The Journal of Open Source Software

see: R Package for Visualizing Statistical Models

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

1 University Medical Center Hamburg-Eppendorf, Germany **2** Ben-Gurion University of the Negev, Israel **3** Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany **4** Department of Psychology, University of South Florida, USA **5** University of Chicago, USA **6** Nanyang Technological University, Singapore

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

The *easystats* is a collection of packages that operate in synergy to provide a consistent and intuitive syntax to work with statistical models in R programming language (R Core Team, 2021). Although these packages return comprehensive numeric summaries of model parameters and performance, complementing such summaries with graphical displays can be more informative. The *see* package provides the necessary tools (*ggplot2* geoms, themes, etc.) to visualize the analytic methods supported in this ecosystem.

Statement of Need

A number of data visualization tools exist in the *ggplot2* framework (Wickham, 2016) for graphically exploring statistical models[1]. A few of them provide ready-made plots or geometric layers for conveniently preparing common visualizations, but without linking them to any particular statistical analyses (e.g., *ggrepel*, *ggalluvial*, *ggridges*, *ggdist*, etc.). A few other packages *do* provide visualizations linked to particular statistical analyses (e.g., *ggpubr*, *tidymv*, *ggstatsplot*, *survminer*, etc.), but they tend to be more specialized. For example, the *ggstatsplot* package (Patil, 2021) offers visualizations only for statistical analysis of one-way factorial designs.

The *see* package, on the other hand, is designed to work with a wide range of statistical models and provides a much more comprehensive *ggplot2*-based visualization toolbox that covers all stages of statistical analysis and reporting. Additionally, as a member of the larger *easystats*-universe, it also embodies a common syntax and philosophy of this ecosystem, lowering the visualization barrier for users of this ecosystem.

Features

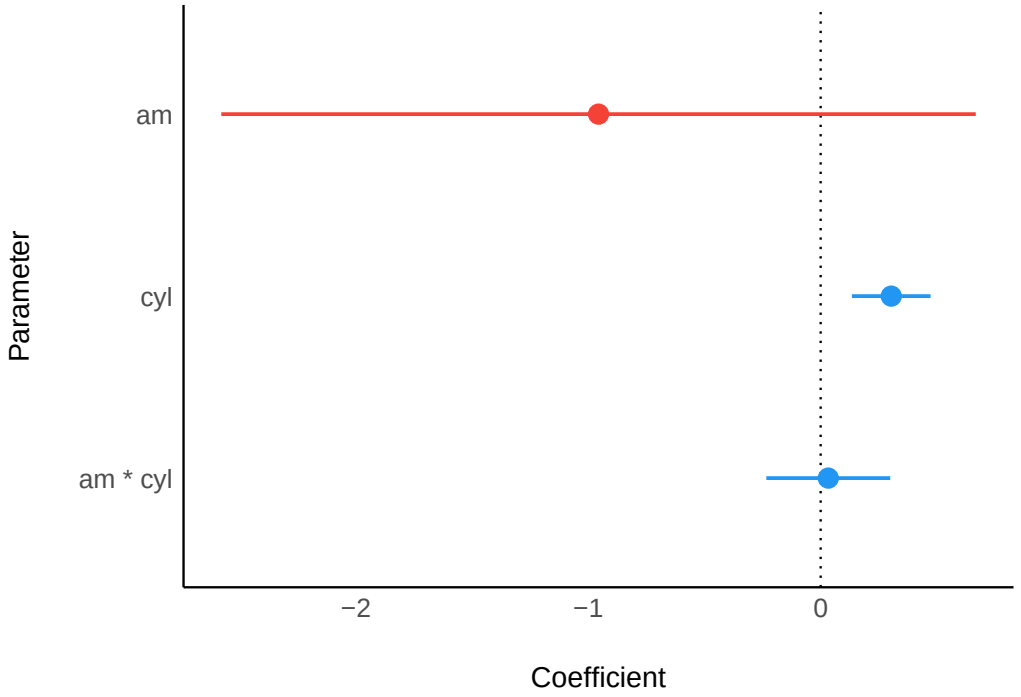
For the sake of brevity, we will discuss only one plotting method for each package. Several other methods are available and the readers are encouraged to see more examples on the package website: https://easystats.github.io/see/

Visualizing parameters of a model

The *parameters* package (Lüdecke, Ben-Shachar, Patil, & Makowski, 2020) converts regression model objects into dataframes. The *see* package can then create dot-and-whisker plots for the extracted regression estimates:

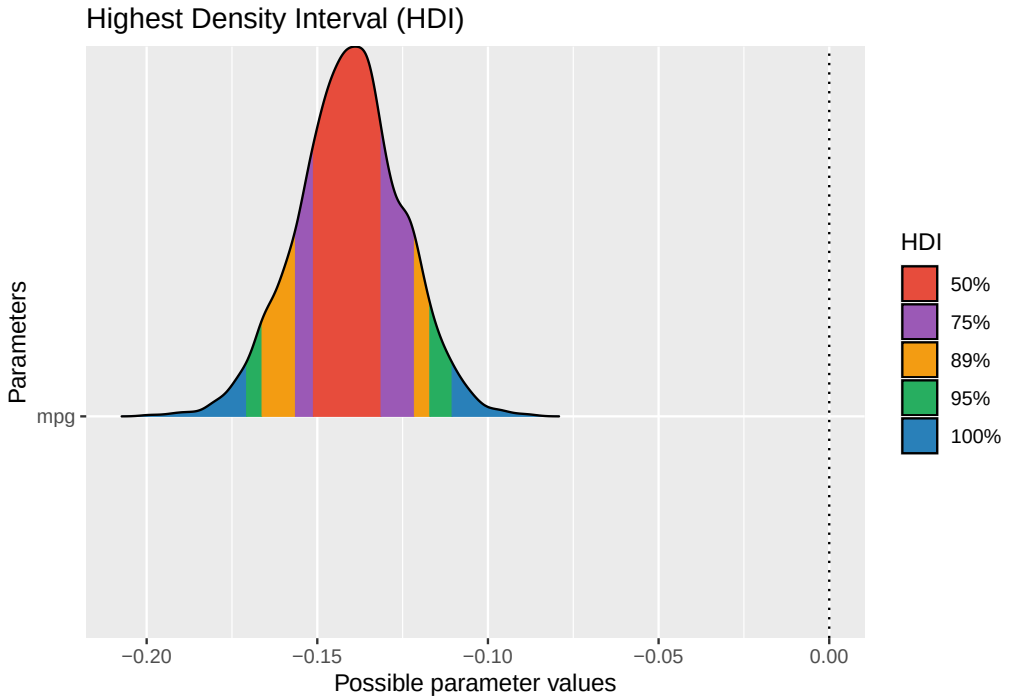
[1] For a comprehensive collection of available packages, see: https://exts.ggplot2.tidyverse.org/gallery/

```
library(parameters)
library(see)
mod <- lm(wt ~ am * cyl, mtcars)
plot(parameters(mod))
```



In case of a Bayesian regression model, which are handled by the *bayestestR* package (Makowski, Ben-Shachar, & Lüdecke, 2019), the *see* package also provides special plotting methods relevant only for Bayesian models (e.g., Highest Density Interval or HDI).

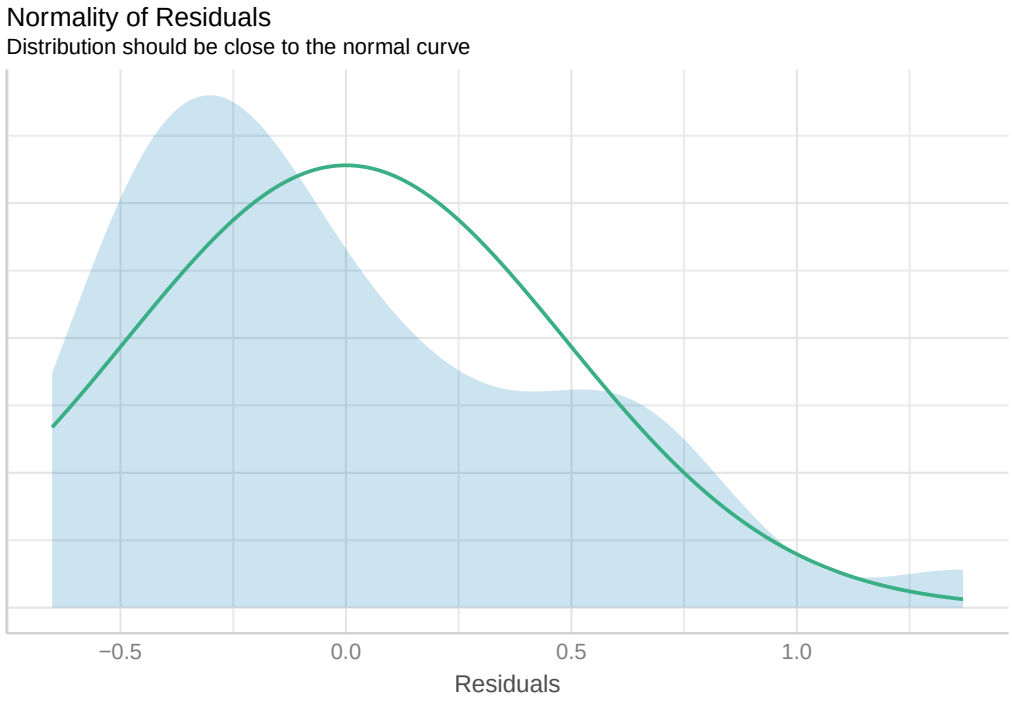
```
library(bayestestR)
library(rstanarm)
mod <- stan_glm(wt ~ mpg, data = mtcars, refresh = 0)
result <- hdi(mod, ci = c(0.5, 0.75, 0.89, 0.95))
plot(result)
```



Visualizing model fit

The *performance* package (Lüdecke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021) provides functions to check assumptions behind regression models. The *see* package can help visualize these assumptions checks (e.g., normality of residuals):

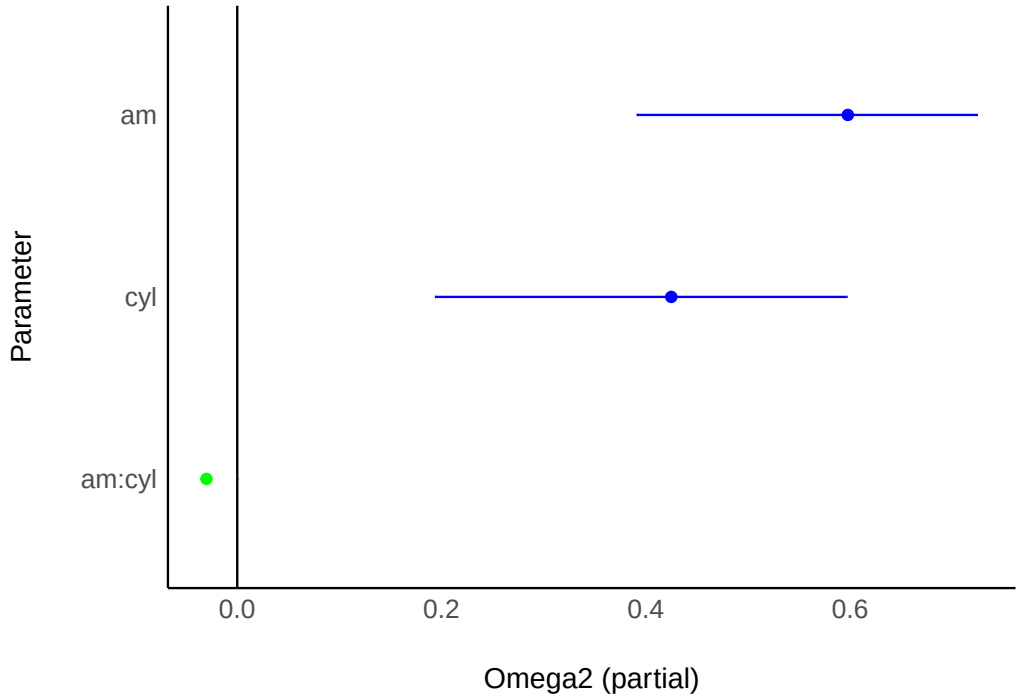
```
library(performance)
mod <- lm(wt ~ mpg, mtcars)
plot(check_normality(mod))
#> Warning: Non-normality of residuals detected (p = 0.016).
```



Visualizing effect sizes

In addition to providing tabular summaries of regression model objects, the *easystats* ecosystem also provides the *effectsize* package (Ben-Shachar, Lüdecke, & Makowski, 2020) to assess practical importance of observed effects by computing appropriate effect size measures. The *see* package can then be used to visualize the magnitude and uncertainty of these effect sizes:

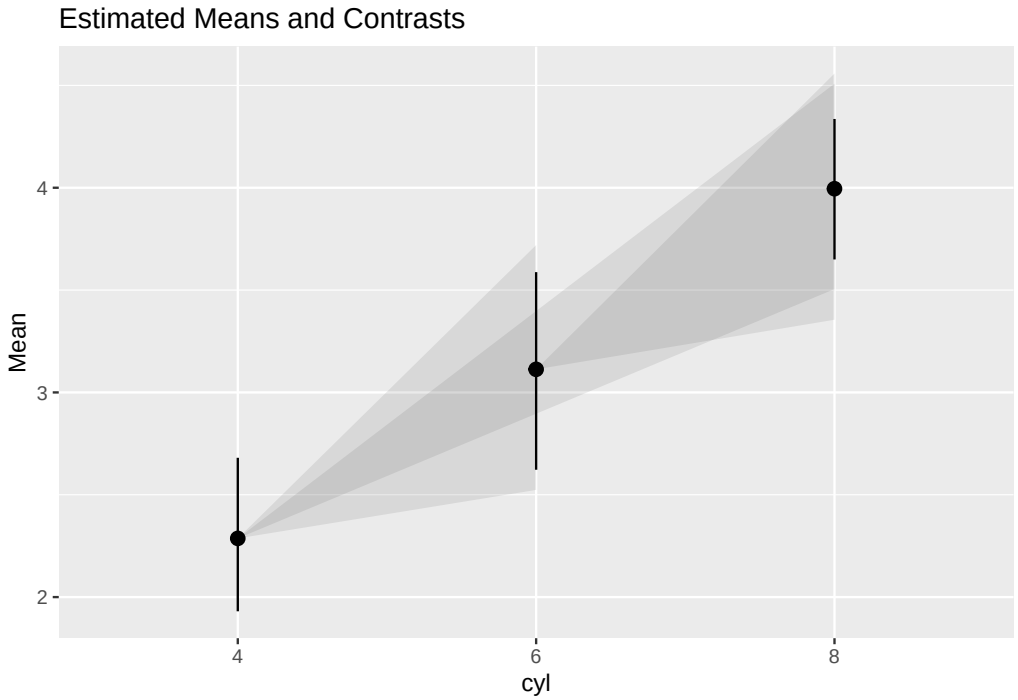
```
library(effectsize)
mod <- aov(wt ~ am * cyl, mtcars)
plot(omega_squared(mod))
```



Visualizing marginal effects

The *modelbased* package (Makowski, Ben-Shachar, Patil, & Lüdecke, 2020a) can be used to compute marginal means for regression models and the *see* package provides functions to visualize these marginal effects.

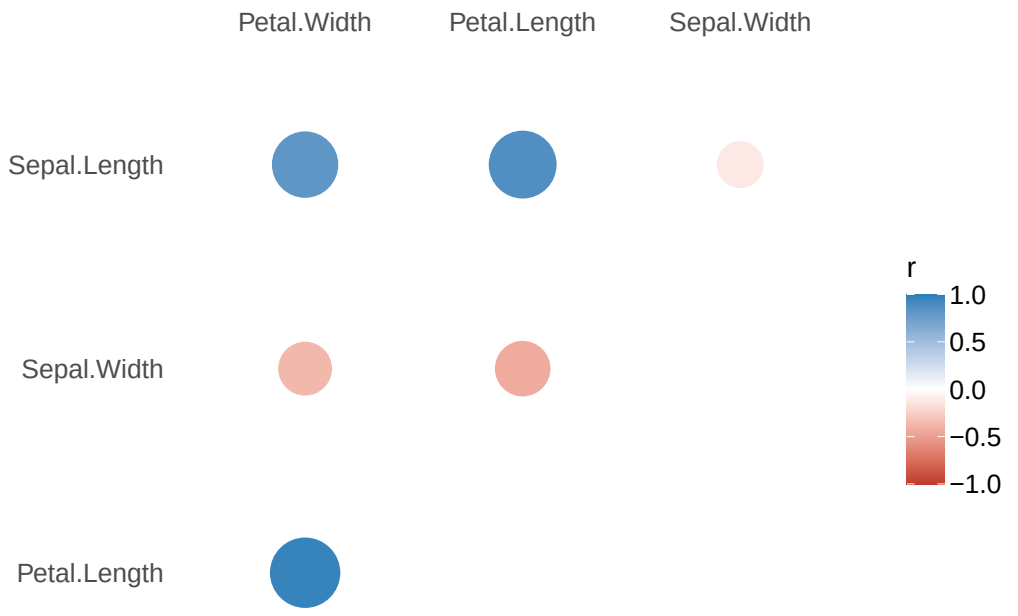
```
library(modelbased)

mod <- stan_glm(wt ~ as.factor(cyl), data = mtcars, refresh = 0)
contrasts <- estimate_contrasts(mod)
means <- estimate_means(mod)

plot(contrasts, means)
```



Visualizing correlation matrices

The *correlation* package (Makowski, Ben-Shachar, Patil, & Lüdecke, 2020b) provides a unified syntax to carry out various kinds of correlation analyses, and the *see* package can easily visualize these correlations in a matrix.

```
library(correlation)
result <- correlation(iris, type = "percentage")
plot(summary(result))
```



Licensing and Availability

see is licensed under the GNU General Public License (v3.0), with all source code stored at GitHub (https://github.com/easystats/see), and 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

see 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. doi:10.21105/joss.02815

Lüdecke, D., Ben-Shachar, M. S., Patil, I., & Makowski, D. (2020). parameters: Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, *5*(53), 2445. doi: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. doi:10.21105/joss.03139

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. doi:10.21105/joss.01541

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

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020a). Estimation of model-based predictions, contrasts and means. *CRAN*. Retrieved from https://github.com/easystats/modelbased

Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. *PsyArxiv*. doi:10.31234/osf.io/p7mku

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

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org