GRINTAR: A DEMONSTRATION OF REPRODUCIBLE ANALYSIS, VISUALIZATION AND DISTRIBUTION OF ERGOMETER EXERCISE DATA

UNIVERSITY OF APPLIED SCIENCES UTRECHT

INNOVATIVE TESTING RESEARCH GROUP
INSTITUTE FOR LIFE SCIENCES & CHEMISTRY

HU UNIVERSITY OF APPLIED SCIENCES UTRECHT
THE NETHERLANDS

Marc Teunis, PhD[1]
Jan-Willem Lankhaar, PhD[2,3]
Shirley Kartaram, MSc[1]
Eric Schoen, PhD[4]
Raymond Pieters, PhD[1,5]

[1] Innovative Testing Research Group
HU University of Applied Sciences Utrecht

[2] Digital Smart Services Research Group
HU University of Applied Sciences Utrecht

[3] Institute for ICT
HU University of Applied Sciences Utrecht

[4] TNO
Zeist, The Netherlands

[5] Institute for Risk Assessment Sciences (IRAS)
Utrecht University

E-mail marc.teunis@hu.nl

Reproducibility of research results is essential for the progress of science, but it often does not keep pace with the explosion of analytical technology. Most analytical tools do provide means to stimulate reproducible analyses (e.g. versioning), but they are often put into practice ineffectively. As a result, scientific outcomes are often poorly reproducible.

The **grintar** *R* package demonstrates how complying with a number of principles during analysis can greatly improve reproducibility. The **grintar** package contains the raw and processed data and analysis and visualization methods from GRINTA!, a recent ergometer exercise study. For preprocessing and analysis, Ridge's *guerilla analytics approach* (Ridge 2014) was followed. All data preprocessing, analysis and visualization steps were automated and stored in the **grintar** package, as well as the raw and resulting data sets. For coding style, Wickham's style (Wickham 2014) was followed. In addition, the package was fully documented.

After scientific publication, the **grintar** package will be published on GitHub to be used for educational and scientific purposes.

The **grintar** package demonstrates that the guerilla analytics approach provides practical and useful guidelines for reproducible data analysis.

References

Ridge, Enda. 2014. *Guerrilla Analytics: A practical approach to working with data.* Morgan Kaufmann.

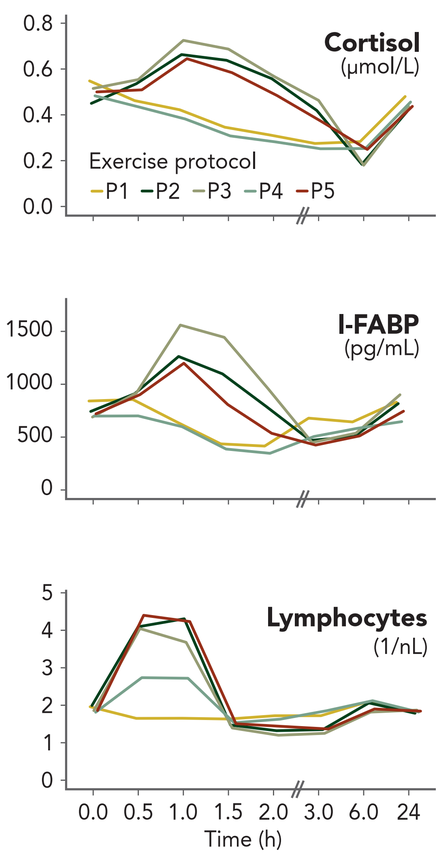
Wickham, Hadley. 2014. *Advanced R.* CRC Press.

MORE ABOUT...

- **The study** » Box ***Grinta! An ergometer exercise study***
- **Guerilla analytics** » Box ***The 7 Principles of guerilla analytics***
- **The package** » Box ***Grintar R package***
- **Guerilla analytics in practice** » Figure ***Data loading illustrated***

GET THE PACKAGE

github.com/uashogeschoolutrecht/grintar



Response of cortisol, intestinal fatty acid binding protein (I-FABP) and lymphocyte concentrations to one hour ergometer exercise tests at varying levels of intensity and in different hydration conditions.

GRINTA! AN ERGOMETER EXERCISE STUDY

Background Bicycle ergometer exercise testing may serve as a human stress model to investigate nutritional claims.

Aim To identify biomarkers for intestinal function and immune response after ergometer exercise testing.

Methods Healthy volunteers (*N* = 15) did one hour bicycle ergometer exercise tests of different intensities and in different hydration conditions. Serum, urine and saliva samples were collected prior to, during and after cycling until 24 h after exercise. About 100 biomarkers of intestinal function, immune response and general physiology were determined in 6 different laboratories. Data were analyzed with a multilevel mixed linear model.

Results Immune response (e.g. NK cells, neutrophils), intestinal function (e.g. intestinal fatty acid binding protein) and general physiology (e.g. cortisol) peak between 0 and 6 h. Kinetic changes were observed at high as well as low exercise intensity.

Conclusion Bicycle ergometer testing can serve as a human stress model to investigate nutritional claims even in persons who are able to exercise at low intensity only.

(publication in preparation)

THE 7 PRINCIPLES OF GUERILLA ANALYTICS

Guerilla analytics offers a practical approach to working with data. It is based on 7 principles:

1. **Clarity** - Space is cheap, confusion is expensive.
2. **Simplicity** - Prefer simple, visual project structures over heavily documented and project-specific rules.
3. **Automation** - Prefer automation with program code over manual graphical methods.
4. **Data provenance** - Maintain a link between data in the file system, data in the analytics environment and data in work products.
5. **Version control** - Version control changes to data and program code.
6. **Knowledge consolidation** - Consolidate team knowledge in version-controlled builds.
7. **Integrity of runs** - Prefer analytics code that runs from start to end.

Guerilla analytics provides practical tips for

- Version control
- Testing
- Workflow management
- Communication

GRINTAR R PACKAGE

A sample of the items available in the package.

Workflow scripts
Scripts that perform tasks from the Guerrila Analytics workflow

`grintar::workflow_load("dataset")`
Loads raw data into the data manipulation environment (DME).

`grintar::workflow_clean_dme("path_dme")`
Deletes existing files from the DME.

`grintar::workflow_convert_xlsx2csv_addrowid()`
Converts a Microsoft Excel file to a comma-seperated value (CSV) and adds a column with row IDs.

Package utility functions
General utility functions

Plots

`grintar::plot_panels(first_author, figure_number)`
Recreates panel graphs from publications.

`grintar::plot_heatmap(ml_grinta, contrasts, param)`
Creates a heatmap of statistical nlme analysis; which contrasts to plot is customizable.

Analysis demos

`grintar::analyze_pca()`
Interactive tutorial on principal component analysis with the GRINTA! data.

`grintar::analyze_nlme()`
An interactive tutorial on non-linear mixed effects models on the GRINTA! data.

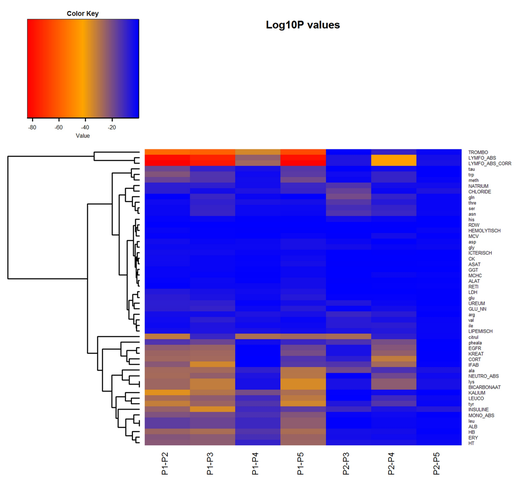
`grintar::analyze_rf()`
An interactive tutorial on Random Forest Decision Tree regression analysis on the GRINTA! data.

Data sets
Data sets available in the package namespace

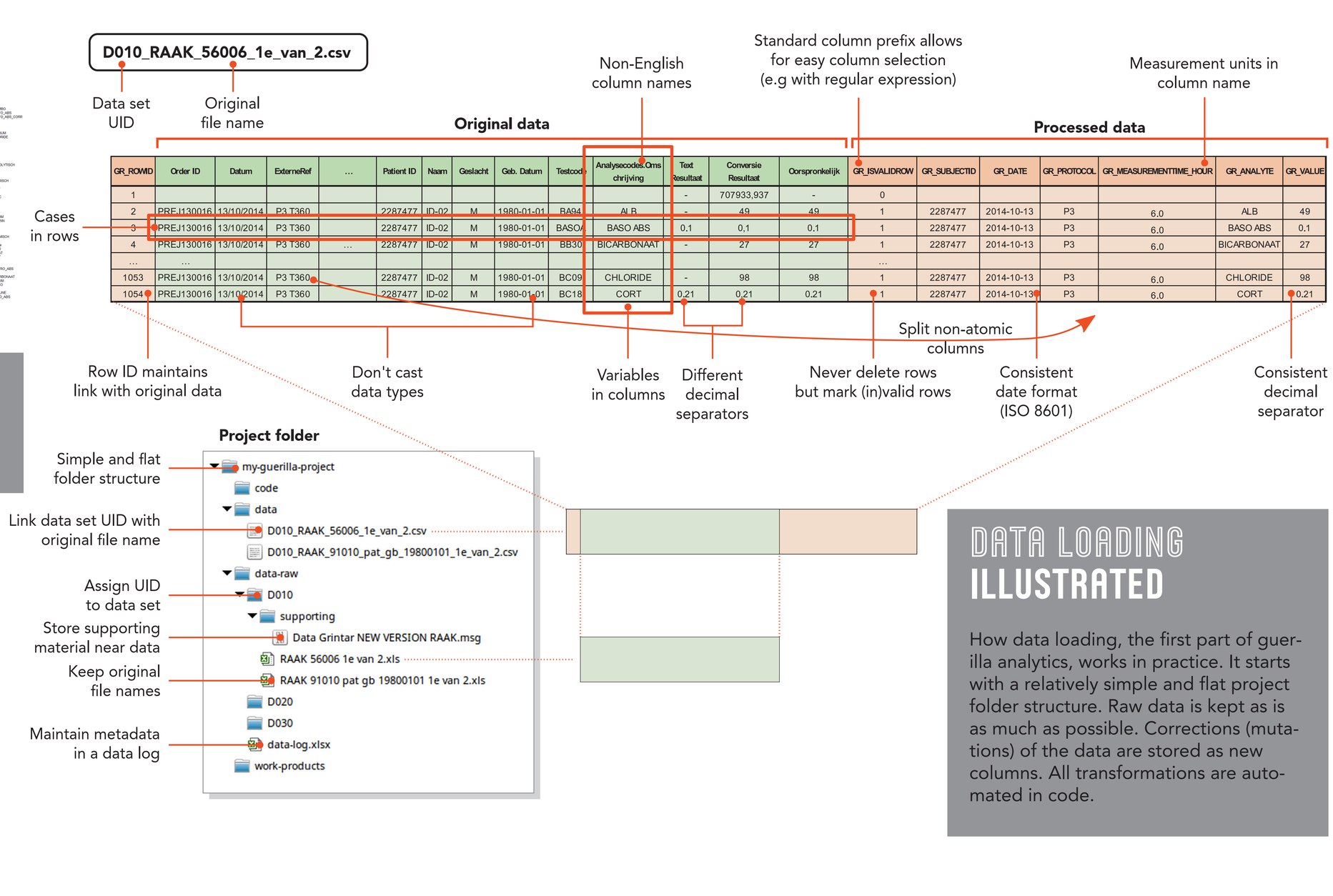
`grintar::data("messy_grinta")`
Illustrate the steps to a tidy version of the data.

`grintar::data("tidy_grinta")`
The tidy version of the messy_grinta dataframe.

`grintar::data("multilevel_grinta")`
Results from a mixed effects models analysis (p and F values for each possible contrast in the experiment, and for all biomarkers determined).



Contrast comparison of different exercise protocols (columns) resulting from ANOVA applied to biomarkers (rows). Clusters show biomarkers that differ between different protocols.



DATA LOADING ILLUSTRATED

How data loading, the first part of guerilla analytics, works in practice. It starts with a relatively simple and flat project folder structure. Raw data is kept as is as much as possible. Corrections (mutations) of the data are stored as new columns. All transformations are automated in code.