

About Writing Dynamic Documents with R

Author Name¹

¹*Department of Geography, University of Zurich, Winterthurerstrasse 190, Zurich
name@geo.uzh.ch*

Abstract

This is the abstract of the template document used to show how to write publications in R with R Markdown and the help of some packages. Based on a concrete usecase this document exemplifies some of the caveats that may occur when writing such document and publish it online on a GIT repository. It also presents typical usecases in Markdown usage and presents some tricks.

Introduction

This example publication is aimed to serve as a motivation on how to create reproducible documents in R and to advocate in general reproducible research.

State of the Art

Various authors in qualitative and quantitative research argue for that as many parts of the research workflow reproducible. Brunsdon (2015) state “Reproducible quantitative research is research that has been documented sufficiently rigorously that a third party can replicate any quantitative results that arise”.

To further motivate you, read (Healy 2016, LeVeque et al. (2012), Baker (2016), Editorial (2016), Pebesma et al. (2012), Vandewalle (2012), Nüst et al. (2011), Buckheit and Donoho (1995), Healy (2011)) or the short and to the point editorial of Editorial (2016).

Case Study - Parc Adula

This case study presents a small subset of a current study conducted at the Department of Geography at the University of Zurich. The study investigates the development of a second Swiss National Park - the *Adula Parc*.

Exploratory topic analysis

For this case study 16 interviews have been carried out. Each of these guided interviews got annotated based on a predefined topic tree. The following plots displays a sample of an output from MXAQDA as software for qualitative data analysis.

Overview on the held interviews and representants:

- *Cantonal Government* (n: 4): Representants of four different involved departments
- *Environmental Organisation* (n: 1): Involved as a stakeholder in the park planning
- *Federal Government* (n: 2): Responsible that the parc follows regulations of ‘Natur- und Heimatschutz’
- *Local* (n: 5): Local representants of the park region
- *Parc Team* (n: 2): Team member involved in the park planning
- *Tourism* (n: 2): Local tourism representants

The following plot presents the frequency of occurrence of a select list of topics that occurred in the interviews. While there seem to be spent more focus on the *Pro Argument* against the *Contra Argument* during the interviews. It is interesting to see that comparing topics on *Tourismus* have far more weight than those on *Biodiversität*.

Table 1: Topic mentions.

Code	Mention
Biodiversitaet	5
Contra Argument	39
Pro Argument	68
Tourismus allgemein	48

Figure presents the frequencies matrix of the topics occurrences across the different interviews. It provides an overview on where and by which representant topics occur.

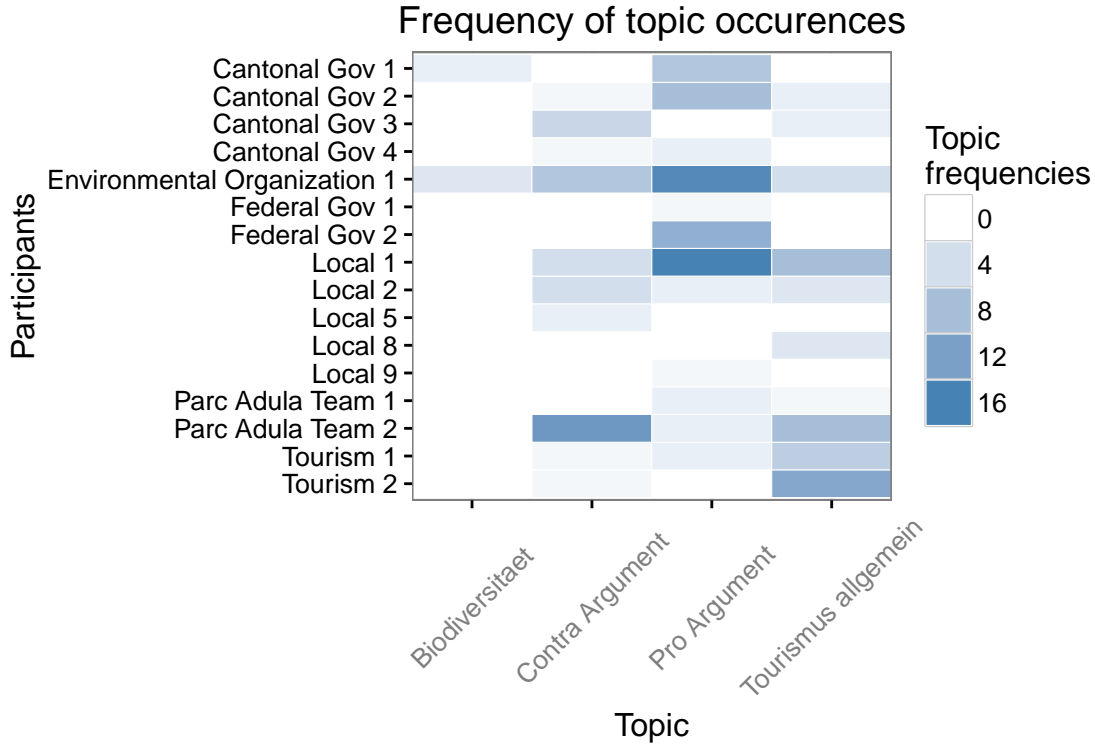


Figure 1: Frequency matrix of a selected list of topics across the various representants

Notes on reproducibility: Depending on the data to analyze privacy plays a role. While for the analysis itself the data is being anonymised, storing the raw or preprocessed data on a public repository may poses issues regarding the privacy of the data.

Google query timeline

Overview on the Google trend evolution on the search query: *Parc Adula* (url, provides a CSV file). The timeline shows overall a small amount of queries for this word combination, with a spike on 2015-11-01. The has been retrieved on August 11, 2016.

Notes on reproducibility: Due to license restrictions of the open geodata it is not possible to store the data on a public Git repository. The included script `R/loadMapData.r` downloads the data directly from the link provided in the geodata catalog infobox of <http://maps.geo.admin.ch>

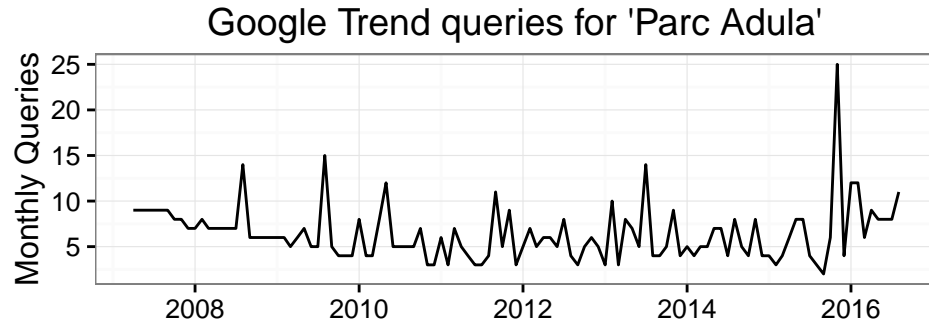


Figure 2: Timeline of queries for Parc Adula set in the Google search engine



Figure 3: Planned perimeter of Parc Adula, Switzerland, Data source: Swisstopo

Case study area

Parc Adula is situated in Switzerland in the border region of the cantons Ticino and Grisons. The map below presents the current outer perimeter of the planned national parc.

Notes on reproducibility: Due to license restrictions of the open geodata it is not possible to store the data on a public Git repository. The included script `R/loadMapData.r` downloads the data directly from the link provided in the geodata catalog infobox of <http://maps.geo.admin.ch>

Discussion & Conclusions

This template based on data of an ongoing presents some typical examples maybe used in a publication written in RMarkdown. It presents the inclusion of data and analysis, features plots, tables, and various markdown elements and shows how to integrate literature. The generated files in *PDF*, *Word* or *HTML* often still need fine some fine-tuning afterwards (particularly in Latex). However, it still presents a great way documenting the research process, that is easily shareable and the generation of the initial drafts.

Acknowledgements

The Reproducible Research workshop was supported by the InnoPool of the Department of Geography, University of Zurich.

References

- Baker M, 2016, 1,500 scientists lift the lid on reproducibility. *Nature*. 533(7604):452–454.
- Brunsdon C, 2015, Quantitative methods I: Reproducible research and quantitative geography. *Progress in Human Geography*. doi: 10.1177/0309132515599625
- Buckheit J, Donoho D, 1995, WaveLab and Reproducible Research. *Wavelets and Statistics*. 10355–81.
- Editorial, 2016, Reality check on reproducibility. *Nature*. 533(7604):437–437.
- Healy K, 2016, The Plain Person’s Guide to Plain Text Social Science. Healy2016
- Healy K, 2011, Choosing Your Workflow Applications. *The Political Methodologist*. 18(2):9–18.
- LeVeque RJ, Mitchell IM, Stodden V, 2012, Reproducible research for scientific computing: Tools and strategies for changing the culture. *Computing in Science & Engineering*. 14(4):13–17.
- Nüst D, Stasch C, Pebesma E, 2011, Connecting R to the sensor Web. In: *Lecture notes in geoinformation and cartography*. 227–246
- Pebesma E, Nüst D, Bivand R, 2012, The R software environment in reproducible geoscientific research. *Eos, Transactions American Geophysical Union*. 93(16):163–163.
- Vandewalle P, 2012, Code Sharing Is Associated with Research Impact in Image Processing. *Computing in Science & Engineering*. 14(4):42–47.