The Rsquared Platform – Reproducible Research for the R Community

Manuel J. A. Eugster

Institut für Statistik Ludwig-Maximiliams-Universität München

BioMed-S Referat, Schloss Höhenried, Bernau, 2011

Joint work with Friedrich Leisch and Torsten Hothorn:

- Leisch, Eugster, and Hothorn. *Executable papers for the R community: The R² platform for reproducible research.* Procedia Computer Science, 4:618–626, 2011. Proceedings of the International Conference on Computational Science, ICCS 2011.
- Finalist of the Executable Paper Grand Challenge.

THE VIEWS EXPRESSED IN THIS TALK DO NOT NECESSARILY REPRESENT THE VIEWS OF MY COLLABORATORS!

THE VIEWS EXPRESSED IN THIS TALK DO NOT NECES CARILY REPRESENT THE VIEW CONTACT OF COLLABORATORS!

Just in case I lose it ...

UNDEF

teruallo quadratorum, & Canones iidein hîc etiam locum habebunt , vt manifeflum eft.

QVÆSTIO VIII.

TROPOSITVM quadratum diuidere in duos quadratos. Imperatum fit vt 16. diuidatur in duos quadratos. Ponatúr primus 1 Q. Oportet igitur 16 -1 Q. æquales effe quadrato. Fingo quadratum à numeris quotquot libuerit, cum defe-Au tot vnitatum quot continet latus ipfius 16. efto à 2 N. - 4. ipfe igitur quadratus crit 4 Q. + 16. - 16 N. hac aquabuntur vnitatibus 16 - 1 Q. Communisadiiciatur vtrimque defectus, & à similibus auferantur similia, fient 5 Q. aquales 16 N. & fit 1 N. # Eritigitnr alter quadratorum #.alter verò 2. & vtriulque fumma eft " feu 16. & vterque quadratus eft.

TON Jin Caze cu Ca ne Faywror SEREN eis duo TETRAJWIOUS. E-חודר (2 שו או דיוה אוופוע פיב לני דב-זרמן שיוטער. אמן דער לאשע ל הרפי חב Sunamews mas. School a ege pora-Sas 15 reif Sura news mas loas ED refalwin. Tracosw + refazuvor Soro 55. Sow on more reind tooswy no bow ostrifis no madeed. isw ss B rend u S. aunos age o repayavos ésay duranear S 4" 15 [rei 4 55 15] Tauta loa Movan 15 rei A Swapews mas. nown wegonelow i reifes, i boo όμοίων όμοια. δωναμεις αρα ε ίσαι acconsist. if strend a Seronos

8٢

3 / 15

Arithmeticorum Lib. II.

85

teruallo quadratorum, & Canones iidem hîc etiam locum habebunt, vi manifeflum eft.

QVÆSTIO VIII.

```
quadratum
                            TON JAn Carden Ta Ta Lavor
TROPOSITVM.
  diuidere in duos quadratos.
In
in
pri
    In the sense of Fermat's Last Theorem:
~ 1
     "I have discovered a truly marvelous source code of
Fii
    this, which this paper is too narrow to contain."
qu
Au
nc
    (*) see Wikipedia (2011)
4
b
```

Communisadiiciatur vtrimque aege o re baywrog ëszu duwa u e wdefectus, & à fimilibus auferantur fimilia, fient 5 Q. aquales 16 N. & fit 1 N. # Eritigitur alter quadratorum # alterverò #. & vtriulque fumma eft<math>ir feu 16. & vterque quadratuseft.

Clearbout's Principle:

"An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."

(*) Buckheit and Donoho (1995) and de Leeuw (2001)

1. Publications which are not reproducible are useless. Source code which computes results shown in papers and the used data sets must be available.

- 1. Publications which are not reproducible are useless. Source code which computes results shown in papers and the used data sets must be available.
- 2. Publications which deal with closed data sets have to be "quasi-reproducible" by providing the analyses and artificial data sets (at best with similar characteristics like the closed data sets).

- 1. Publications which are not reproducible are useless. Source code which computes results shown in papers and the used data sets must be available.
- 2. Publications which deal with closed data sets have to be "quasi-reproducible" by providing the analyses and artificial data sets (at best with similar characteristics like the closed data sets).
- **3.** Providing some source code and, for example, a binary data file on an author's website is not "reproducible". Similar to publications, source code and data sets must be available "forever" and in a standardized way.

- 1. Publications which are not reproducible are useless. Source code which computes results shown in papers and the used data sets must be available.
- 2. Publications which deal with closed data sets have to be "quasi-reproducible" by providing the analyses and artificial data sets (at best with similar characteristics like the closed data sets).
- **3.** Providing some source code and, for example, a binary data file on an author's website is not "reproducible". Similar to publications, source code and data sets must be available "forever" and in a standardized way.
- **4.** A second execution of the analyses on the authors' computers is not "reproducibility". Analyses must be reproducible on an unbiased open platform.







local and global improvements of research



Local Improvements



Figure 1. Experimental Diagram



Figure 2. Experimental Mess

Global Improvements

"We reproduced two analyzes in principle and six partially or with some discrepancies; ten could not be reproduced. The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis."

(*) Ioannidis (2009) via Hothorn and Leisch (2011)

Global Improvements





100 Pine Street, Suite 230, San Francisco, CA 94111 Phone: (415) 777-2750, Fax: (415) 777-2803

Steven L. Shafer, MD Editor-in-Chief

February 25, 2011

To our readers:

In 2009 Dr. Joachim Boldt published a manuscript in *Anesthesia & Analgesia* comparing albumin and hydroxyethyl starch priming cardiopulmonary bypass.¹ The study was retracted in December 2010 for lack of IRB approval.² A subsequent investigation by Klinikum Ludwigshafen determined that the study was fabricated.³

Global Improvements



"Joachim Boldt is at the centre of a criminal investigation amid allegations that he may have forged up to 90 crucial studies on the treatment."

Steven L. Shafe Editor-in-Chief (*) The Telegraph (2011)

In 2009 Dr. Joachim Boldt published a manuscript in *Anesthesia & Analgesia* comparing albumin and hydroxyethyl starch priming cardiopulmonary bypass.¹ The study was retracted in December 2010 for lack of IRB approval.² A subsequent investigation by Klinikum Ludwigshafen determined that the study was fabricated.³

From: Journal Editoral Office Subject: Request to review manuscript #... To: Manuel.Eugster@stat.uni-muenchen.de

Dear Mr. Manuel Eugster,

```
Would you be willing to referee the attached paper by [...]
```

Should you accept, I would be grateful to have your review within [...] days.

Sincerely, The Managing Editor



Rsquared

Reproducible Research for the R Community

Home

About

Services

Welcome to Rsquared, the platform for reproducible research for the R community.

- Submit paper package
- Submit code snippet
- Public paper packages archive

Demonstration



Sweave

Sweave

Packages:

- pkg/
 - DESCRIPTION
 - R/
 - data/
 - demo/
 - inst/
 - doc/

Sweave

Packages:

- pkg/
 - DESCRIPTION
 - R/
 - data/
 - demo/
 - inst/
 - doc/

CRAN

References

- Jonathan Buckheit and David Donoho. WaveLab and reproducible research. Statistics Department, Stanford University, CA, USA, 1995. URL http://www-stat.stanford.edu/~donoho/.
- Jan de Leeuw. Reproducible research: the bottom line. Statistics Program, Unniversity of California, Los Angeles, CA, USA, 2001. URL http://preprints.stat.ucla.edu/.
- Torsten Hothorn and Friedrich Leisch. Case studies in reproducibility. Briefings in Bioinformatics, 12(3):288-300, 2011. doi: 10.1093/bib/bbq084. URL http://bib.oxfordjournals.org/content/12/3/288.abstract.
- John P. A. et al. loannidis. Repeatability of published microarray gene expression analyses. *NATURE GENETICS*, 41(2):149–155, 2009.
- Friedrich Leisch, Manuel J. A. Eugster, and Torsten Hothorn. Executable papers for the R community: The R² platform for reproducible research. *Procedia Computer Science*, 4:618–626, 2011. ISSN 1877-0509. doi: 10.1016/j.procs.2011.04.065. Proceedings of the International Conference on Computational Science, ICCS 2011.
- The Telegraph. Millions of surgery patients at risk in drug research fraud scandal. The Telegraph, Heidi Blake, Holly Watt and Robert Winnett,
 - http://www.telegraph.co.uk/health/8360667/
 - Millions-of-surgery-patients-at-risk-in-drug-research-fraud-scandal. html, visited on July 7, 2011, 2011.
- Wikipedia. Fermat's last theorem. Wikipedia entry, http://en.wikipedia.org/wiki/Fermat's_Last_Theorem, visited on June 24, 2011, 2011.

Images

p7: http://www.funny-gb.de/details.php?image_id=155
p8: http://www.phdcomics.com/comics.php?f=961