The Journal of Open Source Software

funkyheatmap: Visualising data frames with mixed data types

Robrecht Cannoodt*, Louise Deconinck*, Artuur Couckuyt*, Nikolay S. Markov*, Luke Zappia, Malte D. Luecken, Marta Interlandi, Yvan Saeys†, and Wouter Saelens†

VIB Center for Inflammation Research

Ghent University

Data Intuitive

École Polytechnique Fédérale de Lausanne (EPFL)

University of Muenster

Northwestern University

Helmholtz Munich

Technical University of Munich

Institute of Lung Health and Immunity (LHI), Helmholtz Munich, Comprehensive Pneumology Center (CPC-M), Germany; Member of the German Center for Lung Research (DZL)

*: Shared first authors, †: Shared last authors.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

funkyheatmap: Visualising data frames with mixed data types

Summary

The {funkyheatmap} package offers a flexible and user-friendly solution for visualising data frames containing a mixture of categorical, proportional, and text-based data. It simplifies the creation of informative and visually appealing heatmaps while providing extensive customization options to tailor the output. This tool is especially valuable in research settings for summarising and communicating complex results, such as those encountered in benchmarking studies.

The package is available on CRAN and PyPI and has a JavaScript port in development. For detailed examples and vignettes, visit the project website funkyheatmap.github.io.

Statement of need

Data visualisation is fundamental to exploratory data analysis and communicating findings. While powerful tools like ggplot2 (Wickham, 2009), Matplotlib (Hunter, 2007) and D3.js (Bostock et al., 2011) exist, they often require complex scripting to generate comprehensive visualisations for data frames containing a mix of data types. {funkyheatmap} addresses this challenge by:

- **Seamless Handling of Mixed Data**: Automates the selection of appropriate visualisations (rectangles, bars, pie charts, text) based on data types.

- **Customization**: Provides granular control over colours, groupings, geometries, and annotations for tailored results.
- **Accessibility**: Offers a simplified interface for basic use and detailed documentation for advanced customization.

{funkyheatmap} has proven its utility in benchmarking studies within single-cell omics (Li et al., 2023; Luecken et al., n.d., 2022; Saelens et al., 2019; Sang-Aram et al., 2024; Yan & Sun, 2023), but its applications extend to diverse fields where visualisation of mixed data types is needed.

Functionality

Figure 1 showcases the functionality of {funkyheatmap}, namely:

- **Diverse Geometries**: Supports a range of geometries (rectangles, bars, pie charts, text, images) to effectively represent different data types.
- **Hierarchical Categorical Grouping**: Facilitates the organisation of rows and columns into semantic groups with distinct colour palettes.
- **Documentation and Testing**: Includes comprehensive documentation, vignettes, and a test suite for quality and ease of use.

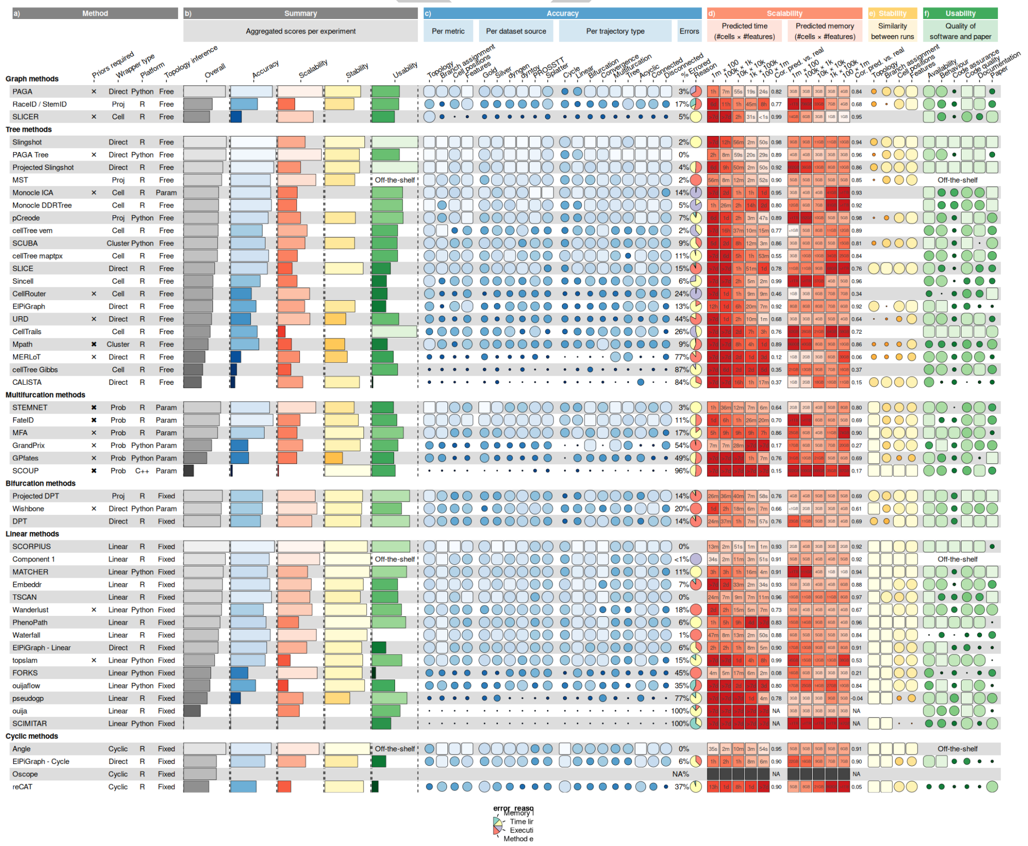

Figure 1: An example of a {funkyheatmap} visualisation using data from a benchmarking study of trajectory inference methods (Saelens et al., 2019).

See Table for more information regarding the recommended geom for different types of data.

Data type	Example	Recommended geom
Numerical data	Scores from 0 to 1	funkyrect
Aggregated data	The mean of scores	bar
Measurement data	3MB or 4h	rect + text overlay
Categorical data	R or Python	text or image
Proportional data	80% success, 10% OOM, 10% failed	pie

Recommended geometries in `{funkyheatmap}` for different data types. The table presents the suggested visualisation methods (geoms) based on the data type of the columns. These recommendations provide a starting point for users to select the most appropriate visual representation for their specific data.

Conclusion

`{funkyheatmap}` streamlines the creation of publication-quality visualisation for mixed data types, empowering researchers and data scientists to communicate their results effectively. The ongoing development of `funkyheatmappy` (Python) and `funkyheatmapjs` (JavaScript) will further expand the accessibility and functionality of this visualisation solution.

Acknowledgements

L.Z. acknowledges support from the Bavarian Ministry of Science and the Arts in the framework of the Bavarian Research Association "ForInter" (Interaction of human brain cells). L.D. acknowledges support from the Research Foundation - Flanders (FWO) (1SF3822N). Y.S. acknowledges support from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" program. A.C. and Y.S. are funded by an FWO TBM grant (Research Foundation – Flanders, T000119N), and Ghent University Special Research Fund (BOF18-GOA-024). W.S. was supported by a Marie Skłodowska-Curie fellowship (101028476).

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(03), 90–95. https://doi.org/10.1109/MCSE.2007.55

Li, H., Zhou, J., Li, Z., Chen, S., Liao, X., Zhang, B., Zhang, R., Wang, Y., Sun, S., & Gao, X. (2023). A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nature Communications*, *14*(1), 1548. https://doi.org/10.1038/s41467-023-37168-7

Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Müller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colomé-Tatché, M., & others. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, *19*(1), 41–50. https://doi.org/10.1038/s41592-021-01336-8

Luecken, M. D., Gigante, S., Burkhardt, D. B., Cannoodt, R., Strobl, D. C., Markov, N. S., Zappia, L., Palla, G., Lewis, W., Dimitrov, D., & others. (n.d.). Defining and benchmarking open problems in single-cell analysis. *Research Square*. https://doi.org/10.21203/rs.3.rs-4181617/v1

Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature Biotechnology*, *37*(5), 547–554. https://doi.org/10.1038/s41587-019-0071-9

Sang-Aram, C., Browaeys, R., Seurinck, R., & Saeys, Y. (2024). Spotless, a reproducible pipeline for benchmarking cell type deconvolution in spatial transcriptomics. *Elife*, *12*, RP88431. https://doi.org/10.1101/2023.03.22.533802

Wickham, H. (2009). *ggplot2*. Springer. https://doi.org/10.1007/978-0-387-98141-3

Yan, L., & Sun, X. (2023). Benchmarking and integration of methods for deconvoluting spatial transcriptomic data. *Bioinformatics*, *39*(1), btac805. https://doi.org/10.1093/bioinformatics/btac805