

Visualizations with statistical details: The ‘ggstatsplot’ approach

Indrajeet Patil¹

1 Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany

DOI:

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted:

Published:

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

Statement of Need

In a typical exploratory data analysis workflow, data visualization and statistical modeling are two different phases: visualization informs modeling, and modeling in its turn can suggest a different visualization method, and so on and so forth ([Wickham & Grolemund, 2016](#)). The central idea of `ggstatsplot` package in R programming language ([R Core Team, 2021](#)) is simple: combine these two phases into one in the form of graphics with statistical details. The package is designed to follow the best practices in **both** data visualization ([Cleveland, 1985](#); [Grant, 2018](#); [Healy, 2018](#); [Tufte, 2001](#); [Wilke, 2019](#)) and (Frequentist/Bayesian) statistical reporting ([Association & others, 1985](#); [Doorn et al., 2020](#)).

Before discussing benefits of this approach, we will see output from one of its functions to understand its behavior.

```
library(ggstatsplot)
library(palmerpenguins) # for 'penguins' dataset

ggbetweenstats(penguins, species, body_mass_g)
```

As can be seen, with a **single** line of code, the function produces details about descriptive statistics, inferential statistics, effect size estimate and its uncertainty, pairwise comparisons, Bayesian hypothesis testing, Bayesian posterior estimate and its uncertainty. Moreover, these details are juxtaposed with informative and well-labeled visualizations, designed to follow best practice guidelines from data visualization research. Without `ggstatsplot`, getting these statistical details and customizing a plot would require more than an hour of work, but, with `ggstatsplot`, it takes a little more than a few seconds. In other words, this package takes away *an* excuse from researchers to thoroughly explore their data - since it lowers the technical and time barriers to entry significantly - and instills good data sanitation/exploration habits.

All statistical analysis in the package is carried out via `statsExpressions` package ([Patil, 2021](#)) and `easystats` ecosystem of packages to achieve this ([Ben-Shachar, Lüdtke, & Makowski, 2020](#); [Lüdtke, Ben-Shachar, Patil, & Makowski, 2020](#); [Lüdtke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021](#); [Lüdtke, Waggoner, & Makowski, 2019](#); [Makowski, Ben-Shachar, & Lüdtke, 2019](#); [Makowski, Ben-Shachar, Patil, & Lüdtke, 2020](#)).

Benefits

This approach- (a) avoids errors in statistical reporting, (b) highlights the importance of the effect by providing effect size measures by default, (c) provides an easy way to

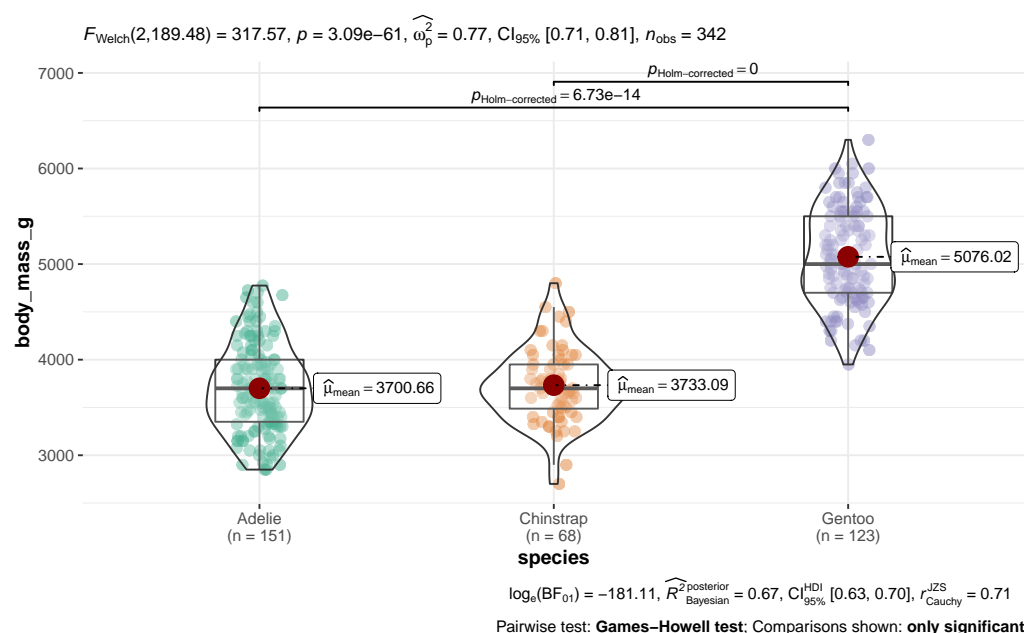


Figure 1: Example plot from the 'ggstatsplot' package illustrates its philosophy of juxtaposing informative visualizations with details from statistical analysis. To see all supported plots and statistical analysis, see the package website: <https://indrajeetpatil.github.io/ggstatsplot/>

evaluate *absence* of an effect using Bayesian framework, (*d*) forces to evaluate statistical assumptions behind chosen analysis in the context of the underlying data, and is (*e*) easy and (*f*) simple enough that somebody with little-to-no coding experience can use it without making an error.

Licensing and Availability

ggstatsplot is licensed under the GNU General Public License (v3.0), with all source code stored at [GitHub](#), 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, by [filing an issue](#). See the package's [Contribution Guidelines](#).

Acknowledgements

I would like to acknowledge the support of Mina Cikara, Fiery Cushman, and Iyad Rahwan during the development of this project. ggstatsplot relies heavily on the [easystats](#) ecosystem, a collaborative project created to facilitate the usage of R for statistical analyses. Thus, I would like to thank the [members](#) of [easystats](#) as well as the users. I would additionally like to thank the users of ggstatsplot for reporting bugs and providing helpful feedback.

References

Association, A. P., & others. (1985). *Publication manual of the american psychological association: American psychological association*. American Psychological Association.

- Ben-Shachar, M. S., Lüdtke, 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](https://doi.org/10.21105/joss.02815)
- Cleveland, W. S. (1985). *The Elements of Graphing Data* (1st edition.). Monterey, Cal: Wadsworth, Inc.
- Doorn, J. van, Bergh, D. van den, Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., et al. (2020). The JASP guidelines for conducting and reporting a bayesian analysis. *Psychonomic Bulletin & Review*, 1–14. doi:[10.3758/s13423-020-01798-5](https://doi.org/10.3758/s13423-020-01798-5)
- Grant, R. (2018). *Data Visualization: Charts, Maps, and Interactive Graphics*. CRC Press.
- Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton University Press.
- Lüdtke, 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](https://doi.org/10.21105/joss.02445)
- Lüdtke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Assessment, testing and comparison of statistical models using R. *Journal of Open Source Software*, 6(59), 3112. doi:[10.31234/osf.io/vtq8f](https://doi.org/10.31234/osf.io/vtq8f)
- Lüdtke, 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. doi:[10.21105/joss.01412](https://doi.org/10.21105/joss.01412)
- Makowski, D., Ben-Shachar, M. S., & Lüdtke, 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](https://doi.org/10.21105/joss.01541)
- Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdtke, D. (2020). Methods and algorithms for correlation analysis in R. *Journal of Open Source Software*, 5(51), 2306. doi:[10.21105/joss.02306](https://doi.org/10.21105/joss.02306)
- Patil, I. (2021). statsExpressions: R package for tidy dataframes and expressions with statistical details. *Journal of Open Source Software*, 6(59), 3111. doi:[10.31234/osf.io/ntbvy](https://doi.org/10.31234/osf.io/ntbvy)
- 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/>
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd edition.). Cheshire, Conn: Graphics Press.
- Wickham, H., & Golemund, G. (2016). *R for Data Science*. O'Reilly Medias.
- Wilke, C. O. (2019). *Fundamentals of Data Visualization*. O'Reilly Media.