



Reproducible Data Analysis 101

@juli_tkotz

Nice to meet you

- Juli Tkotz
- Central Institute for Mental Health,
Mannheim
- PhD student neuropsychology
- Enthusiastic about **open science**
and **coding stuff**



Like, REALLY enthusiastic





What is reproducible data analysis?

Provided data and code,
others can reach the same
analysis results as you.

Is it working?

Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology [→ go to paper](#)



Pepijn Obels¹, Daniël Lakens¹, Nicholas A. Coles², Jaroslav Gottfried³, and Seth A. Green⁴

Abstract

Ongoing technological developments have made it easier than ever before for scientists to share their data, materials, and analysis code. Sharing data and analysis code makes it easier for other researchers to reuse or check published research. However, these benefits will emerge only if researchers can reproduce the analyses reported in published articles and if data are annotated well enough so that it is clear what all variable and value labels mean. Because most researchers are not trained in computational reproducibility, it is important to evaluate current practices to identify those that can be improved. We examined data and code sharing for Registered Reports published in the psychological literature from 2014 to 2018 and attempted to independently computationally reproduce the main results in each article. **Of the 62 articles that met our inclusion criteria, 41 had data available, and 37 had analysis scripts available. Both data and code for 36 of the articles were shared. We could run the scripts for 31 analyses, and we reproduced the main results for 21 articles.** Although the percentage of articles for which both data and code were shared (36 out of 62, or 58%) and the percentage of articles for which main results could be computationally reproduced (21 out of 36, or 58%) were relatively high compared with the percentages found in other studies, there is clear room for improvement. We provide practical recommendations based on our observations and cite examples of good research practices in the studies whose main results we reproduced.

Is it working?

Andrew C. Chang* and Phillip Li†

September 4, 2015

Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not”

Abstract

[→ go to paper](#)

We attempt to replicate 67 papers published in 13 well-regarded economics journals using author-provided replication files that include both data and code. Some journals in our sample require data and code replication files, and other journals do not require such files. Aside from 6 papers that use confidential data, we obtain data and code replication files for 29 of 35 papers (83%) that are required to provide such files as a condition of publication, compared to 11 of 26 papers (42%) that are not required to provide data and code replication files. We successfully replicate the key qualitative result of 22 of 67 papers (33%) without contacting the authors. Excluding the 6 papers that use confidential data and the 2 papers that use software we do not possess, we replicate 29 of 59 papers (49%) with assistance from the authors. Because we are able to replicate less than half of the papers in our sample even with help from the authors, we assert that economics research is usually not replicable. We conclude with recommendations on improving replication of economics research.

What is reproducible data analysis?

Provided data and code,
others can reach the same
analysis results as you.

Dirty Data

| participant | age | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | score1 | score2 |
|-------------|---------|----|------|----|------|----|----|---------|--------|
| I64476XG | 40 | 3 | 2 | 2 | 2 | 1 | 4 | 5 | 5 |
| C376685NP | 33 | 3 | 3 | 4 | 5 | 4 | 4 | 3,75 | 6 |
| Q60968TN | 35 | 1 | 1 | 2 | 4 | 2 | 4 | 1,5 | 2 |
| A231763DX | 29 | 3 | 5 | 5 | 5 | 1 | 2 | 8 | 8 |
| O11728CW | 33 | 5 | 4 | 4 | 4 | 1 | 2 | 9 | 9 |
| Z100815RZ | 39 | 3 | 2 | 2 | -999 | 3 | 3 | 3 | 5 |
| W889138LS | 23 | 1 | 5 | 3 | 4 | 2 | 2 | 5,5 | 6 |
| H665158SI | 25 | 1 | 4 | 1 | 1 | 2 | 3 | 4,5 | 5 |
| T67968QK | 38 | 1 | 1 | 3 | 1 | 3 | | 1,33333 | 2 |
| K336076AD | 35 | 3 | 4 | 1 | 3 | 3 | 2 | 5 | 7 |
| L413131PO | 19 | 4 | 5 | 4 | 5 | 1 | 5 | 9 | 9 |
| R561317MT | 36 | 2 | -666 | 4 | 3 | 3 | 1 | -665,33 | -664 |
| D167267JF | zwanzig | 1 | 3 | 4 | 1 | 1 | 4 | 4 | 4 |
| N53747FJ | 34 | 5 | 4 | 4 | 1 | 3 | 1 | 5,66667 | 9 |
| P98433YE | 23 | 3 | 1 | 4 | 4 | 2 | 3 | 2,5 | 4 |
| V446086IQ | 30 | 2 | 3 | 3 | 5 | 2 | 4 | 4 | 5 |
| S586643UY | 200 | 4 | 3 | 5 | 2 | 3 | 5 | 4,33333 | 7 |

| participant | age | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | score1 | score2 |
|-------------|---------|------|------|------|------|------|------------|---------|--------|
| I64476XG | 40 | ?! 3 | ?! 2 | ?! 2 | ?! 2 | ?! 1 | ?! 4 | ?? 5 | ?? 5 |
| C376685NP | 33 | 3 | 3 | 4 | 5 | 4 | 4 | 3,75 | 6 |
| Q60968TN | 35 | 1 | 1 | 2 | 4 | 2 | 4 | 1,5 | 2 |
| A231763DX | 29 | 3 | 5 | 5 | 5 | 1 | 2 | 8 | 8 |
| O11728CW | 33 | 5 | 4 | 4 | 4 | 1 | 2 | 9 | 9 |
| Z100815RZ | 39 | 3 | 2 | 2 | -999 | ?? 3 | 3 | 3 | 5 |
| W889138LS | 23 | 1 | 5 | 3 | 4 | ?? 2 | 2 | 5,5 | 6 |
| H665158SI | 25 | 1 | 4 | 1 | 1 | 2 | 3 | 4,5 | 5 |
| T67968QK | 38 | 1 | 1 | 3 | 1 | 3 | ?? 1,33333 | | 2 |
| K336076AD | 35 | 3 | 4 | 1 | 3 | 3 | 2 | 5 | 7 |
| L413131PO | 19 | 4 | 5 | ?? 4 | 5 | 1 | 5 | 9 | 9 |
| R561317MT | 36 | 2 | -666 | 4 | 3 | 3 | 1 | -665,33 | -664 |
| D167267JF | zwanzig | ? 1 | 3 | 4 | 1 | 1 | 4 | 4 | 4 |
| N53747FJ | 34 | 5 | 4 | 4 | 1 | 3 | 1 | 5,66667 | 9 |
| P98433YE | 23 | 3 | 1 | 4 | 4 | 2 | 3 | 2,5 | 4 |
| V446086IQ | 30 | ? 2 | 3 | 3 | 5 | 2 | 4 | 4 | 5 |
| S586643UY | 200 | 4 | 3 | 5 | 2 | 3 | 5 | 4,33333 | 7 |

Codebooks

This is a data set from Omnomnom et al. (2021). N = 666 participants answered questions about their pizza preferences. Based on this, the Pizza Personality Scale was developed, with the two personality facets "mushroom" and "calzone". For more information, contact juliane.tkotz@zi-mannheim.de

Missing values are coded as NA

| | | |
|----------------------------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| participant_id | | Character. Participant identification code. A random sequence of characters and numbers. |
| age | | Numeric. Participant age in years. min = 18, max = 40 |
| Q1 | | Pizza survey question 1: "How do you prefer your pizza?" 1 = "very thin" - 5 = "very thick" (whole numbers) |
| Q2 | | Pizza survey question 2: "How do you feel about pineapple on pizza?" 1 = "absolutely disgusting" - 5 = "love it" (whole numbers) |
| Q3 | | Pizza survey question 3: "How often do you eat pizza?" 1 = "never" - 5 = "every day" (whole numbers) |
| Q4 | | Pizza survey question 4: "Do you prefer meat or vegetables on pizza?" 1 = "strongly prefer meat" - 5 = "strongly prefer vegetables" (whole numbers) |
| Q5 | | Pizza survey question 5: "Barbecue sauce on pizza is ..." 1 = "very delicious" - 5 = "very disgusting" (whole numbers) |
| Q6 | | Pizza survey question 6: "How about pizza with pasta as topping?" 1 = "very delicious" - 5 = "very disgusting" (whole numbers) |
| mushroom_personality_score | | Sum score of items Q1, Q4 and Q5, which together make up the mushroom dimension of the pizza personality scale. |

```
codebook.txt - Notepad
File Edit Format View Help
This is a data set from Omnibus.com. Participants answered questions
about their pizza preferences. Data from a personality scale was developed,
with the two personality facets "mushroom" and "calzone". For more information, contact
juliane.tkotz@zi-mannheim.de

Missing values are coded as NA

participant_id ..... participant identification code. A random
age ..... numeric. Participant age in years. min = 18, max = 40
Q1 ..... Pizza survey question 1: "How do you prefer your pizza?
1 = "very thin" - 5 = "very thick" (whole numbers)
Q2 ..... Pizza survey question 2: "How do you feel about pineapple
on pizza?" 1 = "absolutely disgusting" - 5 = "love it"
(whole numbers)
Q3 ..... Pizza survey question 3: "How often do you eat pizza?"
1 = "never" - 5 = "every day" (whole numbers)
Q4 ..... Pizza survey question 4: "Do you prefer meat or
vegetables on pizza?" 1 = "strongly prefer meat" -
5 = "strongly prefer vegetables" (whole numbers)
Q5 ..... Pizza survey question 5: "Barbecue sauce on pizza is ..."
1 = "very delicious" - 5 = "very disgusting" (whole
numbers)
Q6 ..... Pizza survey question 6: "How about pizza with pasta as
topping?" 1 = "very delicious" - 5 = "very disgusting"
(whole numbers)
mushroom_personality_score ..... Sum score of items Q1, Q4 and Q5, which together make
up the mushroom dimension of the pizza personality scale.
```

"Foolproof" file format

Background

How many participants?

the data

What has been done?

How are missing values treated?

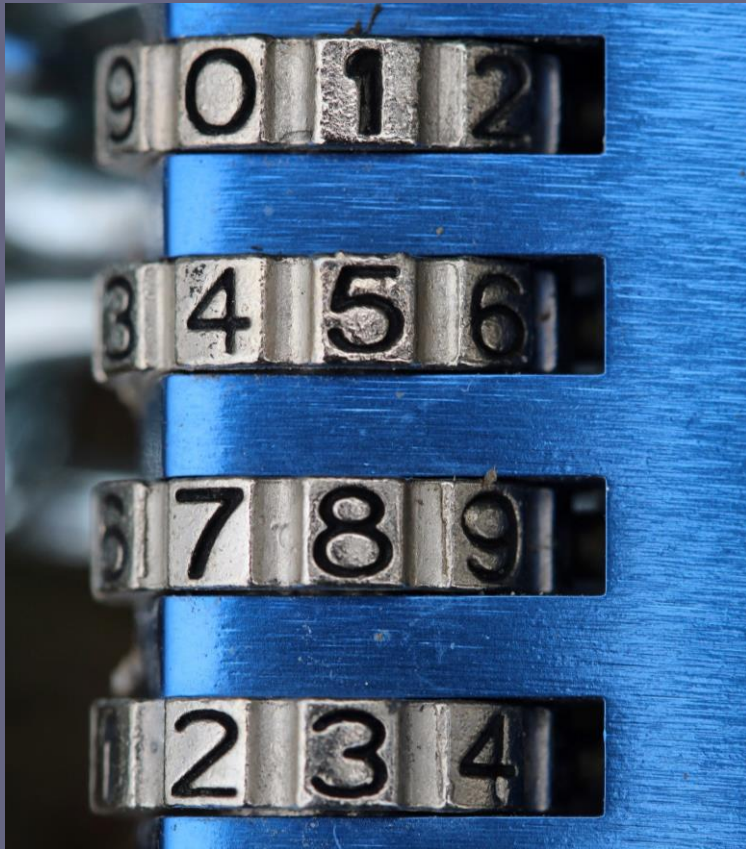
Data type?
Unit?
What values to expect?
What was the question?

Variable names

Variable info

Safe storage

→ [Nature article about data management plans](#)



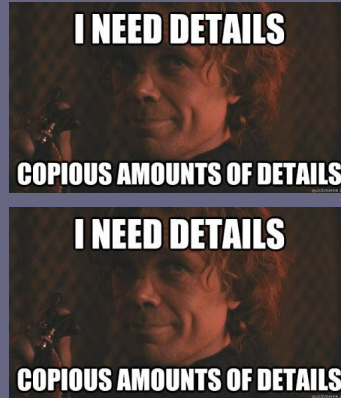
Anne Nygård on Unsplash

- Put your data somewhere safe!
- It should also be (safely!) accessible for others.
- What data repositories are commonly used in your field?
- Do they provide a doi for your data set?
- Upload meta data and/or a codebook.
- Think about data protection.
- Choose a licence.

What is reproducible data analysis?

Provided data and code,
others can reach the same
analysis results as you.

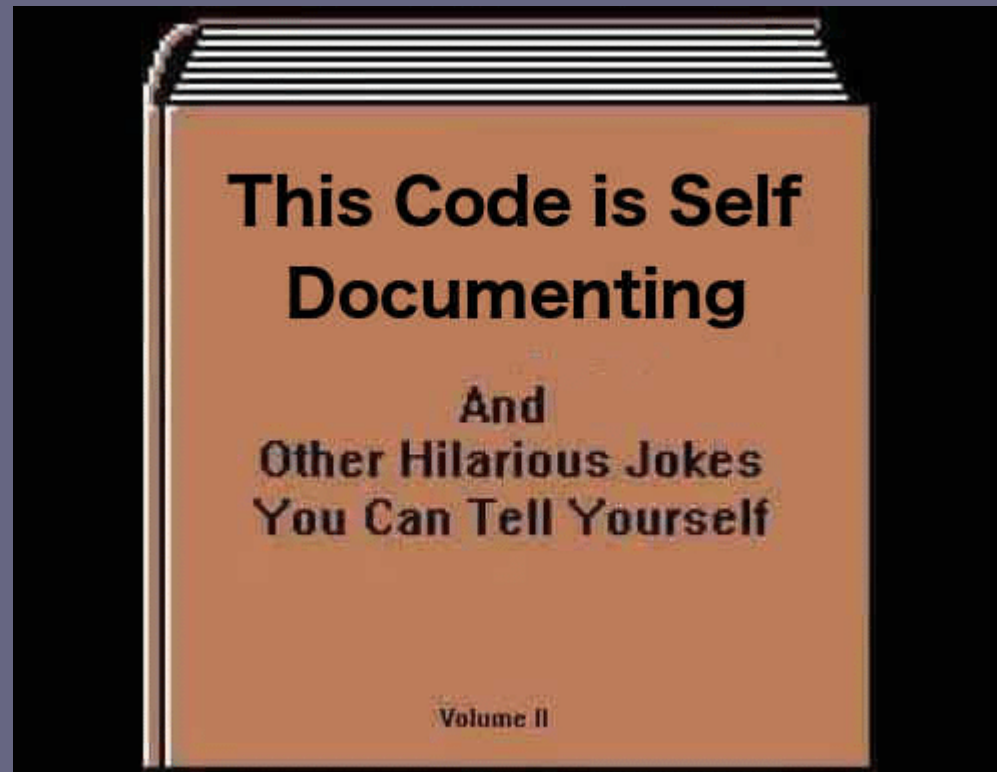
Code vs. point and click



- It provides all the details of your analysis.
- It's easy to run it again and again.
- Because for some analyses, there is no other way.
- Bonus points: You don't need to copy-paste your results into your manuscript.

➡ It's a key ingredient for transparent and reproducible analyses.

Same as for data (but different)



Be structured!



```
R 00_data_processing.R  
R 01_calculate_scores.R  
R 02_descriptive_statistics.R  
R 03_analysis_pizza_preference.R  
R 04_analysis_pizza_regressions.R  
R 05_generate_plots.R
```

BETTER

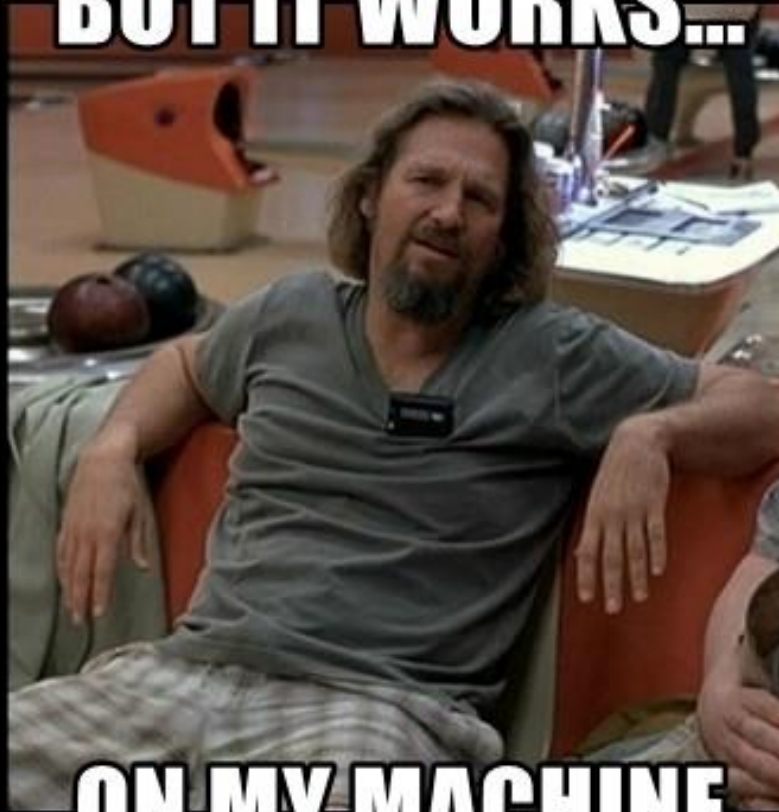
Nomen est omen

```
1
2 a <- 42
3 test <- 9 + a
4 Test <- test + a
5 Test2 <- test + test
6
7 bananapudding <- c(test, Test, Test2)
8
9 ImVeryInconsistent <- vector(
10   mode = "numeric",
11   length = length(bananapudding))
12
13
14 for (dd in seq_along(bananapudding)) {
15   ImVeryInconsistent[dd] <-
16     bananapudding[dd] * 2
17 }
```

- Use meaningful variable names.
 - E.g. df_sales_per_pizza or taxation_rate
- Use a consistent style.
- Structure your code. Code wants to breathe.
- Maybe look at a style guide for your language.

Think about others

BUT IT WORKS...



ON MY MACHINE

memegenerator.net

Make it self-contained

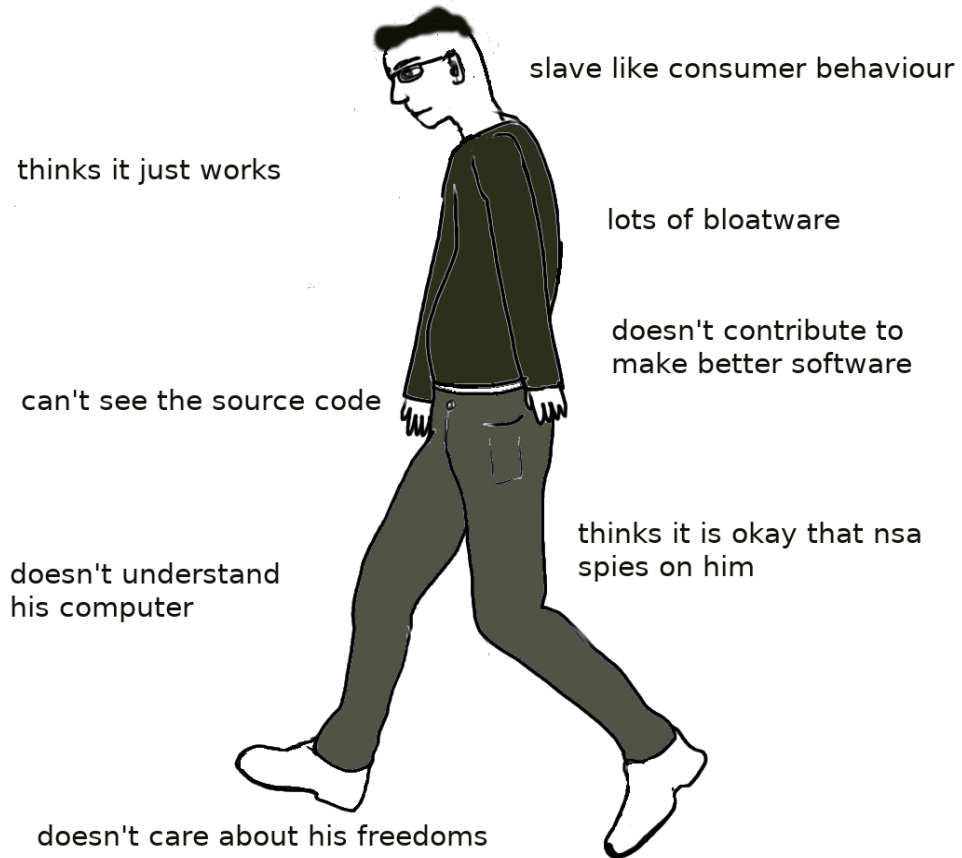


Photo by Fallon Michael on Unsplash

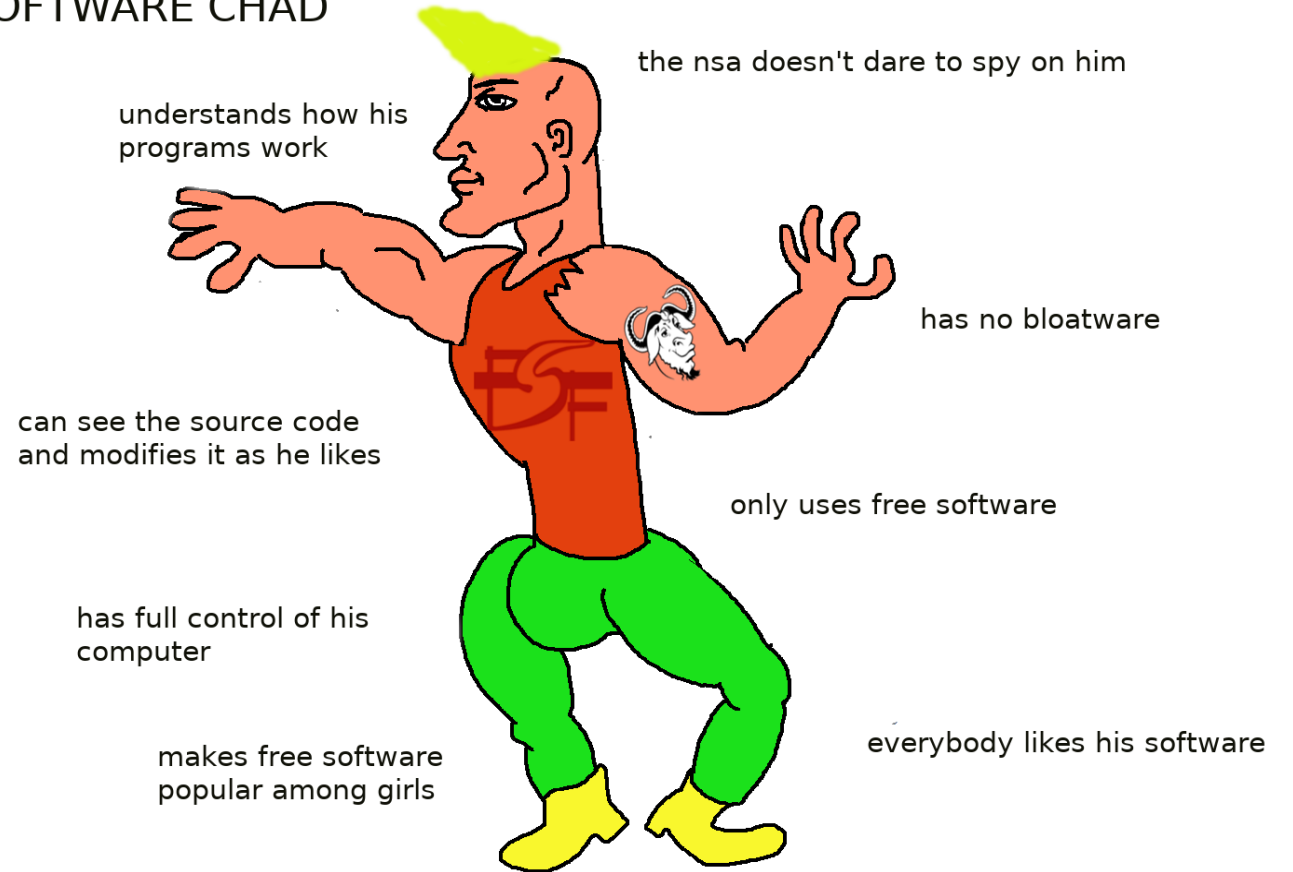
- Don't include anything that is specific to your computer. E.g. file paths.
 - `C:/Users/juli/OneDrive/Dokumente/sleep_study/data_21-04-14.csv`
 - `./data_21-04-14.csv`
- Use relative paths.
- Ideally, everything you need to run your script is in the same place.

Use free/open source software if you can

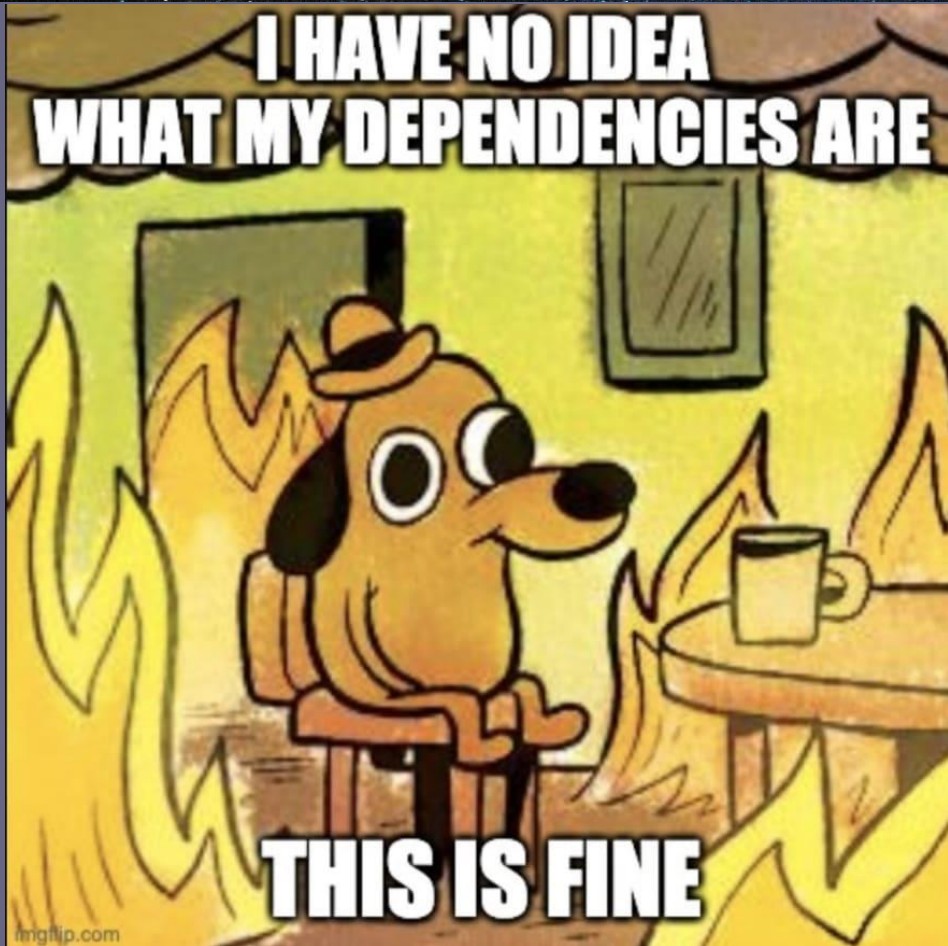
"The Proprietary Virgin"



THE FREE SOFTWARE CHAD



But still ...



- Different operating systems.
- Different software versions.
- Different additional packages.

Different software and OS versions

The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements

[Ed H. B. M. Gronenschild](#), ^{1, 2, *} [Petra Habets](#), ^{1, 2} [Heidi I. L. Jacobs](#), ^{1, 2, 3} [Ron Mengelers](#), ^{1, 2} [Nico Rozendaal](#), ^{1, 2} [Jim van Os](#), ^{1, 2, 4} and [Machteld Marcelis](#) ^{1, 2}

FreeSurfer is a popular software package to measure cortical thickness and volume of neuroanatomical structures. However, little if any is known about measurement reliability across various data processing conditions. Using a set of 30 anatomical T1-weighted 3T MRI scans, we investigated the effects of data processing variables such as FreeSurfer version (v4.3.1, v4.5.0, and v5.0.0), workstation (Macintosh and Hewlett-Packard), and Macintosh operating system version (OSX 10.5 and OSX 10.6). Significant differences were revealed between FreeSurfer version v5.0.0 and the two earlier versions. These differences were on average $8.8 \pm 6.6\%$ (range 1.3–64.0%) (volume) and $2.8 \pm 1.3\%$ (1.1–7.7%) (cortical thickness). About a factor two smaller differences were detected between Macintosh and Hewlett-Packard workstations and between OSX 10.5 and OSX 10.6. The observed differences are similar in magnitude as effect sizes reported in accuracy evaluations and neurodegenerative studies.

The main conclusion is that in the context of an ongoing study, users are discouraged to update to a new major release of either FreeSurfer or operating system or to switch to a different type of workstation without repeating the analysis; results thus give a quantitative support to successive recommendations stated by FreeSurfer developers over the years. Moreover, in view of the large and significant cross-version differences, it is concluded that formal assessment of the accuracy of FreeSurfer is desirable.

[→ go to paper](#)

Code Capsules

Code capsules can bridge these differences.

For example:



Environment

R (3.6.0)

R is a language and environment for statistical computing and graphics

Ubuntu 18.04 R



Additional Packages ?

Customize the selected environment with any other packages you need. You can also use these package managers to install other package managers, such as for different languages. Packages will be installed on the next capsule run. [Learn more.](#)

Package Managers

Packages

apt-get

lmodern 2.004.5-3

pandoc 1.19.2.4-dfsg-1build4

pandoc-citeproc 0.10.5.1-1build4

rstudio-server 1.2.5033

wget 1.19.4-1ubuntu2.2

+ Add

Bioconductor

+ Add

R (CRAN)

binb 0.0.5

dplyr 0.8.5

effsize 0.8.0

ggplot2 3.3.0

knitr 1.28

rmarkdown 2.1

tinytex 0.23

+ Add

R (GitHub)

m-Py/prmisc b010558ead60afb76216cd529ce877c6720e68bd

+ Add

Bonus: Reproducible Manuscripts



263 3.1 Epistemic Trustworthiness

264 Participants placed more epistemic trust in the debaters when reading a neutral debate: Student
265 teachers in the neutral condition ($M = 5.06, SD = 1.00$) perceived the debaters to have more expertise
266 than those in the uncivil condition ($M = 5.06, SD = 1.00$), $t(218.49) = 1.99, p = .047, d = 0.27$.
267 Furthermore, participants reading a neutral debate ($M = 4.76, SD = 1.02$) reported higher ratings of
268 debaters' integrity than those reading an uncivil debate ($M = 4.05, SD = 1.15$), $t(219.41) = 4.87, p <$
269 $.001, d = 0.65$. Additionally, ratings of benevolence were higher in the neutral condition ($M = 4.77,$
270 $SD = 0.98$) than in the uncivil condition ($M = 4.05, SD = 0.89$), $t(214.11) = 5.67, p < .001, d = 0.76$
271 (see Figure 2).

272 We further explored the correlation between the conflict explanation items and the METI subscales,
273 that is, if the perception of various aspects of a conflict was associated with different degrees of
274 epistemic trust. Those who agreed that the debaters in the scenario referred to different research
275 results also thought them to have more expertise, $r(220) = .14, p = .039$. There was no relation with
276 integrity, $r(220) = .07, p = .321$, or benevolence, $r(220) = .03, p = .679$. Assuming personal reasons
277 for the conflict had the strongest relationship with epistemic trust. The more participants perceived
278 the conflict to be personal, the less expertise they assigned to the debaters $r(220) = -.25, p < .001$. In
279 a similar manner, perception of a personal conflict lead to decreased ratings of integrity, $r(220) =$
280 $-.36, p < .001$, and benevolence, $r(220) = -.41, p < .001$. How much participants agreed that the
281 debaters referred to different goals of PAVLOV did not correlate with any of the METI subscales,
282 neither with expertise, $r(220) = .10, p = .122$, nor with integrity, $r(220) = -.00, p = .946$, nor with
283 benevolence $r(220) = -.00, p = .994$. Embracement of the statement that debaters referred to different
284 effects of PAVLOV was not associated with epistemic trust either, neither with expertise, $r(220) =$
285 $.01, p = .863$, nor with integrity, $r(220) = -.06, p = .348$, nor with benevolence $r(220) = -.05, p =$
286 $.475$. Internal consistency of the METI subscales was somewhat lower than initially found by
287 Hendriks et al. (2015), with a Cronbach's α of .87 for expertise, .83 for integrity and .76 for
288 benevolence.

```
## R Markdown to the rescue
```

```
```{r intext_stats, echo = TRUE}
nerd <- read.csv("./data/nerd.csv", sep = "\t")
```
```

```
```{r copy_paste_hell}
include_graphics("./pics/slide_inception.png")
```
```

This example dataset consists of $N = \text{r nrow(nerd)}$ participants with an age range between $\text{r min(nerd[["age"]])}$ and $\text{r max(nerd[["age"]])}$ years. Overall, $\text{r sum(nerd$age > 100)}$ participants reported to be older than 100, so we probably can't trust this data set a lot.



R Markdown saving my a**

3.1 Epistemic Trustworthiness

Participants placed more epistemic trust in the debaters when reading a neutral debate: Student teachers in the neutral condition ($M = 5.06, SD = 1.00$) perceived the debaters to have more expertise than those in the uncivil condition ($M = 5.06, SD = 1.00$), $t(218.49) = 1.99, p = .047, d = 0.27$.

Epistemic Trustworthiness

```
Participants placed more epistemic trust in the debaters when reading a neutral debate: Student teachers in the neutral condition `r print_mean_sd(meti$expertise_sum[meti$condition == "neutral"])` perceived the debaters to have more expertise than those in the uncivil condition `r print_mean_sd(meti$expertise_sum[meti$condition == "neutral"])`, `r print_ttest(t_meti_exp, d_meti_exp)`. Furthermore, participants reading a neutral debate `r print_mean_sd(meti$integrity_sum[meti$condition == "neutral"])` reported higher ratings of debaters' integrity than those reading an uncivil debate `r
```

Further advantages



- Directly render high quality plots.
- Never worry about formatting again.
- Interactive papers.

I'm sorry

**AND THEN I TOLD MY
STUDENTS**



**TO LEARN A NEW PROGRAMMING
LANGUAGE JUST FOR MY LECTURE**

Yes. Yet another thing to learn.

But ...

- Eventually, it saves time. For you and your colleagues.
- We don't really have a choice ... have we?

Do it for future you





Thank you!
@juli_tkotz