How to automate analysis workflow via reproducible reports (using an R ‘template’ package) for recurrent standardized data: a case study

Kirsten Van Hoorde (OpenAnalytics)

European Bioconductor Meeting 2017, December 5-6, 2017

Motivation and goal

Creation of data analysis reports resulted often in **looking up, copy-pasting and re-use of code from previous analyses**: a **time-consuming** approach which is **prone to errors**.

So look for an **alternative** that:

- can deal with recurrent analyses,
- results in a standardized/reproducible analysis,
- is less prone to errors,
- saves time,
- focuses on the important part (i.e. experiment/analysis-specific).

Automate *parts of an analysis workflow (e.g. new type of data) to* ***focus*** *on only analysis-specific part (e.g. experiment).*

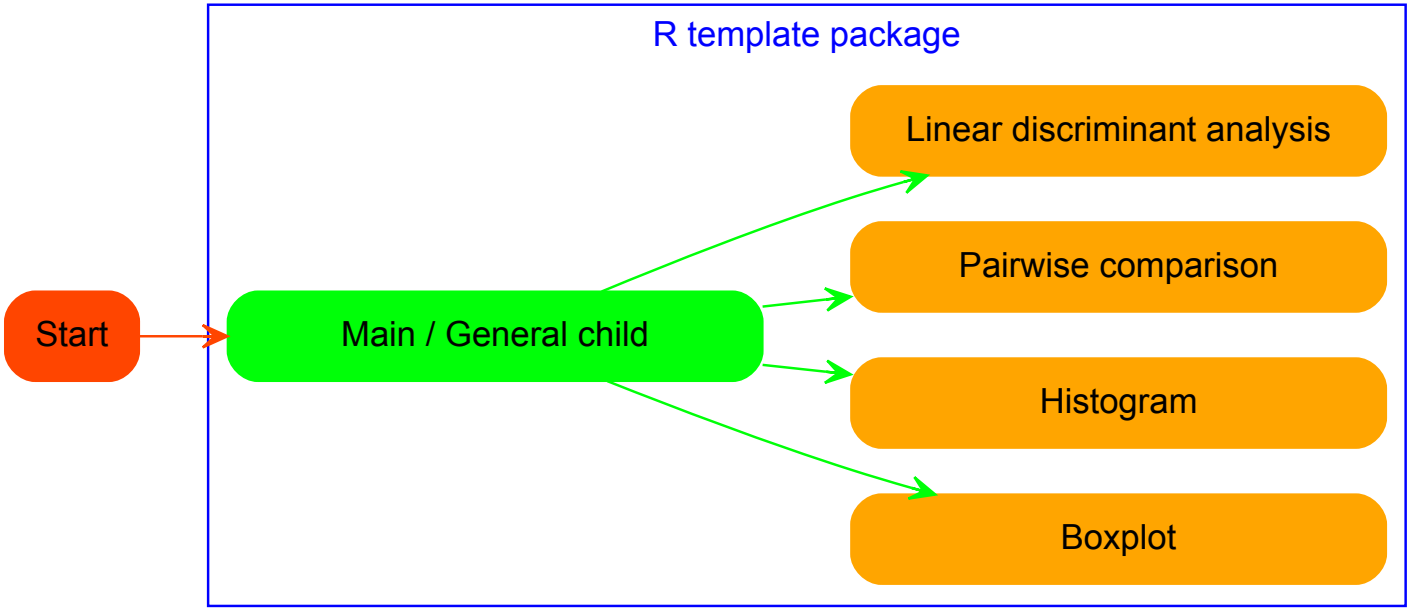
How to start?

- Distinguish the **data(experiment)-specific** (e.g. data, different models, visualization parameters) from the **frequently used analysis** parts.
- Capture the **'consistent' code** in R functions or the equivalent R 'template' documents for reporting.

Our (biological) case studies

- **Data types**
  - Microarray analysis
  - T-cell receptor (TCR) analysis
  - Quantitative polymerase chain reaction (qPCR)
  - …
- **Reporting**
  - Quality control (QC)
  - Modelling
  - Clustering analysis
  - Visualization
  - …

Suggested approach



Analysis-specific parameters - Main structure + calls - Specific (sub-)analyses of interest

R 'template' package

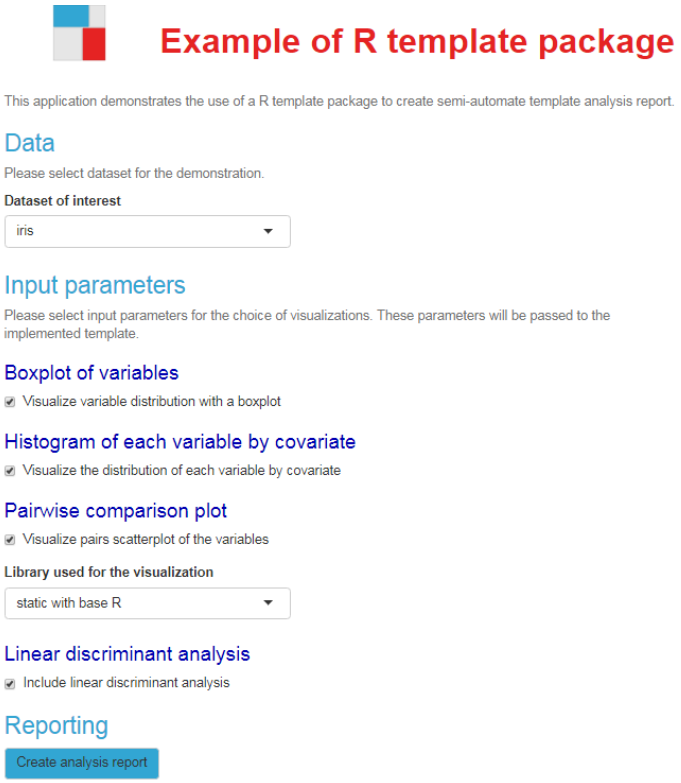
Advantages

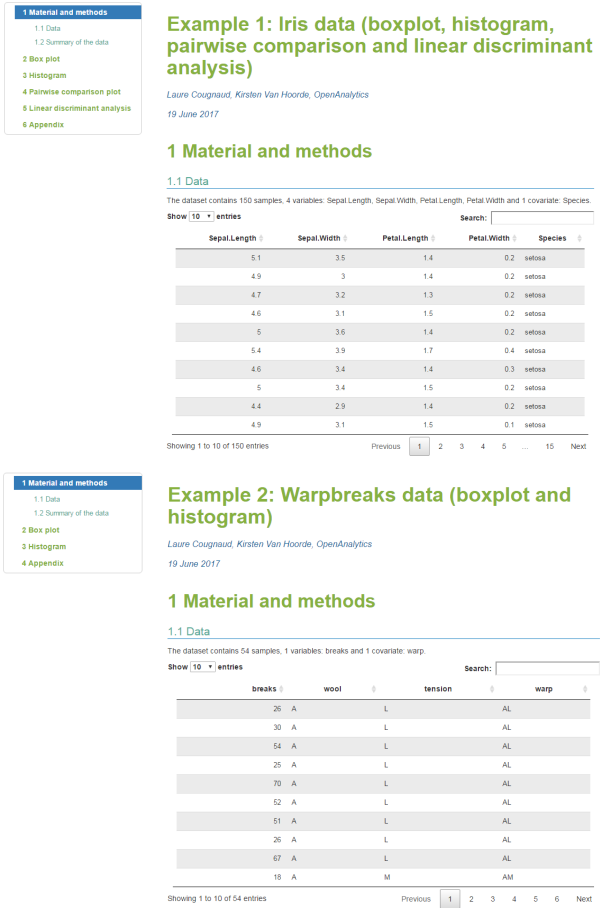
- **Less prone to errors**
- **Time saving**
- **Consistency** across reports
- **Reproducibility** of the entire workflow via versioning
- Easily **adjusting code** (correct bugs - extend analysis - …)
- Opportunities via a **shiny application**

Worthwhile the effort?

A lot of time and analysis reproducibility can be gained when using this package afterwards.

Example R 'template' package





Summary

All the tools are already available thanks to the open R community, so do-it-yourself!

Example R 'template' package (with shiny application example)
https://github.com/openanalytics/useR2017_templatePackageExample

Contact

- laure.cougnaud@openanalytics.eu
- kirsten.vanhoorde@openanalytics.eu

Useful links

- Child documents: knitr and Sweave
- Parametrized and interactive documents