A review of interactive graphics for data exploration and analysis

Tom Elliott

iNZight Analytics Ltd

Iwi affiliations: Ngāti Whakaue, Ngāti Raukawa

Report published 7 February 2025

Available from https://inzight.co.nz/files/Review_of_interactive_graphics_2025.pdf

Abstract

Interactive data visualization has become fundamental to modern data analysis, enabling deeper insights and more intuitive exploration of complex datasets. This paper reviews the evolution and current state of interactive graphics for data exploration and analysis, with particular focus on R-based tools and web technologies. We examine existing frameworks and tools, discussing their strengths and limitations in facilitating interactive data analysis. Building on this review, we propose a new ecosystem that leverages R and web technologies to create privacy-preserving interactive visualization tools, addressing key challenges in secure data access and analysis. Our proposed framework emphasizes the separation of front-end visualization from back-end data processing, enabling fine-grained access control while maintaining rich interactive capabilities for data exploration and statistical analysis.

Supported by the Ngā Puanga Pūtaiao Fellowships from Government funding, administered by the Royal Society Te Apārangi.

1 Introduction

Over the past decades, visualization tools have evolved significantly, shifting from static desktop applications to dynamic web-based platforms. While traditional statistical graphics remain valuable, the emergence of interactive visualization tools has opened new possibilities for data exploration and analysis. Interactive visualization plays a critical role in modern data analysis, as highlighted by numerous researchers including Cook (2007), Theus and Urbanek (2014), Ward (2015), and Young (2011).

The rise of web technologies has accelerated this evolution, making interactive graphics more accessible and easier to distribute. Modern browsers now support sophisticated visualization capabilities through technologies such as SVG, Canvas, and WebGL, complemented by a plethora of JavaScript libraries, for example `D3.js`, that allow developers to create complex interactive graphics. However, many current web-based data visualizations lack a strong analytics engine to drive real-time analysis, focusing instead on creating individual visualizations primarily for storytelling purposes, as often seen in data journalism.

A key feature of interactive data analysis (also referred to as *visual analysis*) is linked graphics. However, linked graphics have become limited in modern data visualisation libraries, and Bartonicek (2024a) gives some reasons for this including changes to how graphics are constructed by combining components, rather than being a single object as they used to be. Another possible reason is that the tools have been developed by web developers rather than statisticians, and the goal was only ever to allow developers to *create* custom charts, rather than a focus on exploration.

This review focuses on interactive graphics for data exploration and analysis, examining free tools and frameworks with particular attention to R and web technologies, addressing the technical aspects of crafting interactive graphics rather than the graphics themselves. We will see the features common to each, as well as identify any unique features. These will then be used to inform the design of a new "framework" that we are working on.

2 Software for interactive data visualization

Interactive data analysis extends beyond mere visualization to encompass entire analytical workflows, including iterative exploration, model fitting, result interpretation, and sharing. Interactive graphics serve as interfaces for statistical computation, allowing users to interact directly with the data rather than just the visual representation.

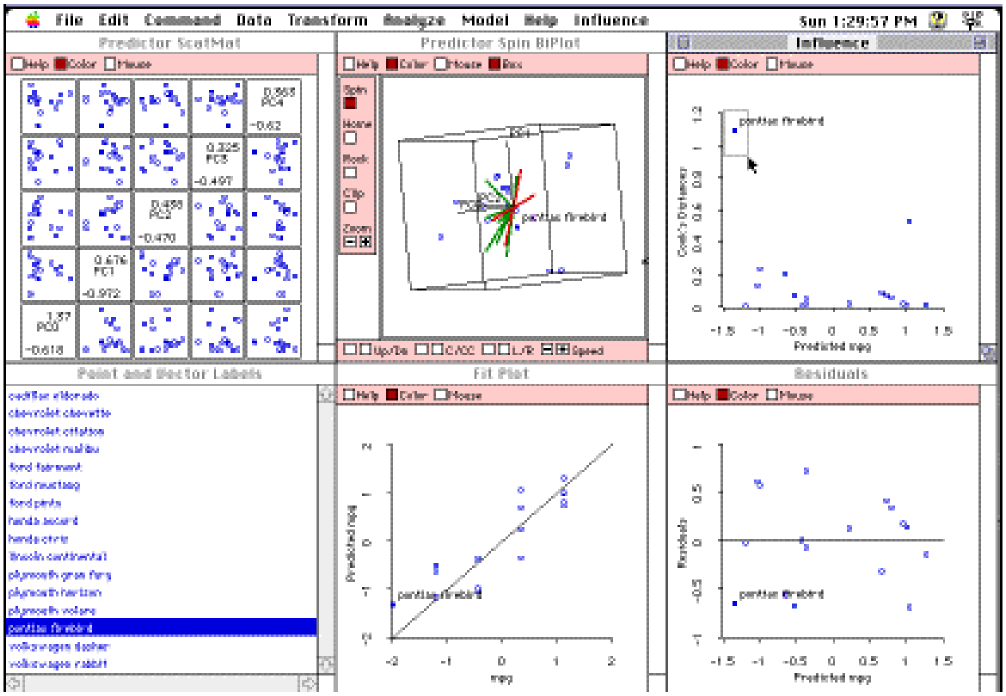

Figure 1: `ViSta` *SpreadPlot* for a regression model. Source: https://www.visualstats.org/vista-frames/statvis.html.

2.1 ViSta (Young & Bann, 1996)

`ViSta` is a desktop-based system that is based on the core concepts of *Visual Statistics* (hence the name). It features fully dynamic and interactive statistical visualisations that are designed to facilitae users in better understanding their data.

`ViSta` does many standard interactive tasks, such as linked graphics (where selecting points or objects in one plot highlights them in another). It also includes several unique features:

- *Workmaps*, which are visual representations of data analysis sessions. Data transformations, analyses, and results are displayed in a directed graph, and users are able to move back and forth through the analysis steps.

- *SpreadPlot*, which is a linked set of interactive graphics that allow users to explore a particular feature or object. Figure 1 shows an example of a SpreadPlot for a regression model.

`ViSta` is a desktop-only application, and the last update was in 2014. This makes it fairly impractical for modern environments, as support and cross-plotform availability are top concerns when choosing software. It also uses a very outdated GUI toolkit, which makes it look very dated.

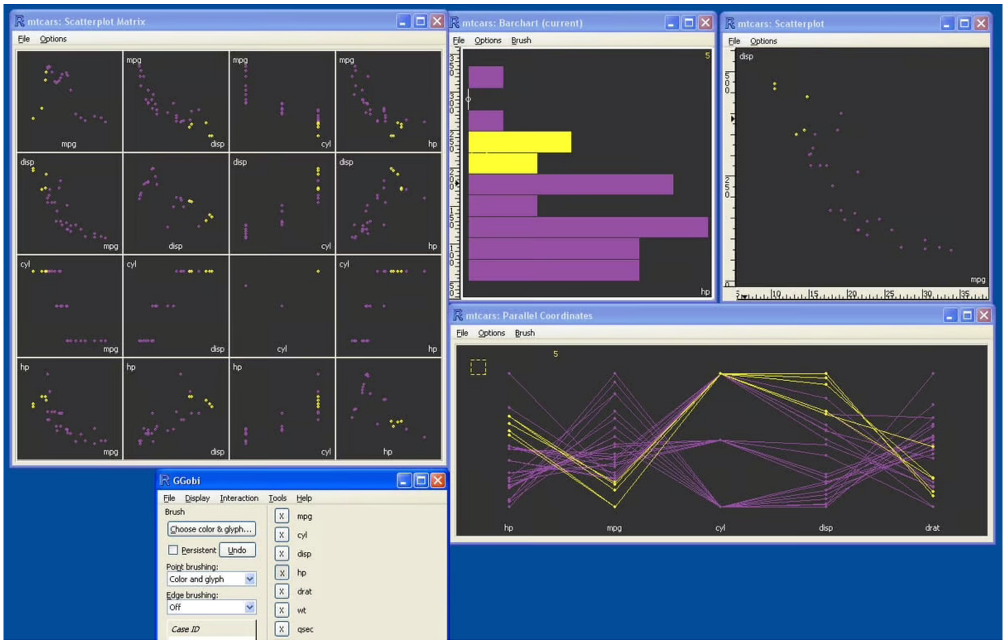

Figure 2: `GGobi` interface windows. Source https://www.datacamp.com/doc/r/interactive.

2.2 GGobi (Swayne et al., 2003)

The open-source package `GGobi` is a desktop-based system that provides dynamic and interactive graphics for exploring high-dimensional data. It focuses on highly dynamic and interactive visualisations, such as *tours*, a way of looking at multi-dimensional data. I also supports other traditional features such as brushing, linked graphics, and identification. Figure 2 shows a panel of linked graphics, including a pairs plot (top left), histogram (top center), scatter plot (top right), and parallel coordinates plot (bottom), as well as a control bannel at the bottom of the screen. Like `ViSta`, `GGobi` is a desktop-only application. The last update was in 2012, and the companion R package `rggobi` was archived from CRAN[1] in 2020.

2.3 Mondrian (Theus & Urbanek, 2014)

Similar to `GGobi`, `Mondrian` is a desktop-based system that provides interactive graphics for data exploration, and has many of the same features. Again, the focus is on interactive graphics, and specifically excels at categorical data, geographical data, and large datasets. Figure 3 shows the `Mondrian` interface, with a scatter plot (left), histogram (top right) and bar plot (bottom right). The software is discributed as a `Java .jar` file, and so it is limited to desktop use. There have been no feature updates since approximately 2011.

[1]The *Comprehensive R Archive Network, https://cran.r-project.org*.

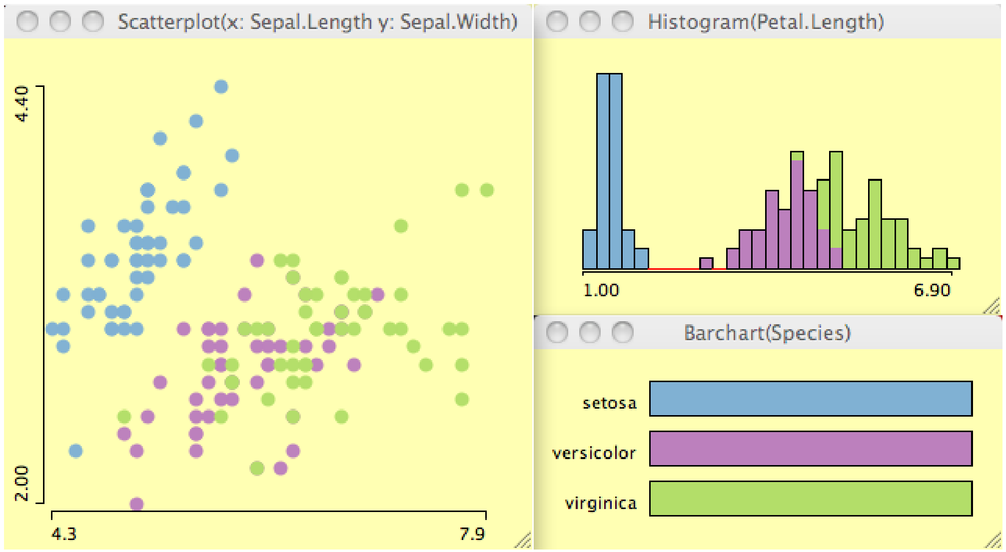

Figure 3: `Mondrian` windows. Source: http://mondrian.theusrus.de/.

2.4 Plotscaper (Bartonicek, 2024b)

A more recent addition to the landscape is `Plotscaper`, a web-based system that allows users to create linked graphics using R. The graphics are displayed in a web browser using TypeScript (a modern reimagining of JavaScript with type safety and many other developer-friendly features). The package is an implementation of new theoretical work on linking not only standalong graphics (e.g., a bar plot) but complex graphics consisting of various components (Bartonicek et al., 2025), as one might build using libraries such as `ggplot2` (Wickham, 2016).

The software supports both standalone web files and R web server deployment. Figure 4 shows an example of a `Plotscaper` landscape, with severl plots linked together with a (single) selection highlighted.

2.5 iNZight (Elliott & Wild, 2023)

The data analytics software `iNZight` is different from the others mentioned in that it does not feature any of the truly "interactive graphics" featuers. However, it does allow users to explore a dataset through individual, static graphics, and more importantly it forms the starting point for the present long-term project.

A key feature of `iNZight` is that it is *variable first*, meaning that the user selects a variable to explore, and the software then displays the appropriate graphics and summaries. This is in contrast to the other software, which typically require users to choose a plot or analysis type first. This can add an barrier to entry for users who are not familiar with data anlysis, such as students or researchers.

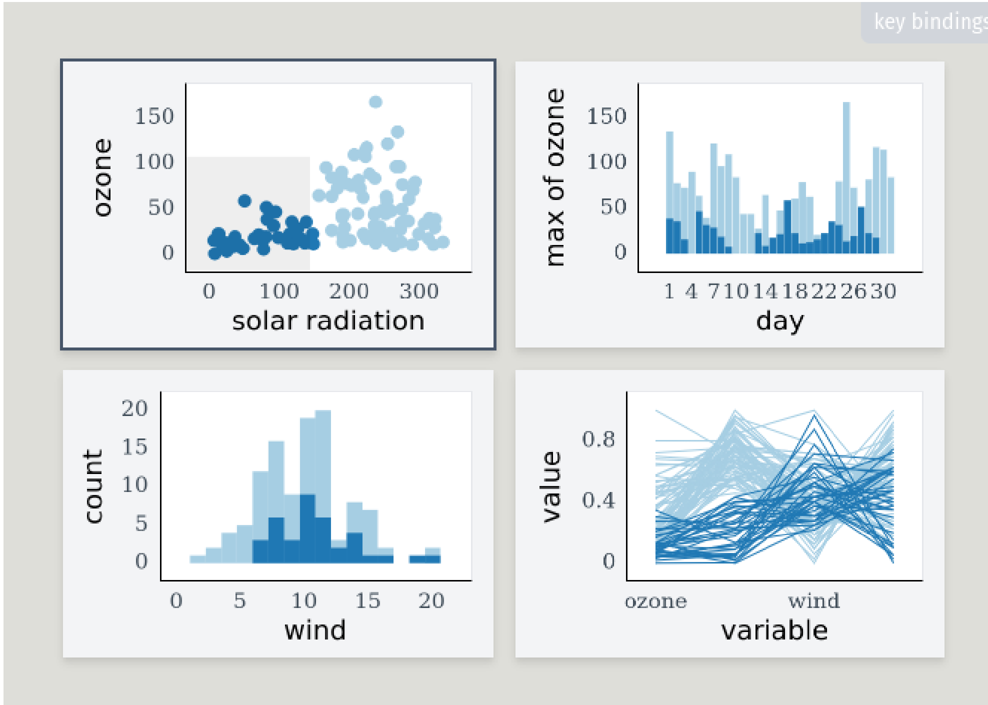

Figure 4: `Plotscaper` plots. Source: https://bartonicek.github.io/plotscaper/articles/plotscaper.html.

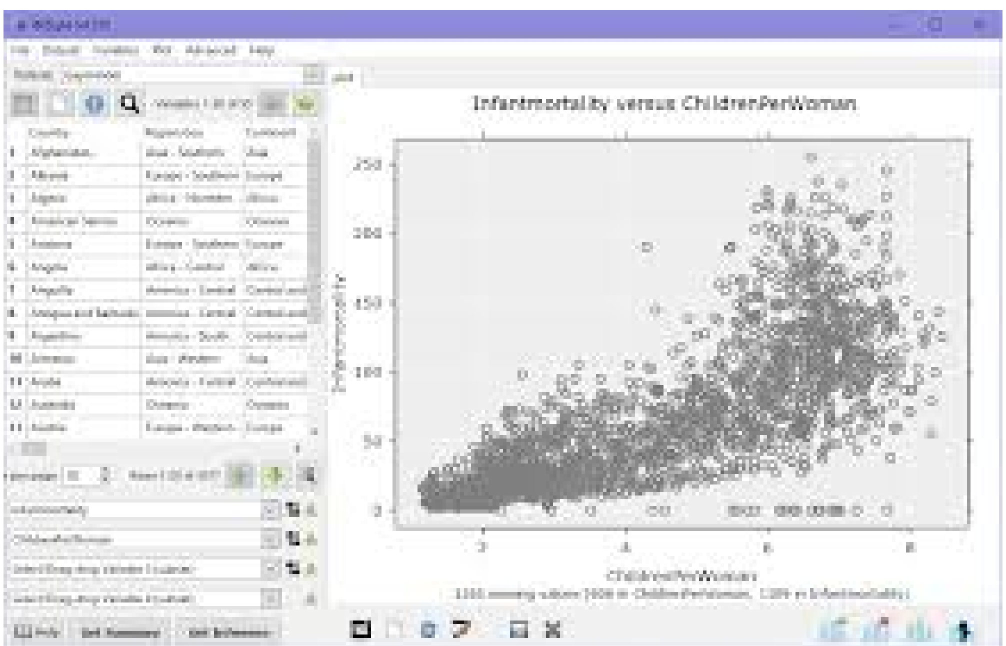

Figure 5: `iNZight` interface. Source: https://inzight.nz/.

`iNZight` comes in both desktop and web-based versions, but these are both in need of a significant rebuild. As part of this, we are looking to take the existing "exploration" capabilities of `iNZight` and add fully interactive features to them. The desktop version of `iNZight` is shown in Figure 5.

2.6 Features of interactive graphics

The key features of interactive graphics for visual analysis include:

- **Linked plots**, in which selections in one plot are reflected in others;
- **Brushing** for interactively selecting and highlighting points or other elements;
- Statistical analysis, inference, and summary generation which can be explored interactively (filtering, identication, linking to graphics, etc);
- **Direct manipulation** of data beyond graphical elements, rather than just the graphics themselves—an example could be transforming a variable in the dataset;
- **Dynamic updates** reflecting data changes or user interactions, which facilitates easy real-time exploration; and
- **Contextual information** and controls, including buttons, tooltips, menus, and other elements that provide additional information or functionality when required.

2.7 Modern web interactive graphics

Current web-based visualization libraries such as D3,[2] Highcharts,[3] Plotly,[4] and Vega[5] enable native interactive graphics with high customization potential. However, these tools often involve significant technical overhead, resulting in implementations that typically focus on single-figure interactivity. Linked graphics remain rare, and typically require manual effort to implement.

That said, the web-based tools are often more accessible to users, as they can be run in a browser without any installation. They are also highly performant, as the graphics are rendered on the user's machine, and modern devices are capable of rendering complex graphics quickly. For this to work, the data must be available to the user, which can be a security concern for sensitive data.

[2] https://d3js.org
[3] https://www.highcharts.com/
[4] https://plotly.com/
[5] https://vega.github.io/vega/

3 A framework for interactive data analysis

We are working on a framework that addresses current limitations while incorporating modern security requirements. The goal of our work is to create a system of related packages that allow implementation of the features discussed in Section 2.6, but with the added flexibilty to include others where necessary.

3.1 Architecture

The primary architecture will be based on `Rserve` (Urbanek, 2003), a binary R server that allows other applications to use the facilities of R. Early related work involves developing a Type-safe JavaScript library for connecting to `Rserve`: an early version of `rserve-ts` is available from https://npmjs.com/package/rserve-ts. We are also working on an R library that makes it easy to develop `Rserve` applications in R, tentatively called `ts` and available from https://github.com/tmelliott/ts.

Using these software as a foundation, we will design several graphics libraries (one for R, another for TypeScript, and potentially some framework-specific libraries such as React). These libaries will enable us to map out an architecture that includes:

- Server-side data analytics engine, allowing complex statistical analysis tasks to take place;
- Front-end graphics and interaction interface, using modern web libraries such as `D3`;
- Client- and server-side linking implementations that allow fast interactivity on public data, while retaining this functionality even when privacy restraints are in place;
- Dynamic computation of statistical information, that is pushed to graphics when the data changes (for example if the dataset is filtered);
- Comprehensive access control system, such that data can be shared with others without revealing the data itself, and access permissions can be set at any desired levels; and
- Permissioned result sharing, allowing users to safely share their results with others without compromising the data.

The framework aims not to replicate complex systems like `D3`, but rather to provide a widget-based graphics library that automatically handles linking between visualization components. Conceptually, graphs will be independent entities with “hooks” that can be connected to a controller, which can either run in the client or on the server.

4 Conclusion

The evolution of interactive data analysis systems reflects the growing importance of dynamic, accessible visualization tools in modern data science. While current web technologies offer sophisticated visualization capabilities, there remains a need for frameworks that balance interactive functionality with robust analytical capabilities and security considerations. Our proposed framework wiull address these challenges by combining the analytical power of R with modern web technologies.

References

Bartonicek, A. (2024a). *The algebra of interaction* (tech. rep.). https://bartonicek.github.io/plotscaper/articles/algebra.html

Bartonicek, A. (2024b). Plotscaper: Explore your data with interactive figures. https://doi.org/10.32614/cran.package.plotscaper

Bartonicek, A., Urbanek, S., & Murrell, P. (2025). No more, no less than sum of its parts: Groups, monoids, and the algebra of graphics, statistics, and interaction. *Journal of Computational and Graphical Statistics*, 1–12. https://doi.org/10.1080/10618600.2024.2429708

Cook, D. (2007). *Interactive and dynamic graphics for data analysis: With r and ggobi* (D. F. Swayne, Ed.) [Includes bibliographical references (p. 177-184) and index]. Springer Science+Business Media, LLC.

Elliott, T., & Wild, C. (2023). iNZight for data analysis [Software version 4.4.0]. https://inzight.nz/

Swayne, D. F., Lang, D. T., Buja, A., & Cook, D. (2003). Ggobi: Evolving from xgobi into an extensible framework for interactive data visualization [Data Visualization]. *Computational Statistics & Data Analysis*, *43*(4), 423–444. https://doi.org/https://doi.org/10.1016/S0167-9473(02)00286-4

Theus, M., & Urbanek, S. (2014). *Interactive Graphics for Data Analysis: Principles and Examples*. Chapman; Hall/CRC. https://doi.org/10.1201/b17187

Urbanek, S. (2003). Rserve – A Fast Way to Provide R Functionality to Applications. *Friedrich Leisch & Achim Zeileis*.

Ward, M. (2015). *Interactive data visualization: Foundations, techniques, and applications* (G. G. Grinstein & D. Keim, Eds.; 2. ed.) [Includes bibliographical references and index]. CRC Press, Taylor & Francis Group, an AK Peters book.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Young, F. W. (2011). *Visual statistics: Seeing data with dynamic interactive graphics* (P. Valero-Mora & M. Friendly, Eds.; Online-Ausg.) [Includes bibliographical references (p. 339-349) and indexes. - Electronic reproduction; Palo Alto, Calif; ebrary; 2011; Available via World Wide Web; Access may be limited to ebrary affiliated libraries]. Wiley-Interscience.

Young, F. W., & Bann, C. M. (1996). *ViSta: The visual statistics system* (tech. rep.). Technical Report 94–1 (c), UNC LL Thurstone Psychometric Laboratory Research Memorandum.