



Univerzita Palackého  
v Olomouci



---

Institut sociálního zdraví  
Univerzity Palackého

Krise reprodukovatelnosti psychologických výzkumů:  
představení programu zajišťujícího vyšší  
reprodukovatelnost datových analýz

**Mgr. et Mgr. Lukáš Novák**



Institut sociálního zdraví  
Univerzity Palackého

## Co znamená termín „výpočetní opakovatelnost“?

Výpočetní opakovatelnost/reprodukovatelnost představuje situaci, kdy je možné zopakovat výsledky analýz za pomoci původních dat a kódu (Valdez et al., 2020, s. 29).

# Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology



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.

Advances in Methods and Practices in Psychological Science  
2020, Vol. 3(2) 229–237  
© The Author(s) 2020

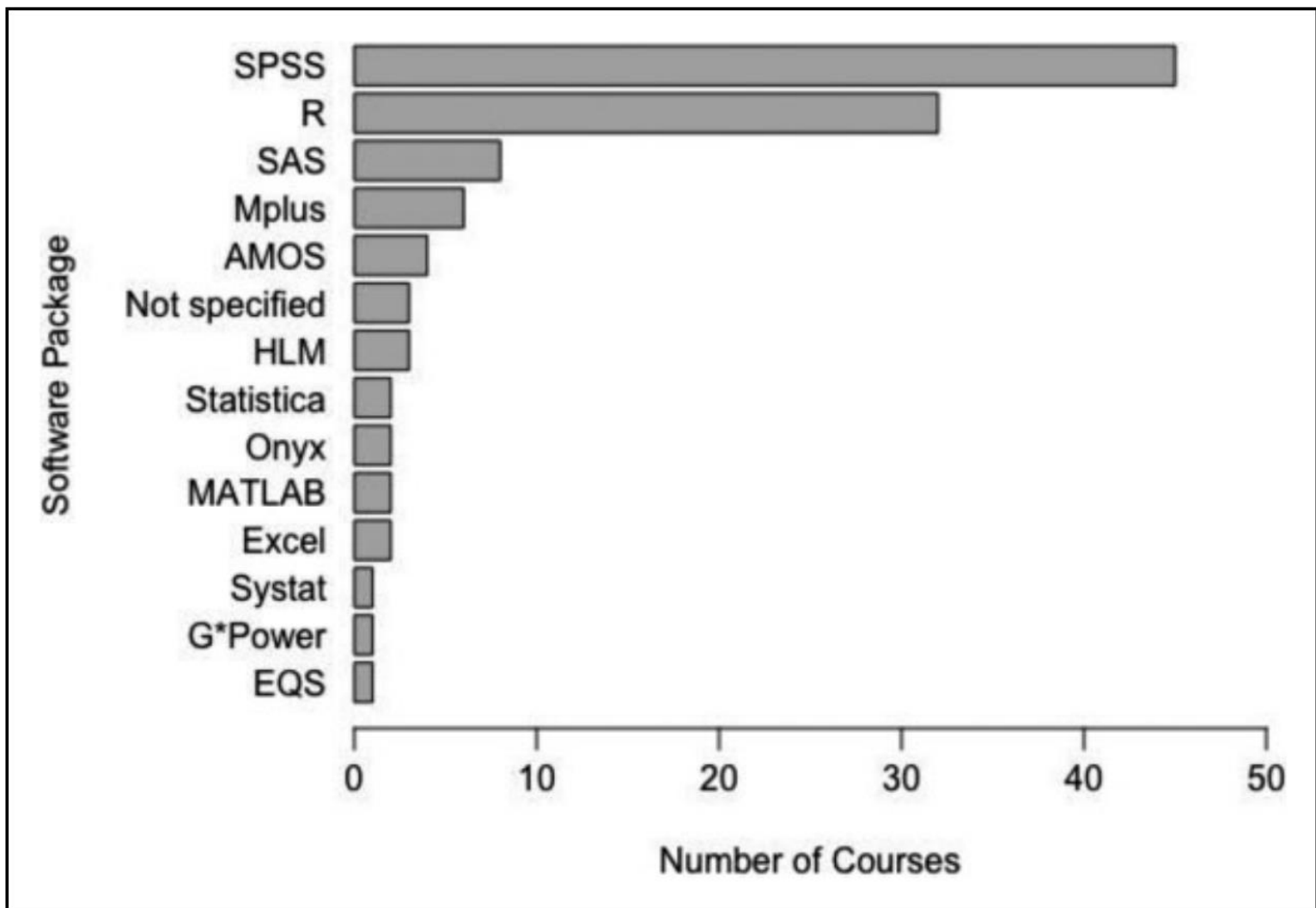


Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/2515245920918872  
[www.psychologicalscience.org/AMPPS](https://www.psychologicalscience.org/AMPPS)





For the 15 articles whose results could not be reproduced, the main reasons were as follows: (a) In 8 cases, code to reproduce some values (e.g., dedicated code to run a macro in SPSS) was missing; (b) in 6 cases, the code gave errors (e.g., variables in the data set were missing, or functions did not run as expected)



**Figure 4.** Number of graduate courses using each software package.



Institut sociálního zdraví  
Univerzity Palackého

# R a R balíčky



tidyr

plotly

ggplot2

tidytext

ggraph

stringr

dplyr

reshape2

tidyquant

dichromat

## list of R Packages

dygraphs

digest

leaflet

MASS

ggmap

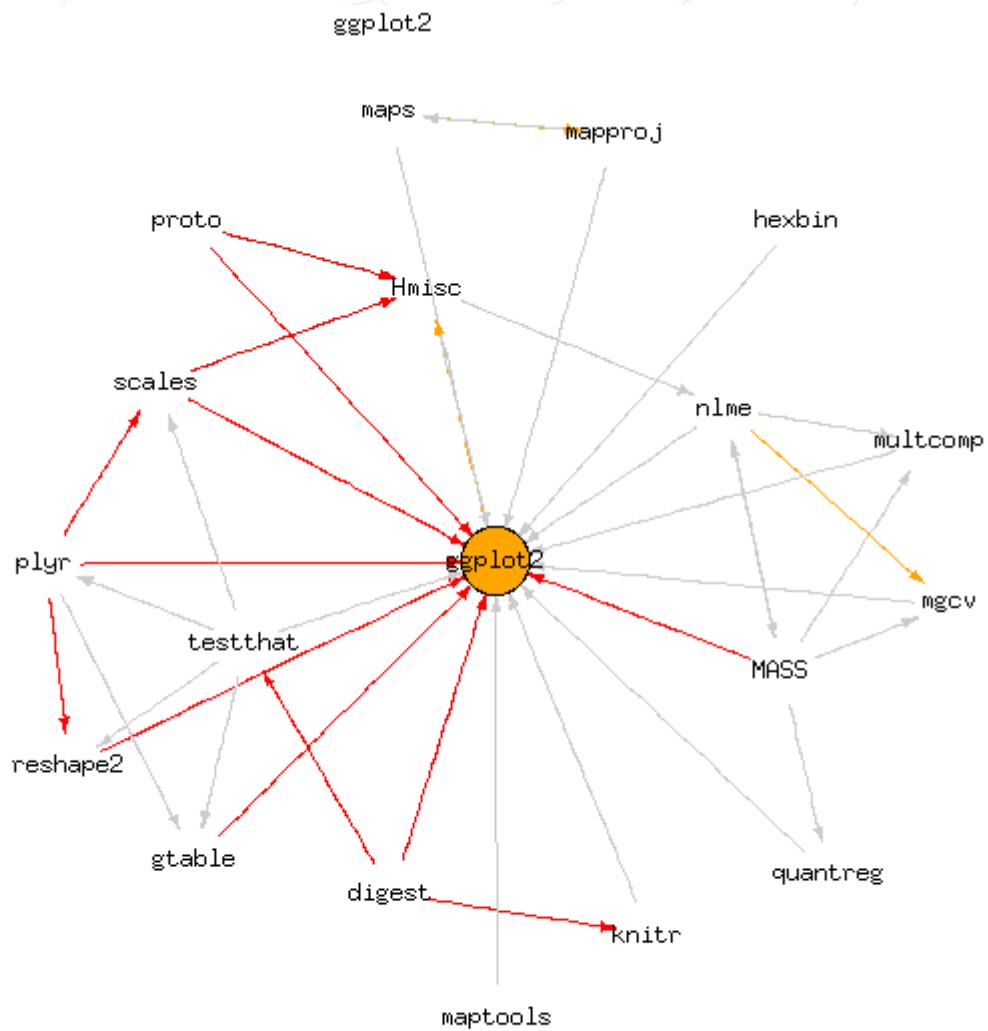
caret

glue

e1071

shiny

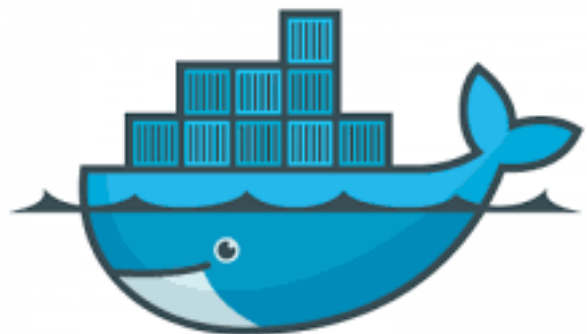
sentimentr







Institut sociálního zdraví  
Univerzity Palackého



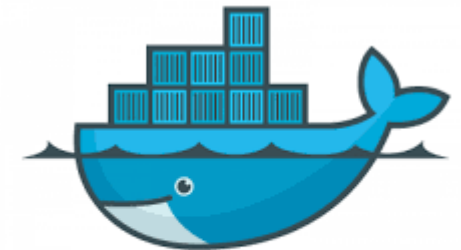
docker



# Docker image a kontejnery



Institut sociálního zdraví  
Univerzity Palackého



docker

Vrstva kontejneru

Čtení a  
upravování

Image

Image 1

Image 2

Image 3

Image 4

Jen pro čtení



Institut sociálního zdraví  
Univerzity Palackého

## Kontrola verzí (GitHub, GitLab)





# R Markdown

The screenshot shows the RStudio editor with a file named 'notebooktest.Rmd'. The code in the editor is as follows:

```
1 ---
2 title: "R Notebook und Python"
3 output:
4   html_notebook: default
5   pdf_document: default
6 ---
7
8 Das ist ein [R Markdown](http://rmarkdown.rstudio.com)
9 Notebook, das auch Python kann. When you execute code
10 within the notebook, the results appear beneath the
11 code.
12
13 ```{python}
14 # -*- coding: utf-8 -*-
15 import sys
16 print (sys.version)
17 print("Hallöchen Python in der R Markdown Welt")
18 ```
```

The console output shows the following:

```
--:-- 569k 98 2107k 98 2076k 0 0 405k 0 0:00:
05 0:00:05 --:--:-- 405k 99 2107k 99 2092k 0 0 374
k 0 0:00:05 0:00:05 --:--:-- 415k100 2107k 100 2107k
0 0 363k 0 0:00:05 0:00:05 --:--:-- 68924
```

The console also displays the following text:

```
The downloaded binary packages are in
/var/folders/0f/9kzwypy12h7220556kxq54gh0000gn/T//Rtmp8vZzTv/do
wnloaded_packages
> library("rmarkdown", lib.loc="/Library/Frameworks/R.framework
/Versions/3.2/Resources/library")
Warnmeldung:
Paket 'rmarkdown' wurde unter R Version 3.2.5 erstellt
> plot(cars)
> plot(cars)
> plot(cars)
>
```

The screenshot shows a web browser displaying the rendered R Markdown notebook. The title is "R Notebook und Python". The content includes the same text as the RStudio editor, followed by the output of the Python code:

```
# -*- coding: utf-8 -*-
import sys
print (sys.version)
print("Hallöchen Python in der R Markdown Welt")
```

2.7.5 (default, Mar 9 2014, 22:15:05)  
[GCC 4.2.1 Compatible Apple LLVM 5.0 (clang-500.0.68)]  
Hallöchen Python in der R Markdown Welt

Aber müß Schwierigkeiten mit den deutschen Ümläuten, da RStudio statt auf mein Anaconda-Python 3 auf das System-Python von Apple zurückgreift. Und das ist ein 2.7.5.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

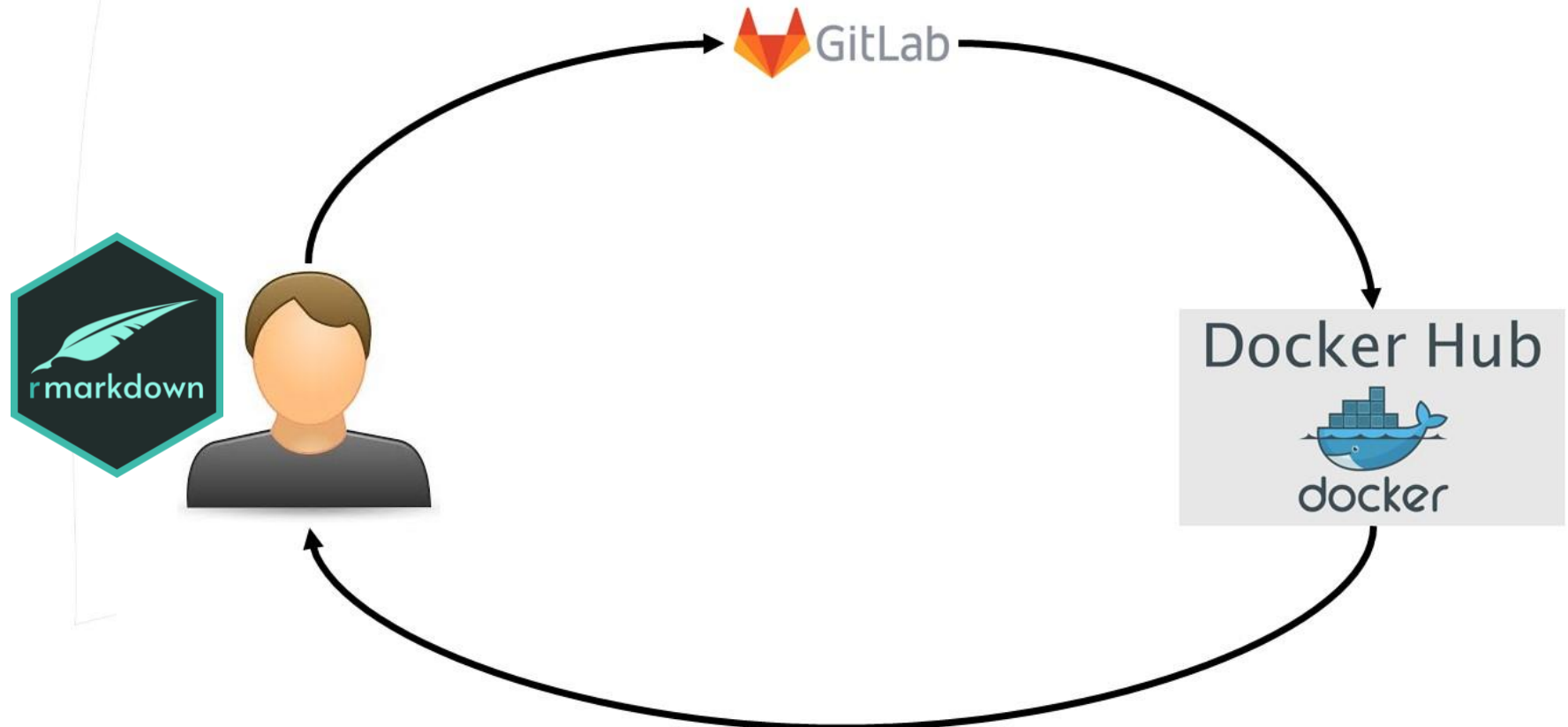
```
plot(cars)
```

The scatter plot shows the relationship between speed (x-axis, 5 to 25) and distance (y-axis, 0 to 120) for the 'cars' dataset. The plot shows a positive correlation, with distance increasing as speed increases.



Institut sociálního zdraví  
Univerzity Palackého

# RRE: R Reproducible Environment





Institut sociálního zdraví  
Univerzity Palackého

# Úložiště REE:

<https://gitlab.com/lukas.novak/rre>





Institut sociálního zdraví  
Univerzity Palackého

# Příklad studie založené na RRE

<https://osf.io/bn3w8/>



## Omezení a silné stránky

- **Silné stránky**
  - Není potřeba přístup k internetu pro práci v R
  - Vysoká výpočetní opakovatelnost
- **Omezení**
  - Zatím jen pro Windows
  - Jeden projekt v daný čas
  - Nízká podpora pro GitHub





## Reference

Calero Valdez, A. (2020). Making reproducible research simple using RMarkdown and the OSF. In Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis: 12th International Conference, SCSM 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22 (pp. 27-44). Springer International Publishing.

Davidson, H., Jabbari, Y., Patton, H., O'Hagan, F., Peters, K., & Cribbie, R. (2019). Statistical software use in Canadian university courses: Current trends and future directions. *Teaching of Psychology*, 46(3), 246-250.

Mikoska, P., Novák, L., Fülep, M., Bok, T., Pilarik, L., Ladmannova, M., & Kořínek, R. (2023, January 28). Standardization and psychometric characteristics of the Overall Depression Severity and Impairment Scale in the Czech Republic. Retrieved from [osf.io/bn3w8](https://osf.io/bn3w8)