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

Marc Teunis, PhD1
Jan-Willem Lankhaar, PhD1,2
Shirley Kartaram, MSC1
Eric Schoen, PhD1
Raymond Pieters, PhD1,3
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 NWO
Zeno, 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 analytic-al technologies. Most analytical tools do provide means to stimulate reproducible analysis (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 also 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

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

More about ...

- The study: Box Grintar: 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/auwshoegausschouten/grintar

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.