datawizard: An R Package for Easy Data Wrangling and Transformations

Daniel Lüdecke[1], **Dominique Makowski**[2], **Mattan S. Ben-Shachar**[3], **Brenton M. Wiernik**[4], **Etienne Bacher**[5], **and Indrajeet Patil**[6]

1 University Medical Center Hamburg-Eppendorf, Germany **2** Nanyang Technological University, Singapore **3** Ben-Gurion University of the Negev, Israel **4** Facebook **5** Luxembourg Institute of Socio-Economic Research, Luxembourg **6** esqLABS GmbH

Summary

The {datawizard} package in the R programming language (R Core Team, 2021) provides a

Statement of Need

The {datawizard} package makes basic data wrangling easier than with base R. Its workflow and syntax are designed to be similar to {tidyverse} (Wickham et al. (2019)), which is a widely used ecosystem of packages, and, therefore, users familiar with this ecosystem can translate their knowledge and use this package. Naturally, one might wonder why recreate data wrangling functionality already present in {tidyverse}.

We believe the following points justify its *raison d'être*

- The easystats (Ben-Shachar et al. (2020); Lüdecke et al. (2020); Lüdecke, Ben-Shachar, et al. (2021); Lüdecke, Patil, et al. (2021); Lüdecke et al. (2019); Makowski et al. (2019); Makowski et al. (2020)) is an ecosystem of packages designed to make statistical analysis easier in R

Data wrangling

Table 1: The table below lists a few key functions offered by *datawizard* for data wrangling. To see the full list, see the package website: https://easystats.github.io/datawizard/

Function	Operation
data_filter()	to select only certain *observations*
data_select()	to select only a few *attributes*
data_extract()	to extract a single *attribute*
data_rename()	to rename attributes
reshape_longer()	to convert data from wide to long
reshape_wider()	to convert data from long to wide
data_join()	to join two data frames

Function	Operation
...	...

Data transformations

Table 2: The table below lists a few key functions offered by *datawizard* for data transformations. To see the full list, see the package website: https://easystats.github.io/datawizard/

Function	Operation
`standardize()`	to center and scale data
`normalize()`	to scale variables to 0-1 range
`adjust()`	to adjust data for effect of other variables
`data_shift()`	to shift numeric value range
`ranktransform()`	to convert numeric values to integer ranks
...	...

Licensing and Availability

datawizard is licensed under the GNU General Public License (v3.0), with all source code openly developed and stored at GitHub (https://github.com/easystats/datawizard), along with a corresponding issue tracker for bug reporting and feature enhancements. In the spirit of honest and open science, we encourage requests, tips for fixes, feature updates, as well as general questions and concerns via direct interaction with contributors and developers.

Acknowledgments

datawizard is part of the collaborative *easystats* ecosystem. Thus, we thank the members of easystats as well as the users.

References

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Lüdecke, D., Ben-Shachar, M. S., Patil, I., & Makowski, D. (2020). Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, *5*(53), 2445. https://doi.org/10.21105/joss.02445

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lüdecke, D., Patil, I., Ben-Shachar, M. S., Wiernik, B. M., Waggoner, P., & Makowski, D. (2021). see: An R package for visualizing statistical models. *Journal of Open Source Software*, *6*(64), 3393. https://doi.org/10.21105/joss.03393

Lüdecke, D., Waggoner, P., & Makowski, D. (2019). insight: A unified interface to access information from model objects in R. *Journal of Open Source Software*, *4*(38), 1412. https://doi.org/10.21105/joss.01412

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020). Methods and algorithms for correlation analysis in R. *Journal of Open Source Software*, *5*(51), 2306. https://doi.org/10.21105/joss.02306

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686