The Journal of Open Source Software

ggmatplot: An R package for data visualization on wide-format data

Xuan Liang[1], **Francis Hui**[1], **Dilinie Seimon**[2], **and Emi Tanaka**[2]

1 Australian National University **2** Monash University

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

The layered grammar of graphics (H. Wickham, 2010), implemented as the `ggplot2` package (Hadley Wickham, 2016) in the statistical language R (R Core Team, 2021), is a powerful tool to create versatile statistical graphics. This graphical system, however, requires input data to be organised in a manner that a data column is mapped to an aesthetic element (e.g. x-coordinate, y-coordinate, color, size), which create friction in constructing plots with an aesthetic element that span multiple columns in the data by requiring users to re-organise the data.

The `ggmatplot`, built upon `ggplot2`, is an R-package that allows quick plotting across the columns of matrics or data with the result returned as a `ggplot` object. The package is inspired by the function `matplot()` in the core R `graphics` system, thus `ggmatplot` can be considered as a `ggplot` version of `matplot` with the benefits of customising the plots as any other `ggplot` objects via `ggplot2` functions.

Statement of need

Data can be tidied in a rectangular form where each row represents an observational unit, each column represents a variable, and each cell represents a value (Hadley Wickham, 2014). What constitutes a variable, hence a column, in a tidy data are dependent upon interpretation.

The increased use of a tidy approach in structuring datasets results in wrangling datasets into wide format: with variables as columns and observations as rows(https://vita.had.co.nz/papers/tidy-data.html). While this introduces a more meaningful mapping of the data into its structure, the grammar of graphics requires wide format data to be wrangled back into long format for visualizing. This can be quite cumbersome when creating simple plots, and converting a tidy data frame back to a long format with multiple variables stacked upon each other can also be considered counter-intuitive.

Therefore, the motivation for `ggmatplot` is to provide a solution that allows `ggplot2` to handle wide format data. Although `ggmatplot` doesn't provide the same flexibility as `ggplot2`, it can be used as a workaround for having to wrangle wide format data into long format and creating simple plots using `ggplot2`.

Examples

Acknowledgements

References

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. (2010). A layered grammar of graphics. *Journal of computational and graphical statistics: a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*.

Wickham, Hadley. (2014). Tidy data. *Journal of Statistical Software, Articles*, *59*(10), 1–23. doi:10.18637/jss.v059.i10

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