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] IRAS
Utrecht University

E-mail marc.teunis@hu.nl

DIAGRAMS

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, 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 (see *GRINTA! An ergometer exercise study*). For preprocessing and analysis, Ridge's *guerilla analytics approach* (Ridge 2014) was followed (see *The 7 principles of guerilla analytics*). For an example of what guerilla analytics means in practice, see *Data loading illustrated*.

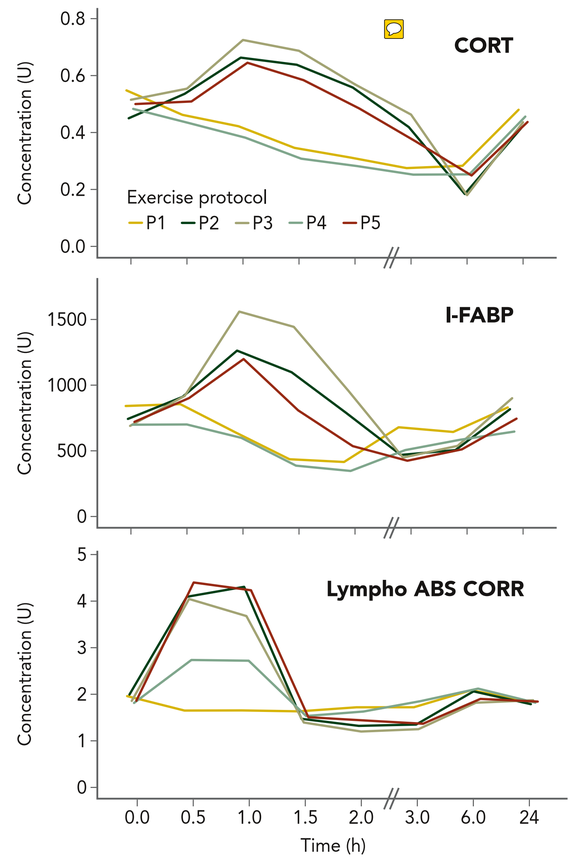
All data preprocessing, analysis and visualization steps were automated and stored in the **grintar** package, as well as the raw and resulting data sets. For coding style, Wickham's style (Wickham 2014) was followed. In addition, the package was fully documented.

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
Wickham
GitHub



Recovery after exercise according to five different exercise protocols. A representative parameter is shown for serum (<matrix1>), urine (<matrix2>) and saliva (<matrix3>).
{Highlights figuren en toelichting protocollen?}

GRINTA! AN ERGOMETER EXERCISE STUDY

Aim Identify biomarkers for recovery after exercise.

Methods Healthy volunteers (N = 15) underwent different cycle ergometer exercise regimens. Serum, urine and saliva samples were collected at baseline, during and after cycling until 24 hours after exercise. From the samples about 100 biological parameters were determined. Samples were analyzed in 6 different laboratories.

Results

(publication in preparation)

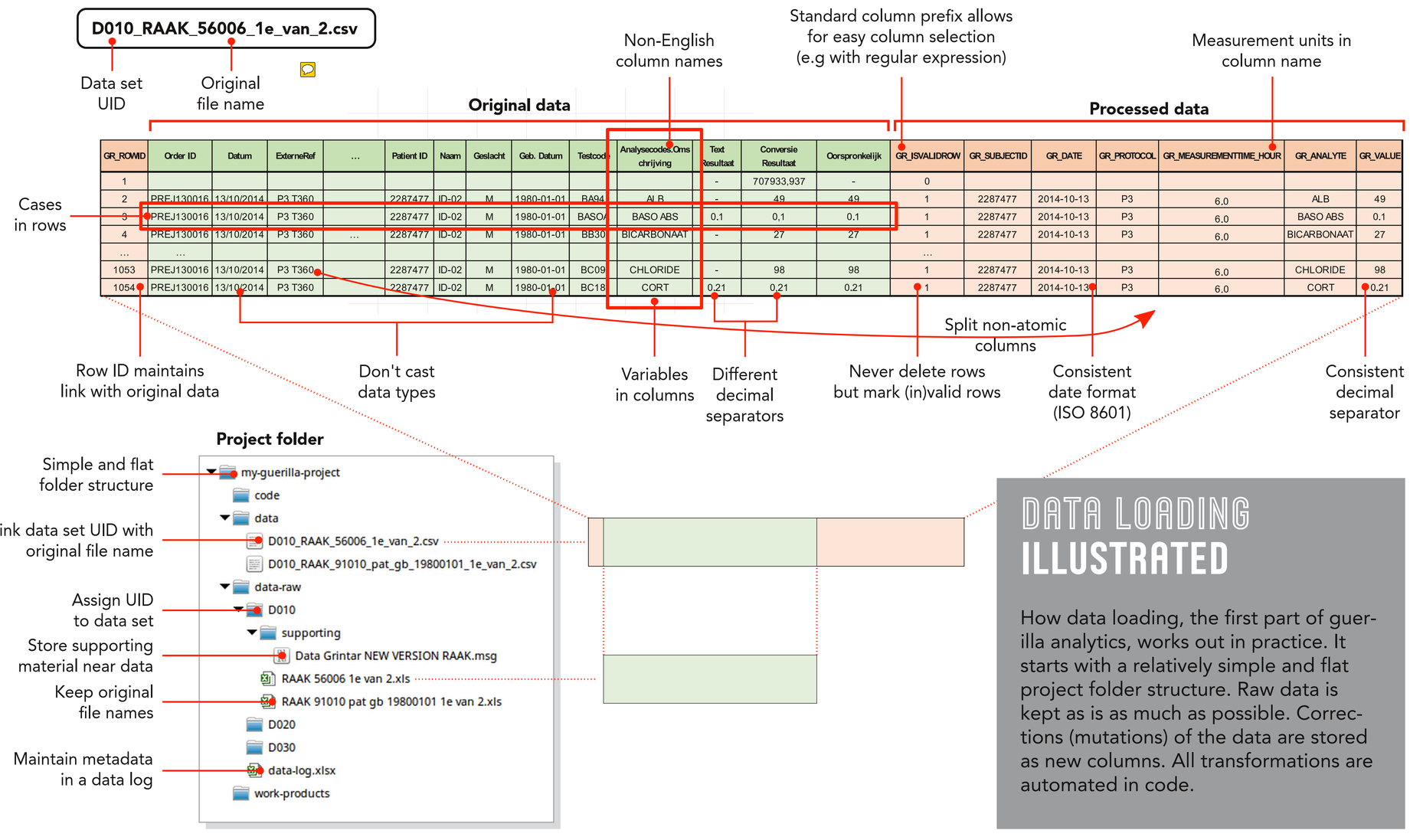
THE 7 PRINCIPLES OF GUERILLA ANALYTICS

Guerilla analytics{ref} 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.

GRINTAR R PACKAGE

Text Marc



DATA LOADING ILLUSTRATED

How data loading, the first part of guerilla analytics, works out 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.