

# “Principled and Efficient Interactive Data Visualization”

## by Bartonicek, Adam

## for a PhD in Statistics

Student ID: 828803059

Email: abar435@aucklanduni.ac.nz

Department of Statistics

Supervisor: Dr. Simon Urbanek

Co-supervisor: Dr. Paul Murrell

Advisory Committee: Drs B. Efron, D. R. Cox, K. Pearson.

Date of enrolment in the programme and expected date of completion

March 29, 2023

This document represents the student’s research proposal after one year of provisional PhD registration. Confirmed PhD registration is now sought.

## 1 Introduction

Humans learn about the world around them by interacting with it. Further, as major part of our cerebral cortex is devoted to visual processing, we learn best by interacting with things we can see. The same applies to data. If we want to learn from our data effectively, we need practical and reliable tools for visualizing data and interacting with it. As such, interactive data visualization is an important pursuit in statistics and data science, and this is evidenced by its rising popularity: currently, many interactive data visualization libraries exists across popular data-science-adjacent programming languages (such R, Python, Julia, and Javascript) and their ecosystems, including D3 (Bostock et al., 2011), plotly (Plotly Inc., 2023), Highcharts (Highcharts Team, 2023), and Vega (Satyanarayan et al., 2015) and Vega-lite (Satyanarayan et al., 2016).

Yet, many of these present-day interactive data visualization libraries tend to suffer from a common set of drawbacks. Specifically, due to historical and other reasons, the libraries tend to fall into two camps:

1. Specialized, high-level suites of ”off-the-shelf” interactive visualizations
2. Generic, highly customizable low-level frameworks.

While the libraries from the first camp are usually approachable to users with less programming experience, their major drawback is that they are not extendable. Conversely, libraries from the second camp lend the user great deal of power and flexibility, however, to use them effectively, the users are required to have significant programming experience and deep knowledge of the specific API, especially when it comes to setting up complex interaction between multiple plots. Further, the users are responsible for making sure that the interaction is coherent - there is nothing stopping them from creating figures with interactive features that are meaningless, confusing, or unpredictable. Finally, the design philosophy behind many of these libraries seems to be oriented towards and used for data presentation rather than data exploration (Batch and Elmqvist, 2017). As such, within the interactive data visualization sphere, there is currently a lack of a mid-level, multi-plot interaction system geared towards data exploration.

Crucially, the abovementioned absence of a mid-level interactive data visualization system is not just an implementation gap. Rather, it seems to be a symptom of an underlying lack of a strong theoretical foundation. Specifically, there do not seem to be any guidelines for how such a mid-level system should operate - what criteria should interactive plots meet so that they can be composed together and behave in predictable and consistent ways. As a consequence, the two options left to the creators of interactive data visualizations systems are either: 1) give the user a limited number of ready-made solutions that are known to behave well, or 2) put the onus of making sure that the interaction is coherent onto the user - which precisely correspond to the two interactive data visualization camps outlined above.

The goal of the proposed project is to lay down the theoretical foundations for such a mid-level interactive data exploration system, as well as implement it within a modern programming language. Specifically, at the theoretical level, I will map out the boundaries and constraints that parts of the system - underlying datastructures, plots, graphical primitives, etc... - need to meet in order to satisfy certain criteria of coherency and consistency, such that the user's interactions with the visualization are meaningful, understandable, and predictable. At the applied level, I will implement the system in **JavaScript** and provide a high-level interface in **R**. Finally, the broader teleological goal of the project is to create a tool that applied statisticians and data scientists can use to understand their data in a convenient, principled, and efficient way.

## 2 Background

### 2.1 What Event Counts as *Interactive* Data Visualization?

It may seem surprising, but despite the widespread popularity of interactive data visualizations, there seems to be little consensus as to what actually makes a data visualization interactive. The term gets used by different researchers in vastly different contexts, sometimes in arguably conflicting ways. As such, for the purposes of the present text, it is important to disambiguate what is meant by "interactive data visualization".

Firstly, there is the issue of whether "interactive data visualization" refers to an object or system, or to an action, undertaken by a human being. Pike et al. (2009) note that "interaction" is an overloaded term that can refer to either the concrete tools which users use to manipulate visual information or to the more abstract "human interaction with information" - the back-and-forth between the user and the visual information presented to them (see also Yi et al., 2007). The more abstract action definition is more often emphasized in the field of Human Computer Interaction (see e.g. Sinha et al., 2010). For the purpose of this text, the object/system definition will be used - interactive data visualizations are concrete objects that are produced by (typically computer-based) systems/pipelines that take in raw data and produce images that can be interpreted and manipulated by humans (Brodbeck et al., 2009).

But even after narrowing down the focus to interactive data visualizations as objects, a lot of conceptual ambiguity remains. Some researchers use a **simple** definition and define interactive visualizations as any visualizations that can be actively manipulated by the user (Brodbeck et al., 2009). Other researchers emphasize time or the **temporal** aspect, with visualizations being "interactive" when there is little lag between the user's input and changes to the visualization (Becker and Cleveland, 1987; Buja et al., 1996). Complicating the matters further, some even make the distinction between "interactive" and "dynamic" manipulation, where interactive manipulation happens discretely, such as when pressing a button or selecting an item from a drop-down menu, whereas dynamic manipulation happens continuously, for example when smoothly moving a slider or by clicking-and-dragging (Rheingans, 2002; Jankun-Kelly et al., 2007). These tentative definitions present relatively little restriction on what counts as interactive visualization: for example, one could

argue that the process of a user typing code into a command line to generate new plots could be considered interactive visualization, as long as it happens fast enough.

Other researchers do not seem to be satisfied with these broad definitions. For many, the defining feature of interactive data visualization is the ability to **query** different parts of the dataset (by e.g. zooming, panning, and filtering), and the reactive propagation of changes between connected or **"linked"** parts of interactive figures (Kehrer et al., 2012; Buja et al., 1996; Keim, 2002; Unwin, 1999). Similarly, in Visual Analytics (VA) research, a distinction is made between "surface-level" (or "low-level") interactions, which manipulate attributes of the visual domain only (e.g. zooming and panning), and **"parametric"** (or "high-level") interactions, which manipulate attributes of mathematical models or algorithms underlying the visualization (Leman et al., 2013; Pike et al., 2009).

There are other ways that the term "interactive data visualization" has been used (for more detailed taxonomies, see Yi et al., 2007), however, short summaries of the key definitions mentioned above are presented in Table 1 and will be referred to later on in the text. What is important is that the different types of interactivity imply very different levels of programming complexity. For example, simply changing the color of a point or a bar in a plot, irrespective of anything else, might be implemented by changing an attribute of the underlying graphical primitive only - the point/bar does not need to know about what data it represents. However, if the change happens in response to (linked) brushing within a different plot, then there does need to be some way of tracking which cases of the data belong to the primitive. Likewise, changing the width of a histogram bar does not affect the graphical attributes of the bar only, but the underlying operation (binning) needs to be recomputed with respect to the new value of the parameter.

Table 1: Definitions of Interactive Data Visualization

Name	Short Definition	Details
Simple	Change happens	User can manipulate the visualization in some way
Temporal	Change happens in real time	There is little lag between the user's input and changes to the visualization
Querying	Change results from subsetting	The user can query different parts of the dataset, interaction is analogous to subsetting rows of the data (e.g. zooming, panning, and filtering)
Linked	Change propagates	Parts of the visualization are connected or "linked", such that interaction with one part produces a change in another (e.g. linked brushing)
Parametric	Change reflects an underlying model	The user can manipulate the parameters of some underlying model (e.g. rotating principal axes in a PCA scatterplot, changing the width of a histogram bar)
(Cognitive)	Change is caused and perceived by a human	The user engages in a back-and-forth with the visual information presented to them

The conceptual ambiguity about what gets called "interactive" visualization matters because it leads to radically different implementations in software packages (see Sections 2.2 and 2.3). For example, the **R Graph Gallery** page on **Interactive Charts** (Holtz, 2022) features several examples of interactive visualizations, however, none of them meet the linked and parametric definitions of

interactivity outlined in Table 1. Interactive data visualization systems that exist within the open source data visualization ecosystem differ significantly in the amount of features and flexibility they offer, as well as in how much responsibility or "house-keeping" for maintaining the interactive state they offload onto the user.

## 2.2 Brief History of Interactive Data Visualization in Statistics

Static data visualization has a rich and intricate history (see e.g. Dix and Ellis, 1998; Friendly, 2006; Friendly and Wainer, 2021; Young et al., 2011). Briefly, for a long time, it was considered at best an auxiliary field, however, at the end of 1950's, a series of developments lead to a great increase in its prominence. Firstly, at the theoretical level, the work of Tukey (1962; 1977) and Bertin (1967) established data visualization as valuable discipline in its own right. Secondly, at the applied level, the development of personal computers (see e.g. Abbate, 1999) and high-level programming languages, most notably FORTRAN in 1954 (Backus, 1978), made production of figures easy and accessible to the wider public. Combined, these developments lead to a surge in the use and dissemination of data visualizations.

Final development in the field of static data visualization that is important to mention was the Grammar of Graphics introduced by Leland Wilkinson (2012). Prior to Wilkinson's work, data visualization systems tended to come in two flavors: low-level ones, in which the users had to create visualizations from scratch using graphical primitives, and high-level ones, in which the users could select from a limited range of ready-made visualization types. Wilkinson, building upon the work of Bertin and Tukey, developed theory for a mid-level visualization system - Grammar of Graphics - which allows the users to specify a broad range of statistical graphics by declaratively combining abstract plot attributes such as aesthetics, scales, coordinates, and geometric objects (Wilkinson, 2012). Grammar of Graphics has been successfully implemented in several software packages, most notably the popular **ggplot2** R package (Wickham, 2010) and the proprietary software **Tableau** (Tableau Team, 2023).

As static visualization entered the computer age, interactive data visualization would not be left far behind. Early systems appeared in the 1960's and 1970's, and tended to be specialized for one specific task. For example, Fowlkes (1969) used interactive visualization to show how probability densities reacted to change of parameters and transformations, and Kruskal (1965) used interactive visualization to showcase his multi-dimensional scaling algorithm (a way of embedding objects within a common space based on pairwise distance measurements). The first "general-purpose" system was **PRIM-9** (Fisher-Keller et al., 1974), which allowed for exploration of high-dimensional data in scatterplots using projection, rotation, subsetting and masking. Later systems grew on to become even more general and ambitious. For example, **MacSpin** (Donoho et al., 1988) and **XGobi** (Swayne et al., 1998) provided features such as interactive scaling, rotation, linked selection (or "brushing"), and interactive plotting of smooth fits in scatterplots, as well as interactive parallel coordinate plots and grand tours.

Following the turn of the 21st century, interactive data visualization systems saw even scope and flexibility and also began to be integrated into general-purpose statistical computing software. The successor system to XGobi, **GGobi** (Swayne et al., 2003) was made to be directly embeddable in R. Java-based **Mondrian** (Theus, 2002) allowed for sophisticated linked interaction between many different types of plots including scatterplots, histograms, barplots, scatterplot, mosaic plots, parallel coordinates plots, and maps. Finally, **iPlots** (Urbanek and Theus, 2003) implemented a general framework for interactive plotting that was not only embedded in R but could be directly programmatically manipulated, and was later further expanded and made performant for big data in **iPlots eXtreme** (Urbanek, 2011).

Common features that these statistics-based interactive data visualization systems supported

were: 1) multiple ready-made plot-types 2) interaction between multiple plots with shared underlying data and state, and 3) interactive manipulation of model parameters. They tended to be oriented towards scientific audience, with data exploration as the primary goal.

## 2.3 The Web and Current Age Interactive Data Visualization

The developments in interactive data visualization within the field of Statistics were paralleled by those within Computer Science. Most notably, the rise of Web technologies in the mid 1990's and the appearance of JavaScript in 1995 as a high-level general-purpose programming language for the Web (for a description of the history, see e.g. Wirfs-Brock and Eich, 2020), created an extremely versatile platform for highly-reactive and portable applications. JavaScript was created with the explicit purpose of making the Web interactive, and the fast dissemination of standardized web-browsers meant that online interactive applications could be accessed by anyone, from anywhere. Interactive data visualization became just one of many technologies highly sought after within the fledgling Web ecosystem.

The very early systems such as **Prefuse** in 2005 (Heer et al., 2005) and **Flare** in 2008 (Burleson, 2020) relied on plugins (Java and Adobe Flash Player, respectively). However, in the early 2010's, several true Web-native JavaScript-based interactive data visualization systems emerged. The most prominent among these is **D3.js** (Bostock et al., 2011). D3 is a broad and general framework for manipulating HTML documents using data and displaying it with visualizations. It is in popular use still to this day and has spawned a number of specialized, higher-level data visualization libraries that abstract away the details and provide an interface to commonly used statistical plots, such as the also very prominent **plotly.js** (Plotly Inc., 2022). Importantly, in the D3 model, D3's engine only renders graphics: interactivity is left the user. If a user wants to create a visualization that's interactive, she has to write JavaScript functions that take care of updating the underlying parameters. A radically different approach altogether was taken by **Vega** (Satyanarayan et al., 2015). In Vega, all aspects of visualization including interactivity are specified declaratively, via a JSON-like specification object. As such, Vega carries a large number of reactive primitives. Similar to D3 and plotly, Vega also spawned its own higher-level libraries in **Vega-lite** (Satyanarayan et al., 2016), and **Altair** (VanderPlas et al., 2018). A final popular interactive data visualization library in JavaScript is **Highcharts** (Highcharts Team, 2023). Similar to Vega-lite, Highcharts is a high-level declarative framework in which plots are created based on a JSON-like specification object.

These contemporary web-based interactive data visualization systems offer a great deal of expressiveness, however, it does seem to come at a cost. Specifically, the low-level frameworks like D3 and Vega allow the users to create almost arbitrarily complex interactive figures, and even the higher-level frameworks like plotly, Vega-lite, Altair, and Highcharts still offer a great deal of customizability. Yet, a lot of code and programmer time is required to create complex interactive figures, even in these higher-level frameworks. As a result, most examples that show up on the web or on the libraries' own showcase pages typically feature only shallow interactivity within a single plot - such as zooming, panning, and pop-up labels - and examples of more complex multi-plot interaction such as linked brushing or cross-filtering are far less common. Further, the coherence of interactivity is left to the user: there is nothing stopping them from creating figures with meaningless interactive features. As such, it is hard to call the frameworks like Vega-lite, Altair, and plotly true mid-level systems, the way that ggplot2 can be called a mid-level system for static graphics.

The target audience of the web-based systems is also very different from that of the statistical systems (discussed in Section 2.2). Most real-world use examples come from online news articles and business/government dashboards, with considerably fewer appearing in scientific outlets. Likewise, the focus seems to be more on data presentation rather than data exploration - communicating findings once they have been found rather than discovering them in the first place. This seems to

make sense given the design of these systems - it is worth it to create complex interactive visualizations when there is the guarantee that many people will see the result; however, conversely, it may not be economical to do so with  $n = 1$  (i.e. when the user is the only person who will see the visualization). Data presentation and data exploration are both very important pursuits, however, it does seem that the market for interactive data visualizations for EDA is currently underserved, and this may explain why seemingly few data scientists use interactive visualizations to explore their data (Batch and Elmqvist, 2017).

## 2.4 Specialization vs. Generality

To re-state the common thread from the previous sections, over the last 30 years or so, interactive data visualization has undergone a divergent evolution within two largely independent branches: statistical and web-based/computer science. Different goals and features were prioritized within each branch. The statistical branch focused on creating specialized systems for scientific data exploration. Users could easily create complex interactive figures by picking from a limited range of pre-made, "out-of-the-box" plots that were designed to behave coherently and consistently when composed together. Conversely, in the web-based branch, greater focus was put on generality and data presentation. Users were given a great deal of power and flexibility to create arbitrarily complex interactive figures from scratch, however, the onus for ensuring the interactivity was coherent and consistent was put on them.

The difference between the two branches represents a fundamental tension between specialization and generality. Specialized systems are easy to use but hard to extend; generic systems are extensible by definition but require time and effort to use effectively.

## 2.5 A Very Brief and Inadequate Introduction to Category Theory

# 3 So What's New?

This project seeks to develop a theory for a mid-level interactive data exploration system and fill its unoccupied niche within the interactive data visualization sphere. To accomplish this, the following goals will be actioned:

1. Map out boundaries and constraints that a mid-level interactive data visualization system should meet
  - (a) Describe what contracts do graphical primitives and their underlying statistical summaries need to uphold in order to be interactable with in a coherent and predictable way
2. Develop a theoretical foundation for the system
  - (a) Use concepts and syntax of category theory to describe the necessary structure
  - (b) Consider links back to implementation via functional programming
3. Implement the system in JavaScript and publish as an `npm` package
  - (a) Use plain ES2020 as much as possible, possibly incorporate a reactive programming library and functional programming utility library
  - (b) Use the HTML `canvas` element as a drawing utility, explore using WebGL and other frameworks for performance
  - (c) Provide a full documentation for any functions, classes, etc...

4. Implement a high-level interface to the system in R and publish on CRAN
  - (a) Provide a full documentation for all functions
  - (b) Provide a package vignette
  - (c) Pass all CRAN checks
5. Publish an article in a international peer-reviewed journal such as the *Journal of Statistical Software* or the *R journal*
6. Present the system at an R developer or data visualization conference such as useR! or IEEE VIS

## 4 Data and Special Needs

The use of the system will be tested out on datasets available in the public domain such as the `diamonds` dataset in R. For these data sets, no ethics or disclosure statements are necessary. Possible collaboration with working scientists and corresponding data access may be negotiated at a later date.

The student's own personal computer should be sufficient to produce the developing and testing the system. Should extra computing resources become necessary, the student should be able to inquire and be granted access to the **Ihaka server** (`ihaka.stat.auckland.ac.nz.`) using SSH connection. Other options such as the New Zealand eScience Infrastructure (**NeSI**) should be available too.

The packages produced as part of the project will be published for free as open-source software, under the MIT license (The Open Source Initiative, 2023). No additional ethics approvals are necessary.

## 5 Budget

Development of open-source software is not very budget-intensive. The work related to the project will be funded by the University of Auckland Doctoral Scholarship. Expenses related to conferences (registration fees, travel) may be funded via the Postgraduate Research Student Support (PReSS) account, with an annual allocation of \$1200 NZD.

## 6 Objectives and Goals

The objective(s) of this research project are to...

List *specific* goals/tasks that will be undertaken as part of the proposed research.

1. I wish to derive the first three moments of the Slash distribution. This will involve working out the characteristic function first.
  - a. Then work out the observed information matrix based on ?.
  - b. Then work out the expected information matrix (?).
  - c. Apply the EM algorithm (?) to estimate  $\lambda$ .
  - d. Write a R package to implement my method. It will be written in S4 and use object oriented methods (?).
  - e. Apply the method to the radiation data set (?).

- f. Extend biplots (?) for my data.
- g. Release the package on CRAN (cf. ?).
- h. Publish my results in at least two papers, earmarked for JASA and JRSS-B. Also an applications paper in Biometrics. See Section 7.

2. Solve the Fisher-Behrens problem (???).

## 7 Deliverables and Program Schedule

Table 2: Timeline for my thesis. Itemize the list of deliverables with specific dates so that you can make concerted effort to achieve them. Here are *some* activities—fill in more.

Date	Activity
2018-04-01	Provisional PhD registration (PhD in Statistics).
2018-05-01	(Optional) Updated my personal webpage at <a href="http://www.stat.auckland.ac.nz/~myStudentName">www.stat.auckland.ac.nz/~myStudentName</a> .
2018-05-20	Attended one of the Doctoral Skills Programme's Induction Days.
2018-09-01	Gave my first talk at NZSA conference. Won first prize.
2018-11-01	Gave a talk to PhD Talks Day.
2019-01-15	Submit my first paper to <i>Annals of Statistics</i> (co-authored with supervisor).
2019-05-01	Presented my research progress to a departmental seminar.
2020-01-20	Achieve Goal 1.
2020-08-20	Achieve Goal 2.
2020-10-23	Submit my second paper to <i>Annals of Applied Statistics</i> (co-authored with supervisor).
2020-01-15	Submit my third paper to <i>Biometrics</i> (sole authorship).
2020-02-01	Submit my PhD thesis.

I have fulfilled all my first year requirements. These are:

1. All AFA provisional goals:

- (a) *Full thesis proposal normally completed within 6 months*: Here it is!
- (b) *Completion of one substantial piece of written work within 12 months*: Wrote 90 percent of Chapter 1 of my thesis, submitted one journal article, wrote 50 percent of Chapter 2 of my thesis.
- (c) *Presentation of research progress to a departmental seminar*: 2019-05-01.
- (d) *Approval of the full thesis proposal by the appropriate departmental/faculty postgraduate committee*: Being done.
- (e) *Ethics approval(s)/permissions obtained for the research (if required)*: Not necessary.
- (f) *Attendance at one of the Doctoral Skills Programme's Induction Days*: 2018-05-20.

2. Although optional, it is a good idea to try update a personal webpage (ideally [www.stat.auckland.ac.nz/~](http://www.stat.auckland.ac.nz/~) giving details of my thesis, links to other research resources in my topic, and some personal stuff to make it interesting.

3. I have diligently attend as many Statistics Department seminars as I could. They are given in Table 3. This is much more than the minimum quota set by the department. Consequently



there should be no problem due to this when getting my annual report signed off<sup>1</sup>. The 2018-10-08 seminar was particularly useful because it gave me an idea on how to solve one of my problems.

4. I did STATS 730 in Semester 1 of 2018 and obtained an A+.
5. I did STATS 782 in Semester 2 of 2018 and obtained an A.

Table 3: Departmental seminars I have attended (top part of the table). Talks from another UoA department are in the middle tier. Conferences and workshops are in the bottom tier, e.g., 2019-05-29 event was an all-day workshop. **Note:** I filled in the required document and submitted it within the required time period after each seminar I attended.

Date	Speaker	Title
2018-04-08	David Brillinger	Random trajectories, some theory and applications
2018-02-30	David Cox	Frequentist statistics as a theory of inductive inference
2018-10-08	David Matthews	Estimating diagnostic test likelihood ratios
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
yyyy-mm-dd	Speaker Name	Title of Talk
2018-06-16	James B. Conant	“Geography is not a university subject” (geo Department)
yyyy-mm-dd	Speaker Name	Title of Talk
2019-05-29	David Siegmund	Workshop on ‘Genetic Mapping’ at UoA.

## Appendix A

Delete or replace the contents of this section with any appendices you may have.

The *University of Auckland Statute and Guidelines for the Degree of Doctor of Philosophy (PhD)* (2016) reads<sup>2</sup>:

“The PhD degree is awarded for a formal and systematic exposition of a coherent programme of advanced research work. The work is carried out over the period of enrolment for the degree, and in the opinion of the examiners and the Board of Graduate Studies, satisfies all of the following criteria:

- (i) is an original contribution to knowledge or understanding in its field, and
- (ii) meets internationally recognised standards for such work, and

---

<sup>1</sup>If insufficient seminars have been attended then sign off will occur *after* the minimum number is reached.

<sup>2</sup>Regulation 1, Preamble, item (e).

- (iii) demonstrates knowledge of the literature relevant to the subject and the field or fields to which the subject belongs, and the ability to exercise critical and analytical judgement of it, and
- (iv) is satisfactory in its methodology, in the quality and coherence of its expression, and in its scholarly presentation and format.”

## References

- Abbate, J., Sep. 1999. Getting small: a short history of the personal computer. *Proc. IEEE* 87 (9), 1695–1698.
- Backus, J., 1978. The history of fortran i, ii, and iii. *ACM Sigplan Notices* 13 (8), 165–180.
- Batch, A., Elmqvist, N., 2017. The interactive visualization gap in initial exploratory data analysis. *IEEE transactions on visualization and computer graphics* 24 (1), 278–287.
- Becker, R. A., Cleveland, W. S., 1987. Brushing scatterplots. *Technometrics* 29 (2), 127–142.
- Bertin, J., 1967. *Sémiologie Graphique: Les diagrammes, les réseaux, les cartes*. Gauthier-Villars.
- Bostock, M., Ogievetsky, V., Heer, J., 2011. D<sup>3</sup> data-driven documents. *IEEE transactions on visualization and computer graphics* 17 (12), 2301–2309.
- Brodbeck, D., Mazza, R., Lalanne, D., 2009. Interactive Visualization - A Survey. In: *Human Machine Interaction*. Springer, Berlin, Germany, pp. 27–46.
- Buja, A., Cook, D., Swayne, D. F., 1996. Interactive high-dimensional data visualization. *Journal of computational and graphical statistics* 5 (1), 78–99.
- Burleson, T., Nov. 2020. Flare | Data Visualization for the Web. [Online; accessed 18. Oct. 2022]. URL <https://blokt.com/tool/prefuse-flare>
- Dix, A., Ellis, G., May 1998. Starting simple: adding value to static visualisation through simple interaction. In: *AVI '98: Proceedings of the working conference on Advanced visual interfaces*. Association for Computing Machinery, New York, NY, USA, pp. 124–134.
- Donoho, A. W., Donoho, D. L., Gasko, M., 1988. Macspin: Dynamic graphics on a desktop computer. *IEEE Computer Graphics and applications* 8 (4), 51–58.
- FisherKeller, M. A., Friedman, J. H., Tukey, J. W., 1974. An interactive multidimensional data display and analysis system. Tech. rep., SLAC National Accelerator Lab., Menlo Park, CA (United States).
- Fowlkes, E., 1969. User’s manual for a system fo active probability plotting on graphic-2. Technical Memorandum, AT&T Bell Labs, Murray Hill, NJ .
- Friendly, M., 2006. A brief history of data visualization. In: Chen, C., Härdle, W., Unwin, A. (Eds.), *Handbook of Computational Statistics: Data Visualization*. Vol. III. Springer-Verlag, Heidelberg, pp. ???–???, (In press).
- Friendly, M., Wainer, H., 2021. *A history of data visualization and graphic communication*. Harvard University Press.

- Heer, J., Card, S. K., Landay, J. A., Apr. 2005. *prefuse: a toolkit for interactive information visualization*. In: CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 421–430.
- Highcharts Team, Mar. 2023. Interactive javascript charts library. [Online; accessed 15. Mar. 2023]. URL <https://www.highcharts.com>
- Holtz, Y., Aug. 2022. Interactive charts | the R Graph Gallery. [Online; accessed 22. Aug. 2022]. URL <https://r-graph-gallery.com/interactive-charts.html>
- Jankun-Kelly, T., Ma, K.-L., Gertz, M., 2007. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics* 13 (2), 357–369.
- Kehrer, J., Boubela, R. N., Filzmoser, P., Piringer, H., 2012. A generic model for the integration of interactive visualization and statistical computing using r. In: 2012 IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE, pp. 233–234.
- Keim, D. A., 2002. Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics* 8 (1), 1–8.
- Kruskal, J. B., Oct. 1965. Multidimensional Scaling. [Online; accessed 13. Oct. 2022]. URL <https://community.amstat.org/jointscsg-section/media/videos>
- Leman, S. C., House, L., Maiti, D., Endert, A., North, C., 2013. Visual to parametric interaction (v2pi). *PloS one* 8 (3), e50474.
- Pike, W. A., Stasko, J., Chang, R., O’connell, T. A., 2009. The science of interaction. *Information visualization* 8 (4), 263–274.
- Plotly Inc., Aug. 2022. Part 4. Interactive Graphing and Crossfiltering | Dash for Python Documentation | Plotly. [Online; accessed 22. Aug. 2022]. URL <https://dash.plotly.com/interactive-graphing>
- Plotly Inc., Mar. 2023. Plotly: Low-Code Data App Development. [Online; accessed 15. Mar. 2023]. URL <https://plotly.com>
- Rheingans, P., 2002. Are we there yet? exploring with dynamic visualization. *IEEE Computer Graphics and Applications* 22 (1), 6–10.
- Satyanarayan, A., Moritz, D., Wongsuphasawat, K., Heer, J., 2016. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics* 23 (1), 341–350.
- Satyanarayan, A., Russell, R., Hoffswell, J., Heer, J., 2015. Reactive vega: A streaming dataflow architecture for declarative interactive visualization. *IEEE transactions on visualization and computer graphics* 22 (1), 659–668.
- Sinha, G., Shahi, R., Shankar, M., 2010. Human computer interaction. In: 2010 3rd International Conference on Emerging Trends in Engineering and Technology. IEEE, pp. 1–4.
- Swayne, D. F., Cook, D., Buja, A., Mar. 1998. XGobi: Interactive Dynamic Data Visualization in the X Window System. *J. Comput. Graph. Stat.* 7 (1), 113–130.
- Swayne, D. F., Lang, D. T., Buja, A., Cook, D., Aug. 2003. GGobi: evolving from XGobi into an extensible framework for interactive data visualization. *Comput. Statist. Data Anal.* 43 (4), 423–444.

- Tableau Team, Mar. 2023. Tableau: Business Intelligence and Analytics Software. [Online; accessed 16. Mar. 2023].  
URL <https://www.tableau.com>
- The Open Source Initiative, Feb. 2023. The MIT License. [Online; accessed 29. Mar. 2023].  
URL <https://opensource.org/license/mit>
- Theus, M., Nov. 2002. Interactive Data Visualization using Mondrian. *J. Stat. Soft.* 7, 1–9.
- Tukey, J. W., 1962. The future of data analysis. *The annals of mathematical statistics* 33 (1), 1–67.
- Tukey, J. W., et al., 1977. *Exploratory data analysis*. Vol. 2. Reading, MA.
- Unwin, A., Mar. 1999. Requirements for interactive graphics software for exploratory data analysis. *Comput. Statist.* 14 (1), 7–22.
- Urbanek, S., 2011. iplots extreme: next-generation interactive graphics design and implementation of modern interactive graphics. *Computational Statistics* 26 (3), 381–393.
- Urbanek, S., Theus, M., 2003. iplots: high interaction graphics for r. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Citeseer.
- VanderPlas, J., Granger, B., Heer, J., Moritz, D., Wongsuphasawat, K., Satyanarayan, A., Lees, E., Timofeev, I., Welsh, B., Sievert, S., 2018. Altair: interactive statistical visualizations for python. *Journal of open source software* 3 (32), 1057.
- Wickham, H., 2010. A layered grammar of graphics. *Journal of Computational and Graphical Statistics* 19 (1), 3–28.
- Wilkinson, L., 2012. *The grammar of graphics*. Springer.
- Wirfs-Brock, A., Eich, B., Jun. 2020. JavaScript: the first 20 years. *Proc. ACM Program. Lang.* 4 (HOPL), 1–189.
- Yi, J. S., ah Kang, Y., Stasko, J., Jacko, J. A., 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics* 13 (6), 1224–1231.
- Young, F. W., Valero-Mora, P. M., Friendly, M., 2011. *Visual statistics: seeing data with dynamic interactive graphics*. John Wiley & Sons.