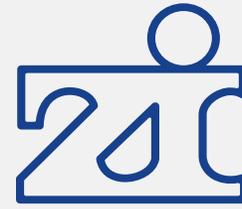


Juli Tkotz

Dep. Clinical Psychology, RG Psychology and  
Neurobiology of Sleep and Memory



Zentralinstitut für  
Seelische Gesundheit

Landesstiftung  
des öffentlichen Rechts

Make your code reproducible  
Mannheim Open Science Day, 20.10.2020



# Why code?



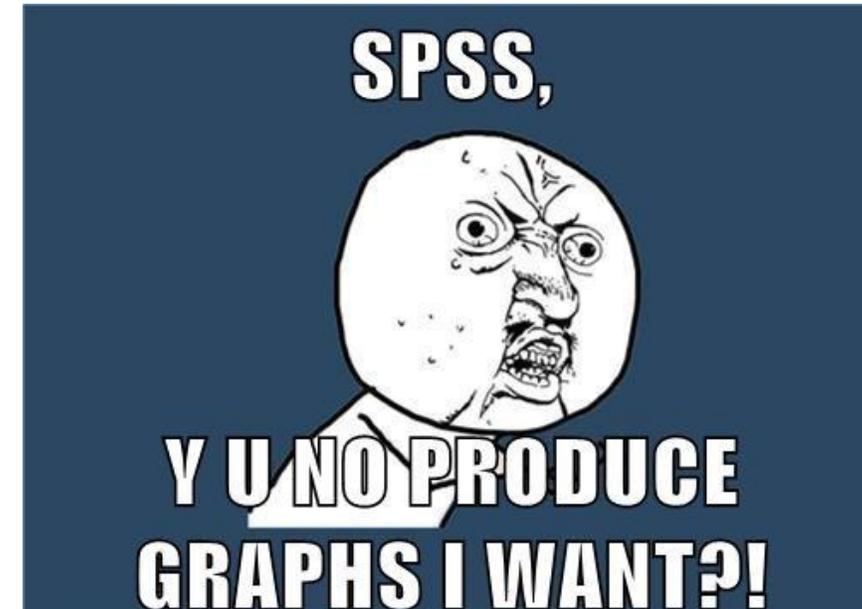
- It provides all the details of your analysis.

# Why code?

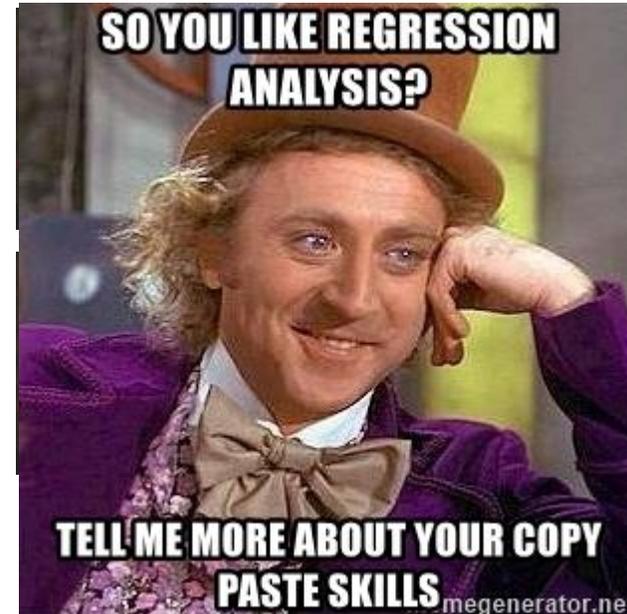


- It provides all the details of your analysis.
- It's easy to run it again and again.

- 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.



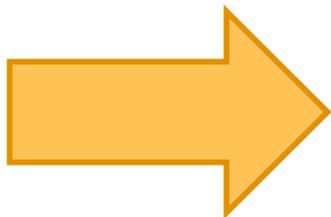
# Why code?



- 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 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 science.

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



Pepijn Obels<sup>1</sup>, Daniël Lakens<sup>1</sup>, Nicholas A. Coles<sup>2</sup>, Jaroslav Gottfried<sup>3</sup>, and Seth A. Green<sup>4</sup>

<sup>1</sup>Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology; <sup>2</sup>Department of Psychology, University of Tennessee; <sup>3</sup>Department of Psychology, Masaryk University; and <sup>4</sup>Department of Psychology, Princeton University

### 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 Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not” [→ go to paper](#)

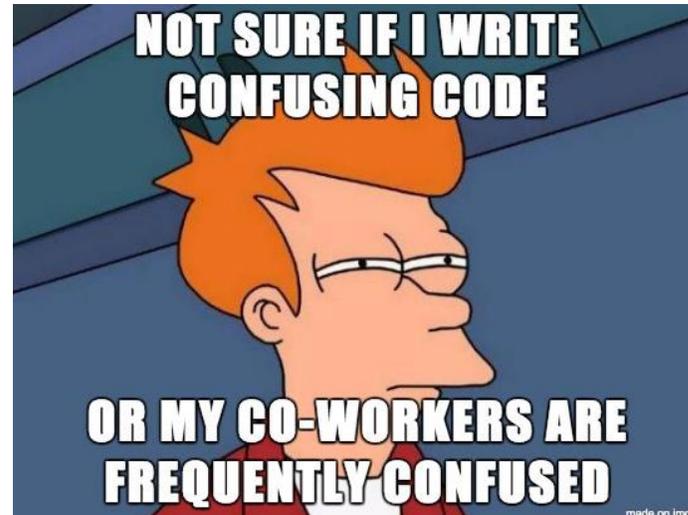
Andrew C. Chang\* and Phillip Li†

September 4, 2015

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.



No formal computer science training.



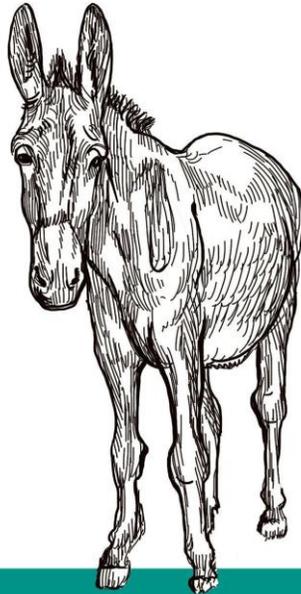
No training in coding conventions or good practices.

Colleague: I don't understand what your code does, can you write some documentation for it?  
Me:



Usage of pre-existing scripts written by colleagues ages ago.

*Where's the fun in just knowing what the code is supposed to do?*



*Essential*

## Excuses for Not Writing Documentation

○ RLY?

@ThePracticalDev

<https://www.reddit.com/r/orlybooks/>

Document. The Hell. Out of it.

```
3 #' What the function is doing
4 #'
5 #' @param argument1 Describe the type of the argument and what it is doing.
6 #' @param argument2 Describe the type of the argument and what it is doing.
7 #'
8 #' @details
9 #'
10 #' More details about the function and what to watch out for.
11 #'
12 #' @example
13 #'
14 #' # Example of how to use this function
15 #'
16 my_function <- function(argument1, argument2) {
17   x <- (x + argument1) * argument2 # more comments if necessary
18   return(x)
19 }
```

CODEBOOK - Editor

Datei Bearbeiten Format Ansicht Hilfe

Some general information about the data set. Then, describe what each column contains, e.g. how a variable was measured and which values are to be expected.

participant_ID .....	Participant identification number. Structure: Experiment name_three digits_two characters, e.g. APOLLO11_666_XX
age .....	Participants' age in years. Min: 18; max: 67. Note: Missing values are coded as -333.
home_planet .....	Character; the planet participants live on. Valid entries are: "earth", "mars", "jupiter"

Agree on standards, like e.g. BIDS

### About BIDS

Neuroimaging experiments result in complicated data that can be arranged in many different ways. So far there is no consensus how to organize and share data obtained in neuroimaging experiments. Even two researchers working in the same lab can opt to arrange their data in a different way. Lack of consensus (or a standard) leads to misunderstandings and time wasted on rearranging data or rewriting scripts expecting certain structure. With the Brain Imaging Data Structure (BIDS), we describe a simple and easy to adopt way of organizing neuroimaging and behavioral data.



```
dicomdir/
├── 1208200617178_22/
│   ├── 1208200617178_22_8973.dcm
│   ├── 1208200617178_22_8943.dcm
│   ├── 1208200617178_22_2973.dcm
│   ├── 1208200617178_22_8923.dcm
│   ├── 1208200617178_22_4473.dcm
│   ├── 1208200617178_22_8783.dcm
│   ├── 1208200617178_22_7328.dcm
│   ├── 1208200617178_22_9264.dcm
│   ├── 1208200617178_22_9967.dcm
│   ├── 1208200617178_22_3894.dcm
│   └── 1208200617178_22_3899.dcm
├── 1208200617178_23/
├── 1208200617178_24/
└── 1208200617178_25/

my_dataset/
├── participants.tsv
├── sub-01/
│   ├── anat/
│   │   └── sub-01_T1w.nii.gz
│   ├── func/
│   │   ├── sub-01_task-rest_bold.nii.gz
│   │   └── sub-01_task-rest_bold.json
│   └── dwi/
│       ├── sub-01_dwi.nii.gz
│       ├── sub-01_dwi.json
│       ├── sub-01_dwi.bval
│       └── sub-01_dwi.bvec
├── sub-02/
├── sub-03/
└── sub-04/
```

Invest time to learn about good coding practices, e.g. 10 good coding practices for data scientists.

There are probably style guides for your language, e.g. in Advanced R by Hadley Wickham.

## Style guide

Good coding style is like using correct punctuation. You can manage without it, but it sure makes things easier to read. As with styles of punctuation, there are many possible variations. The following guide describes the style that I use (in this book and elsewhere). It is based on Google's [R style guide](#), with a few tweaks. You don't have to use my style, but you really should use a consistent style.

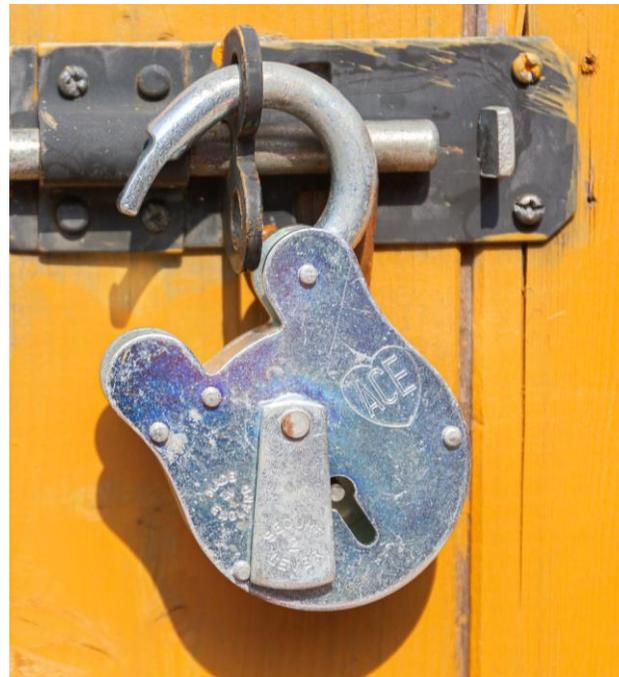
## Missing software

```
> library(brms)  
Error in library(brms) : there is no package called 'brms'
```

### MATLAB

- EUR 2.000 ⓘ  
Perpetual license
- EUR 800 ⓘ  
Annual license

Buy now



Part of the solution: open source software.

Check out [this amazing list](#) on GitHub!

# More specific problems

Don't use absolute paths!

```
20  
21 data <- read.csv("C:\\Users\\Jane.Doe\\my_project\\my_data.csv")  
22
```

```
Error in file(file, "rt") : cannot open the connection  
In addition: Warning message:  
In file(file, "rt") :  
cannot open file 'C:\\Users\\Jane.Doe\\my_project\\my_data.csv': No such file or directory
```

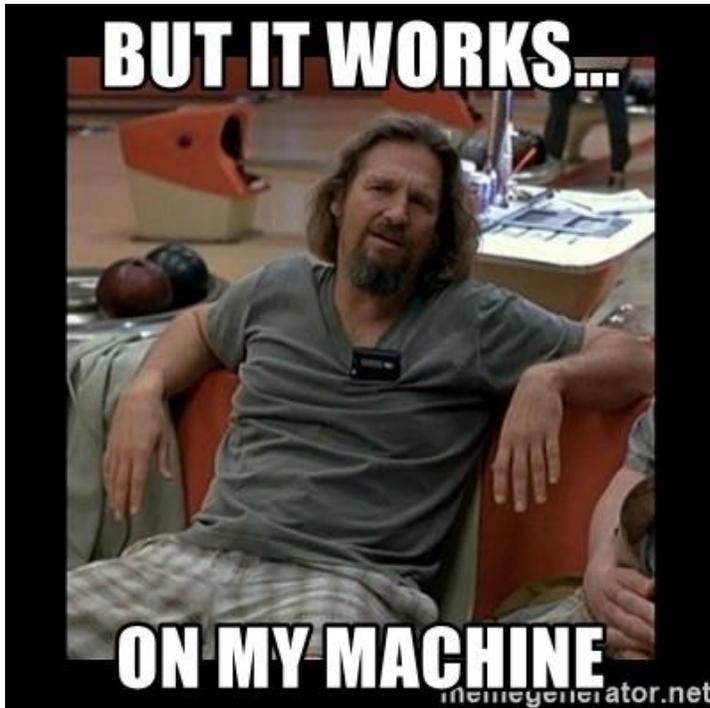
```
23  
24 data <- read.csv(".\\my_data.csv")  
25
```

Use relative paths in a project-oriented workflow.

But still ...



Zentralinstitut für  
Seelische Gesundheit  
Landesstiftung  
des öffentlichen Rechts



Code capsules can bridge  
these differences.  
For example:



People have different  
software (versions), different  
operating systems and  
different backgrounds.

## Environment

### R (3.6.0)

R is a language and environment for statistical computing and graphics

Ubuntu 18.04 R



Time machine!

## 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

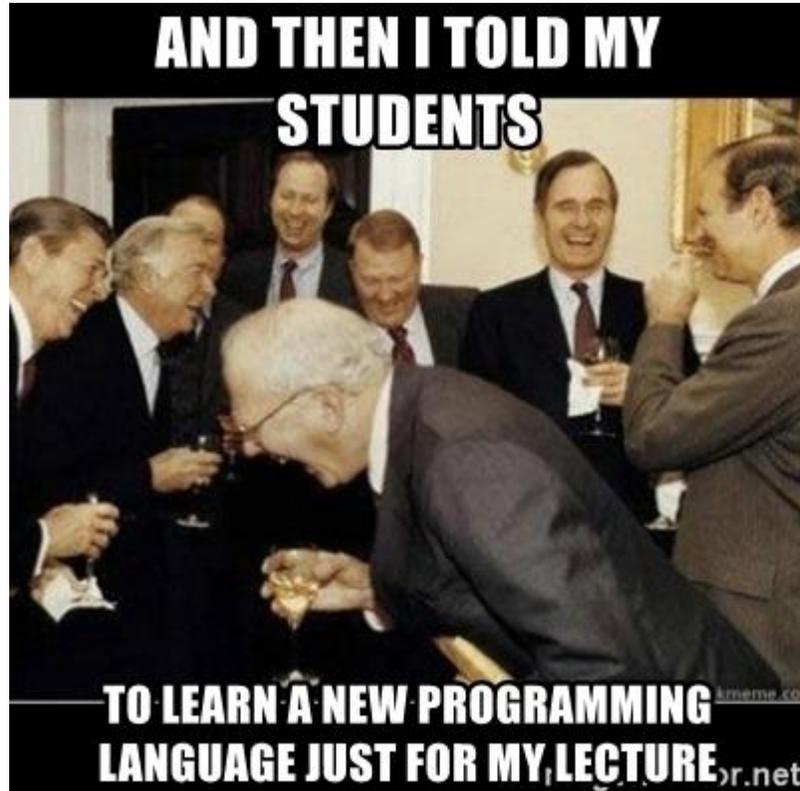
tinytex 0.23

+ Add

#### R (GitHub)

m-Py/prmisc b010558ead60afb76216cd529ce877c6720e68bd

+ Add



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?

Thank you for your attention

[juliane.tkotz@zi-mannheim.de](mailto:juliane.tkotz@zi-mannheim.de)

