 2015). The current "reproducibility crisis" (Open Science Collaboration, 2015) is partly due to the fact that is often hard or even impossible at all to reproduce the results of an (empirical) research article. Not only data and code must be available as a minium requisite for reproducibility, but it must also be possible to easily connect some piece of code for data analysis, its results, and the respective figure in the research manuscript. However, often enough researchers separate these three components in different documents, thereby making a reproduction difficult. In this poster, we demonstrate a simple "recipe" for producing a reproducible research report based on literate programming. In this approach, data, code and text will be processed in the same file. The final paper is distilled out of this combined file. Thus, reproducing as well as updating the paper (e.g., in case of additional data) is straight-forward. We explain the use of well-established tools for that purpose. Specifically, we show how to mix data, R code, text, and text formating using Latex in one document using a *noweb* dialect called *knitr* (Xie, 2013).

REFERENCES

Knuth, D. E. (1984). Literate Programming. *The Computer Journal*, *27*(2), 97–111. http://doi.org/10.1145/2351676.2351679

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). http://doi.org/10.1126/science.aac4716

Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance*, *12*(3), 30–32. http://doi.org/10.1111/j.1740-9713.2015.00827.x

Xie, Y. (2013). *Dynamic Documents with R and knitr* (Vol. 29). CRC Press.