

# Advancing Exploratory Data Analysis in R: The gg1d Package for Multidimensional Data Visualization

Sam El-Kamand<sup>1</sup>, Julian M. W. Quinn<sup>1</sup>, and Mark J. Cowley<sup>1,2</sup>

<sup>1</sup> Children's Cancer Institute, Australia <sup>2</sup> School of Clinical Medicine, UNSW Medicine & Health, Australia ¶ Corresponding author

DOI: 10.xxxxxx/draft

## Software

- Review
- Repository
- Archive

Editor: Open Journals

## Reviewers:

- @openjournals

Submitted: 01 January 1970

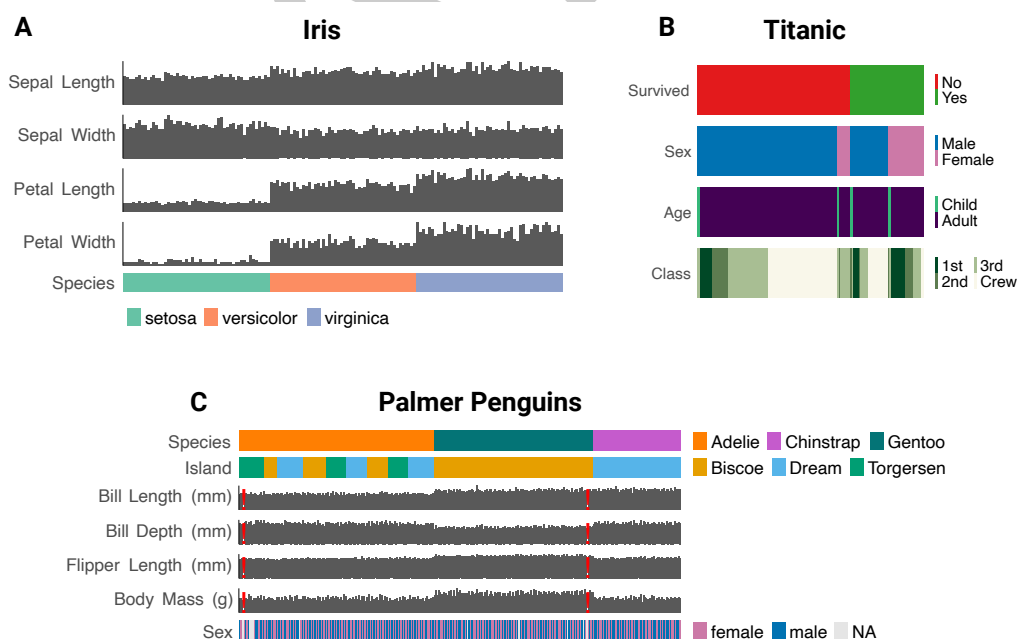
Published: unpublished

## License

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

## Summary

Exploratory data analysis (EDA) involves examining the relationships between both categorical and quantitative features. The gg1d R package streamlines EDA by providing a turnkey approach to visualising n-dimensional data which graphically reveals correlative or associative relationships between 2 or more features (Figure 1). gg1d represents all dataset features as distinct, vertically aligned bar or tile plots, with plot types auto-selected based on whether variables are categorical or numeric. It reduces both the code and time required to detect complex multi-feature relationships that would otherwise only be found through statistical modelling or thorough manual review (Figure 2, Figure 3).



**Figure 1:** gg1d visualizations of common datasets revealing: A) Petals of the setosa species of iris are drastically smaller than other iris species; B) The majority of individuals who perished during the Titanic disaster were adult males; C) *Gentoo* penguins from Biscoe Island (dark green) have shallower bill depths than *Chinstrap* or *Adelie* penguins, despite their increased body mass. Exclamation marks indicate missing values.

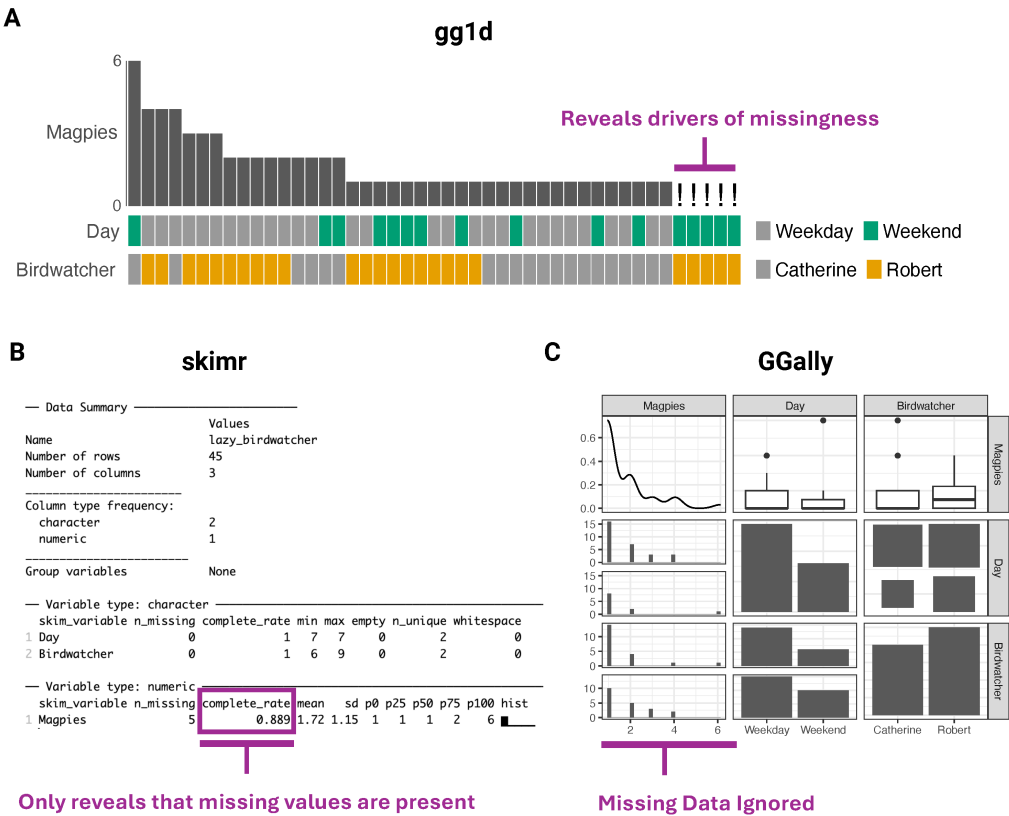
15    **Statement of Need**

16    The R ecosystem already includes popular EDA packages such as `skimr`, which textually  
17    summarizes completeness and descriptive statistics for individual features (1-dimensional), and  
18    `GGally`, which graphically describes pairwise feature correlations (2-dimensional). `gg1d` is an  
19    n-dimensional generalization with key advantages over other EDA packages, most notably its  
20    ability to reveal more complex multidimensional patterns (Figure 2, Figure 3).

Feature	gg1d	Complex Heatmap	Data Explorer	skimr	GGally
Automatic Plot Generation	✓	✗	✓	✓	✓
Automatic plot selection by variable type	✓	✗	✓	✗	✓
Interactive Visualisations	✓	✗	✗	✗	✗
Supports cross-linking with other datasets	✓	✗	✗	✗	✗
Composable with Patchwork	✓	✓	✗	✗	✓
Describes features contribution to total variance (PCA)	✗	✗	✓	✗	✗
Generates Publication Quality Figures	✓	✓	✗	✗	✓
Reveals missingness dependent on multiple features	✓	✗	✗	✗	✗

**Figure 2:** Comparison of R packages that create visualisations commonly used for exploratory data analysis.

21    The benefits of `gg1d` are exemplified when visualizing the artificial Lazy Birdwatcher dataset,  
22    which records magpie observations by two birdwatchers (Figure 3). One birdwatcher does  
23    not work on weekends, creating a missing data pattern dependent on both birdwatcher and  
24    day of the week. This multidimensional pattern becomes immediately apparent from `gg1d`  
25    output, whereas it is difficult to detect using only one-dimensional EDA tools like `skimr` or  
26    two-dimensional tools like `ggpairs` from the `GGally` package.



**Figure 3:** Visualisation of the Lazy Birdwatcher dataset using the **gg1d** package reveals a pattern of missingness dependent on multiple variables, Birdwatcher and Day (A). This pattern is difficult to detect using one-dimensional EDA tools like **skimr** (B) or two-dimensional tools like **ggpairs** from the **GGally** package (C).

We developed **gg1d** for the visualisation of clinical and multiomics data and anticipate it will prove valuable for any exploratory EDA activities. Further examples of **gg1d** visualisations are available in the [gg1d gallery](#).

## Acknowledgements

We thank the developers of the packages integral to **gg1d**, especially David Gohel for **ggiraph** (Gohel & Skintzos, 2024), which enables its interactivity, and Thomas Lin Pedersen for **patchwork** (Pedersen, 2024) and **ggplot2** maintenance. We also acknowledge Hadley Wickham and all contributors to **ggplot2** (Wickham, 2016).

## References

- Gohel, D., & Skintzos, P. (2024). *Ggiraph: Make 'ggplot2' graphics interactive*. <https://doi.org/10.32614/CRAN.package.ggiraph>
- Pedersen, T. L. (2024). *Patchwork: The composer of plots*. <https://doi.org/10.32614/cran.package.patchwork>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://doi.org/10.32614/CRAN.package.ggplot2>