



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

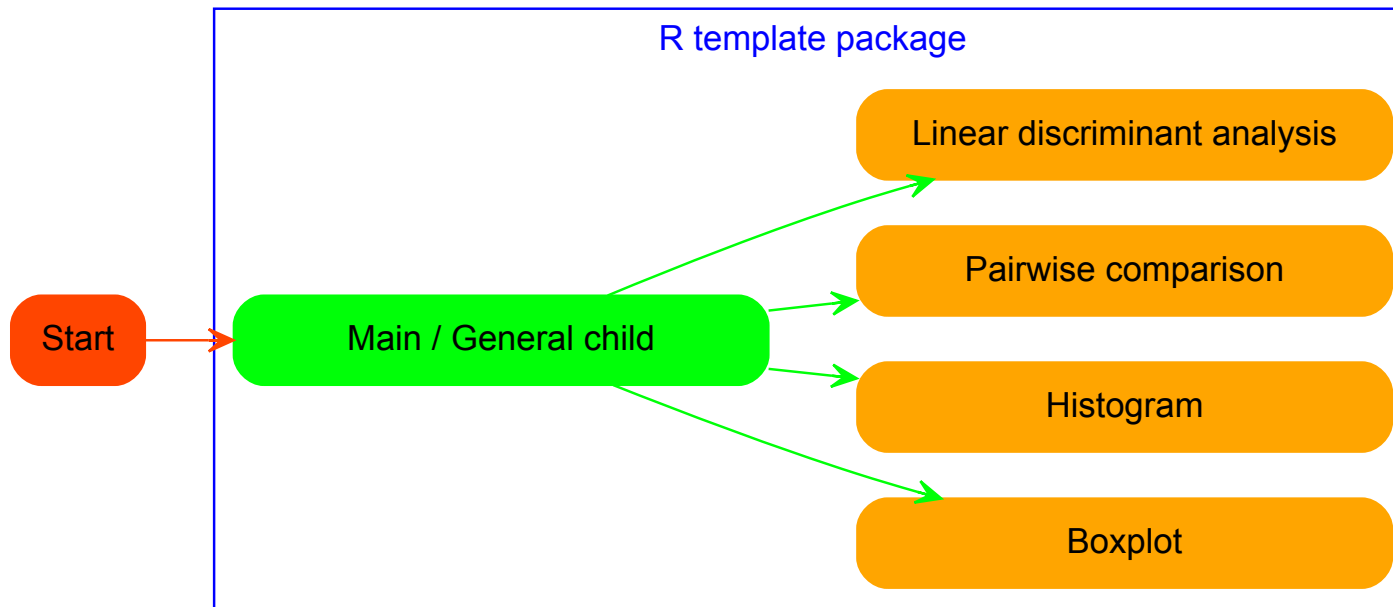
- **Data types**

- 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



Example of R template package

This application demonstrates the use of a R template package to create semi-automate template analysis report.

Data

Please select dataset for the demonstration.

Dataset of interest

iris

Input parameters

Please select input parameters for the choice of visualizations. These parameters will be passed to the implemented template.

Boxplot of variables

☒ Visualize variable distribution with a boxplot

Histogram of each variable by covariate

☒ Visualize the distribution of each variable by covariate

Pairwise comparison plot

☒ Visualize pairs scatterplot of the variables

Library used for the visualization

static with base R

Linear discriminant analysis

☒ Include linear discriminant analysis

Reporting

Create analysis report

1 Material and methods
1.1 Data
1.2 Summary of the data
2 Box plot
3 Histogram
4 Pairwise comparison plot
5 Linear discriminant analysis
6 Appendix

Example 1: Iris data (boxplot, histogram, pairwise comparison and linear discriminant analysis)

Laure Cougnaud, Kirsten Van Hoorde, OpenAnalytics

19 June 2017

1 Material and methods

1.1 Data

The dataset contains 150 samples, 4 variables: Sepal.Length, Sepal.Width, Petal.Length and 1 covariate: Species.

Show 10 ▾	entries	Search:								
Sepal.Length ▾	Sepal.Width ▾	Petal.Length ▾	Petal.Width ▾	Species ▾						
5.1	3.5	1.4	0.2	setosa						
4.9	3	1.4	0.2	setosa						
4.7	3.2	1.3	0.2	setosa						
4.6	3.1	1.5	0.2	setosa						
5	3.6	1.4	0.2	setosa						
5.4	3.9	1.7	0.4	setosa						
4.6	3.4	1.4	0.3	setosa						
5	3.4	1.5	0.2	setosa						
4.4	2.9	1.4	0.2	setosa						
4.9	3.1	1.5	0.1	setosa						
Showing 1 to 10 of 150 entries		Previous	1	2	3	4	5	...	15	Next

Showing 1 to 10 of 150 entries

1 Material and methods
1.1 Data
1.2 Summary of the data
2 Box plot
3 Histogram
4 Appendix

Example 2: Warpbreaks data (boxplot and histogram)

Laure Cougnaud, Kirsten Van Hoorde, OpenAnalytics

19 June 2017

1 Material and methods

1.1 Data

The dataset contains 54 samples, 1 variables: breaks and 1 covariate: warp.

Show 10 entries

Search:

breaks	wool	tension	warp
26	A	L	AL
30	A	L	AL
54	A	L	AL
25	A	L	AL
70	A	L	AL
52	A	L	AL
51	A	L	AL
26	A	L	AL
67	A	L	AL
18	A	M	AM

Showing 1 to 10 of 54 entries

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