

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

Xuan Liang¹, Francis K. C. Hui¹, Dilinie Seimon², and Emi Tanaka²

DOI:

1 The Australian National University 2 Monash University

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

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 and popular 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 original 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 matrices 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

Input data to construct plots with `ggplot2` require data to be organised in a manner that maps data columns to aesthetic elements. This required form is, however, not consistent with some prevalent data formats, specifically multivariate data where the rows correspond to observational units and contiguous columns correspond to the responses of the same measurement units. Examples of common utilization of this form include gene expression data in bioinformatics, where rows correspond to samples, column names correspond to genes, and cell values correspond to the expression level; and multi-abundance data in ecology, where rows correspond to site, column names correspond to species name, and cell values correspond to count. Consequently, plotting with `ggplot2` interrupts the workflow of a user that is trying to quickly explore these types of data. The `ggmatplot` R-package provides a solution to this common friction in producing plots with `ggplot2`.

Examples

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). In some cases, what constitutes a variable (or observational unit), hence a column (or row), in a tidy data can be dependent upon interpretation or downstream interest (e.g. Tables 1 and 2 can be both considered as tidy data), but a clear violation of tidy data principles is when the column names contain data values, e.g. Table 3 contain the name of the species across a number of column names.

```
library(ggmatplot)
ggmatplot(x = select(wide_df, contains("rating")),
```

Table 1: Restaurant rating data in "tidy" form

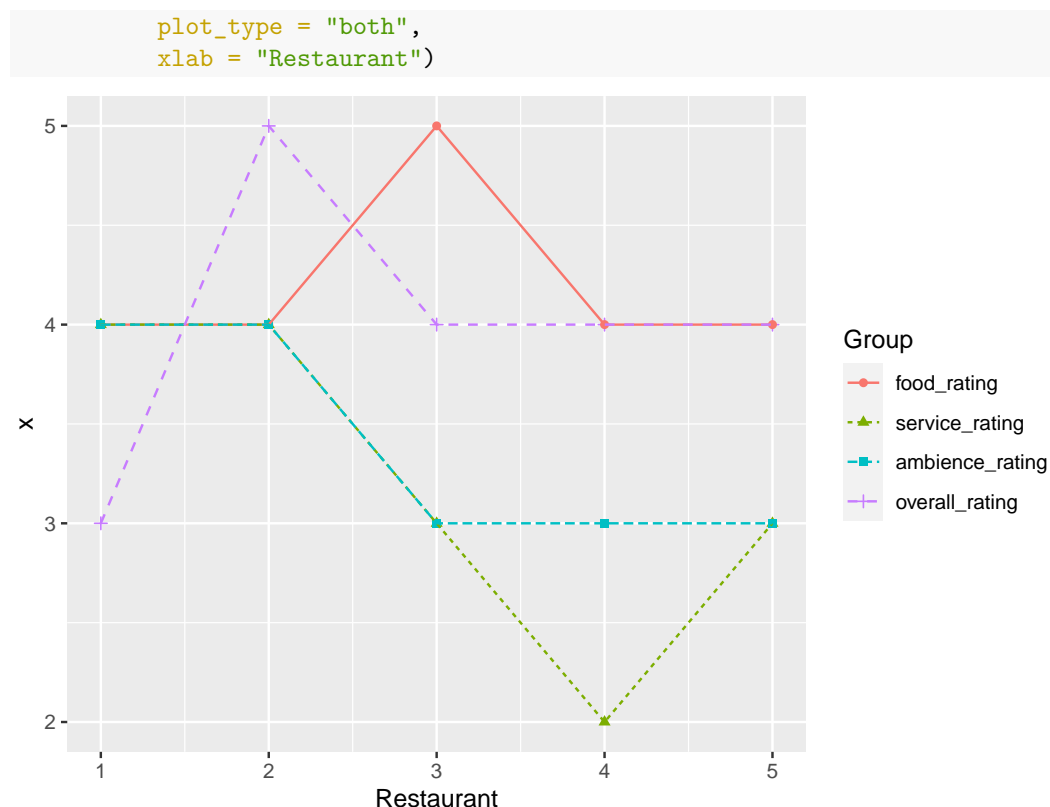
Restaurant	Rating			
	Food	Service	Ambience	Overall
R1	4	3	4	4
R2	4	5	4	4
R3	3	4	5	3
R4	2	4	4	3
R5	3	4	4	3

Table 2: Restaurant rating data in "molten" form

Restauant	Rating type	Rating
R1	food	4
R1	service	3
R1	ambience	4
R1	overall	4
R2	food	4
R2	service	5
R2	ambience	4
R2	overall	4
R3	food	3
R3	service	4
R3	ambience	5
R3	overall	3
R4	food	2
R4	service	4
R4	ambience	4
R4	overall	3
R5	food	3
R5	service	4
R5	ambience	4
R5	overall	3

Table 3: Spider abundance data with environmental covariates.

Site	Environment covariates		Species abundance				
	Soil dry mass	Moss	Alopcune	Arctlute	Pardpull	Trocterr	Zoraspin
1	2.3321	3.0445	10	0	45	57	4
2	3.0493	1.0986	2	0	37	65	9
3	2.5572	2.3979	20	0	45	66	1



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](https://doi.org/10.18637/jss.v059.i10)
- Wickham, Hadley. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>