Click and Brush: A Novel Way of Finding Correlations and Relationships in Visualizations

Michael A.E.Wright and Jonathan C. Roberts

Computing Laboratory, University of Kent, Canterbury, Kent, UK. maew2@kent.ac.uk, j.c.roberts@kent.ac.uk

Abstract

Brushing is a commonly used interaction technique that allows users to select items or an area of the visualization as the user moves the mouse. Brushing is often used with multiple views, where any co-related information is simultaneously highlighted in these linked views. Brushing is a direct manipulation technique where interaction is performed directly with the visualization. Similar effects can be engendered through indirect manipulation, using dynamic query sliders. In fact, such indirect manipulation can be more effective as the user is able to both highlight and constrain interesting features. In this paper we present a new brushing technique, called 'Click and Brush', that allows the user to both highlight correlations and constrain intersections in the data using direct manipulation. Users are able to highlight data (brush), fix this subset (click) and explore further intersections (subsets) of this data through further brushing operations.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Methodology and Techniques]: Interaction Techniques H.5.2 [User Interfaces]: Interaction styles (direct manipulation)

1. Introduction

Visualizations provide the user with a way of graphically representing their data, from simple 2D scatterplots to complex 3D volume representations. The aim of these visualizations is to display vast amount of data in a way that the user can more easily understand. It is well known that by interacting with the representation users gain a deeper understanding of the data.

There are two styles of interaction: indirect and direct manipulation. With indirect manipulation the user controls the functionality of the system by using sliders, menus, buttons etc, thus the users interaction (in the case of a visualization system) changes the content of the displayed information. The FilmFinder system [AS94] is a seminal example of indirect manipulation; this tool used an array of sliders to constrain the information being displayed. For example, the user could choose to only display films that were of a particular genre, included a specific actor, was of a certain length and directed by a specific director. On the other hand, in a direct manipulation interface the methods that the user operates the system are transparent - the user directly interacts with the displayed information [Shn83]. There are many examples of

direct manipulation, brushing is one example where by moving the mouse over a point on the visualization causes that data point to change colour (which also selects the point). Another example is the method of gesturing, such as used by Graffiti, to input text into a keyboardless PDA. Obviously, as the user interface is transparent, the requirements of the technique are that it is easy, memorable and intuitive to use.

Direct manipulation techniques and particularly brushing, are powerful techniques for interaction and exploration of visualizations. They give the user the sense of actually manipulating and controlling the visualization as opposed to indirect manipulation techniques that give the user an abstract sense of interaction. Li et al. [LBS*03] compared brushing histograms (direct manipulation) to dynamic query sliders (indirect manipulation) to see which technique users found more useful and intuitive to use when exploring generalized geographic information visualization. They concluded that "brushing ... [is] superior for more complex discovery tasks, and more highly rated by users for relationship identification".

However, indirect manipulation techniques do have advantages. Techniques such as dynamic query sliders as de-

© The Eurographics Association 2005.



scribed by Ahlberg et al [AS94] allow the user to filter and constrain what information is presented. By adjusting the values of various parameters in this way the user is able to explore the relationships between different variables. This constraint operation is difficult to achieve using direct manipulation techniques.

In this paper we present a new direct manipulation technique that incorporates both brushing and dynamic filtering. The technique enables users to explore relationships between variables. This technique, called 'click and brush', gives the user the ability to brush (to select some elements), then click (to fix the subset), then dynamically brush over different aspects of the visualization system (such as other elements or menu items), which further constrains the information being displayed in the subset. As we shall see, the constraint operation can have different behaviours depending on the representation of the data. Click and Brush builds on considerable research that has been undertaken into brushing and multiple coordinated view visualizations. Click and Brush has been developed into an extended version of MailView [FRB05].

The rest of this paper is organized as follows, in section 2 we will look at some related an background research, in section 3 we will describe the 'Click and Brush' interaction technique, in section 4 we will give some examples of how the technique can be applied and finally, in section 5 we will conclude and suggest some area for further research.

2. Related Work and Background Research

In the early years of computer graphics John Tukey was interested in data analysis; together with Friedman and Fisherkeller [MAFT75] they demonstrated the PRIM-9 system. This system allowed the user to explore visualizations by interactively highlighting regions of interest. This technique was not classified as brushing however the concepts are similar. Becker and Cleveland in 1987 [BC87] introduced the idea of brushing to interact with scatterplot visualizations where the action of moving (brushing) the mouse over elements of the graph selected or highlighted those elements. They stated that brushing is a dynamic graphical method for analyzing data in higher dimensions.

There are three main characteristics of brushing. 1. How the brush is activated. This may be (a) clicking first and dragging something, often a bounding box is dragged (b) hovering over an element to select it, or (c) hovering or wandering over multiple elements each element being accumulated to a selection list. 2. The type and shape of the brush can differ; this concept is similar to brushes used in decorating where the brush can be big or small with fine or coarse hairs. Likewise in visualization, brushes can have different forms such that they can be a point, line or area. This is named implantation by Bertin [Ber81]. For example, Becker and Cleveland define a vertical brush (used on two-dimensional

scatterplots) where the brush area moves in synchrony with the x-coordinate of the mouse pointer (irrespective of the y-coordinate value). 3. The operation that is enacted after a selection, for example, (i) changing the color of brushed elements i.e. changing the retinal variable mapping of the selected elements, (ii) selecting elements based on multiple criteria, (iii) removing selected elements from the visualization (i.e displaying every element other than those that are selected), or (iv) displaying extra information about a selected element; these are known as highlight, shadow highlight, delete and label respectively by Becker and Cleveland [BC87].

Interaction with visualizations is extremely important. Dix and Ellis [DE98] stated that "[adding]... interaction to static visualizations ... [makes them] more powerful and easier to understand". Direct manipulation techniques, such as brushing, allow users to interact, manipulate and explore visualizations. This promotes the discovery of relationships, structure and correlations in the data. Much research has been undertaken into brushing since Becker and Cleveland's seminal paper. This has resulted in many new brushing techniques that allow the user to explore visualizations in new and novel ways. These new brushes change various characteristics of the aforementioned brushing characteristics.

One such brushing variant is angular brushing [HLD02], which selects lines from parallel coordinate graphs depending on the angle the line makes with an axis. This technique allows users to specify a minimum angle X and a maximum angle Y, and selects all the lines within that constraint. Angular brushing provides a novel way of highlighting relationships within the data by exploiting the characteristics of parallel coordinate graphs. It extends the brushing metaphor to allow selection based on the properties of the visualization rather than selecting specific lines or points directly under the brush head.

Another technique, presented by Wong and Bergeron [WB97], extends the brushing metaphor for use in 3D volume data sets. They propose a focus+context method that renders the whole volume data using a coarse resolution, with the features brushed by the user at a higher resolution. Their motivation was to increase rendering time on large datasets while depicting high resolution where required. They explain this using an example of brushing over a 3D rendering of a human head, such that if the user selects the bone, the bone and joint structures are rendered in high resolution while other features such as skin is rendered from a much lower resolution dataset. Similarly, Wong and Bergeron [WB96] present an approach to aid exploration of multidimensional wavelets. Scientific visualizations often use wavelets when visualizing data as they provide a good algorithm to approximate data sets. Their brushing interaction allows the user to brush over an area of the wavelet visualization, which results in a higher resolution rendering of the data within that area superimposed on to the visualization.

Smooth brushing is another extension of the brushing metaphor developed to cope with the smooth distribution of data elements within simulation data. Doleisch and Hauser [Doleisch02] used smooth brushing to aid in defining an elements degree of focus in multiple view visualizations of 3D flow simulation data. They use Focus+Context techniques and linking to alter the visualization views so that the user is able to gain a better understanding of the data. Smooth brushing enables the user to select a region of interest in ndimensions which is then assigned a DOI (Degree of Interest) value, in this case either a 0 meaning it is part of the context or 1 meaning that the element is in focus. This DOI function then determines what or how elements in the different visualizations are displayed to the user. This notion of assigning values to elements to describe there visibility was also used in the XmdvTool [Ward97] with ramped boundaries which are used to allow highlighting of elements that are partially included within a selection by a bounding box

The brushing techniques mentioned so far are spatially dependent – the user selects a set of elements based on the coordinates and location of where the information is displayed. This can be a disadvantage because the user is only exploring through one dimension. Consider the visualization process, the data is first enhanced and filtered, then mapped and finally displayed to the screen. In order to select elements by a user driven brush the system needs to calculate an inverse mapping i.e. from screen space to data space. In this case, each of the elements selected are contiguous by how they are displayed. On the other hand, indirect manipulation, such as dynamic queries, allows non-spatial and spatially non-contiguous elements to be selected. Selective Dynamic Manipulation provides one such compromise between indirect and direct manipulation.

SDM (Selective Dynamic Manipulation) [CRMK95] enables interaction to be performed by directly manipulating handles, which are attached to objects within the visualization. Users are able to move the handles to adjust its value, which causes some or all of other visualized objects to change accordingly. This technique is similar to dynamic query slider as it allows the user to test scenarios by adjusting values of parameters, which then change the visualization to reflect these changes. Like dynamic query sliders this type of manipulation is useful as the user can see relationships between the objects being visualized. Ward [War97] names this "Direct, user-specified" brushing, he details an example of a bounding-box brush, where the user can click on the centre of the brush to change its location, whereas by selecting on the boundary the user can change its size.

Another method of brushing across dimensions can be performed through brushing in a multiple view environment. The user can brush in one dimension in one view, and see the results of their brushing operation highlighted in other dimensions in further views. For example, Siiritola [Sii03] looked at brushing multivariate data in two

complimentary visualizations using a parallel coordinate graph and a reorderable matrix. Brushing multiple linked views has been used in many systems, including Ward in his XmdvTool [MW95] [War94] HyperSlice [vWvL93] Klinker [Kli93] and Lowekamp and Rheingans [LRY02]. Users are able to brush elements or subsets in one view, to be highlighted in all co-related views. This style of linked brushing is of particular benefit when comparing and searching for trends. However, brushing multiple views can be confusing to the user as it may not be clear how a point in one view relates to another view.

In addition to brushing multiple views, it is possible to have multiple brushes. Each of the brushes can be used independently and highlight the selected data in different colors [War97]. This can occur either by keeping the brush heads visible, or the subsequent brush operation can be sequentially added. Furthermore, logical and other intersection operations may be performed on these multiple brushes, this is also known as *Compound Brushing*. For example, Chen [Che04] developed a flexible compound brushing system based on higraphs where various components of the system can be linked together via various logical operations and expressions (e.g. AND, OR, XOR, LESS, EQUAL etc). The higraphs control which information is compound and the operation of the intersections.

3. Click and Brush

In this section we introduce the click and brush concept and detail some contributory operations and tools that enable click and brush.

3.1. The Click and Brush concept

Click and brush is a *compound brush* operation that is enacted by direct manipulation. The concept allows the user to drill down into the data using multiple constraints. We summarise the Click and brush operation by the following four processes.

- The user brushes and selects some elements. This can be achieved through any type, size or form of brush. The selected elements are highlighted to the user. This can be implemented in a number of different ways. For example, selection can be achieved by dragging a bounding box around elements of interest or continuously brushing over elements.
- The user fixes that information into a selection-list by clicking on any highlighted item. The data in the selection-list may be displayed to the user.
- 3. The user can brush further elements to show intersections and correlations with the newly brushed information; these subsequent brush operations affect the fixed elements (either they are highlighted in a different colour or only the elements that intersect remain). The results, of

this further brushing operation, dynamically update as the user brushes over other elements or any brushable items.

4. Finally, the user can either fix more elements, or backstep in a history of selections to return to previous selections, or delete the selection-list completely.

Important concepts are further expanded and clarified below.

Selection-list The selection-list contains details of every item that has been fixed after a brushing operation.

Fixing This operation stores the current highlighted elements into the selection-list. It is useful to allow the user to brush a list of interesting but (perhaps) non-spatially adjacent items.

Selection-table The elements in the selection list are presented to the user in the form of a table (a *selection-table*). As we shall see in the examples below the user can constrain the selection list based on further brushing operations of various aspects of this selection-table. The information in the table is updated dynamically whenever the user brushes new information.

Multiple views Especially views of different types, i.e. multiform views [Rob05], are used to stimulate and produce the interactions and brushing operations. Such diverse views allows the user to explore the information through different perspectives. For example the main visualization could be a 2D scatterplot, another a 3D view, finally a tabular view with labels or text fields containing the subsequent information.

Further brushing After some brushed data has been stored and fixed in the selection-table the user can perform further brushing operations. This allows the user to drill down into the data using the direct compound brushing model. For instance, given a selection-list S, and that a user subsequently brushes over an element 'a', then each element in S that shares something in common with 'a' are highlighted to the user. This further highlight could be depicted in various ways, e.g. (1) the new information is depicted in an additional view (a specialization of the information), (2) the new highlight is shown in a different colour, or (3) only the new information is shown in the current view, i.e. the non-intersecting elements are hidden or shaded in some way. Alternatively, the user can brush over any element of the selection-table; the various are intersections appropriately displayed.

Brushable items There are various aspects of the system that could be usefully made brushable, to engender compound operations. For example, any element of any visualization, multiple items could be used, or any column/row title of the data-table.

Intersections There are many operations that can be usefully applied for compound brushing; as detailed at the end of the related work, section 2.

3.2. Discussion and Issues

The main difference between Click and Brush and compound brushing [War94] [Che04] is that Click and brush is a direct manipulation tool. For instance, the work by Chen uses an external higraph that the user manipulates. The novelty of click and brush is that it allows easy selection of intersecting data that is normally only achieved using indirect manipulation techniques such as dynamic query sliders. Techniques such as angular brushing allow the users to perform data filtering on top of the displayed visualization and qualitative brushing allows users to select features of interest, which are then filtered and then mapped to the display. Click and brush on the other hand combines these two different types of brushing interaction, by selecting subsets visually on the display and by then filtering out data from that subset which can then be mapped back to the display. In this way correlating or intersecting data can be quickly and easily extracted from the visualization, which enables users to identify relationships and patterns within the subset.

There are a number of issues that need to be considered in the implementation of click and brush. Issues such as coordinating or linking the brush action to the different types of visualization views, how is highlighting achieved and speed considerations. Click and brush requires that are two or more types of visualization and coordination between these multivariate views is an important consideration as an interaction in one has to have a logical resulting action in another.

Highlighting is another issue that requires careful consideration and depends very much on the type of visualization that is used. In our implementation we use scatterplots as the main visualization with a data table to display parameters. To highlight elements we allow the user to select the type of highlighting to be used which can either be color or using a specialization view. However, there are alternatives such as focusing and context, masking and distortion.

The final implementation issue we discuss is how the data is stored so the rapid querying to filter out intersecting data is achieved. Brushing is a real time interaction technique and in a system a large amount of data being visualized using a brute force approach to query the data will not achieve the desired interaction speeds. Therefore, a way of representing the data so that rapid querying is achieved is needed. As in the paper by Frau, Roberts and Boukhelifa [FRB05] we use an associated array to store lists of similar data points allowing rapid querying and interaction which is performed in real time.

4. Implementation and Worked Examples

'Click and Brush' has been implemented in an email visualization tool Figure 1. The tool allows users to visualize an email archive in multiple scatterplots with the x-axis being the date and the y-axis the time. Emails are plotted on the

scatterplot based on the date and time which they are received by the user. Naturally, there is a great deal more data about an email that can be extracted, such as subject threads, count of words, frequency analysis of common words, to details of the size of the email or number and type of attachments. Our selection-table currently displays the To, From, Subject and Date fields of any selected data points.

All of the views are coordinated together, such that any brushing operation in one view causes all corresponding points in related views to be highlighted. The Figure 1 depicts five main areas, where, (a) is the menu, (b) is the selection-table view, (c) shows information about a particular email, and (d, e) depict scatter plot views of the information (which can display the same information or perhaps one is a zoom-in of the other, or can display a specialization of the information). Using this tool we describe two worked examples in which 'click and brush' can be used to explore the visualization and find or test hypothesis about the underlying data. We apply the tool to an email archive of the authors own email.

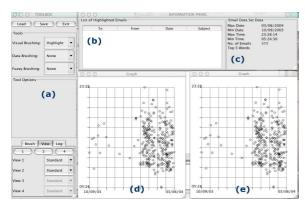


Figure 1: Screen shot of the email visualization tool used to demonstrate the Click and Brush principles. The figure depicts the tool's five main areas: (a) the menu which controls the brushing and intersection operations, (b) is the selectiontable view, which depicts the current selected information, (c) shows information about a particular email, and (d, e) depict scatter plot views of the information (which can display the same information or perhaps one is a zoom-in of the other, or can display a specialization of the information).

4.1. Example 1

In the first example we load in the Laboratory meetings archive folder into the visualization system. In this example we are interested in a smaller date range in this folder so we select a subset of data points, through directly zooming into the plot (Figure 2 A). From this focused view we can brush over the elements; in fact, we notice that on the left side of the visualization (at an earlier date) more of the emails are

sent by one person than on the right (at a later date). To test this hypothesis we click on one of the data points on the left and choose to highlight in another colour rather than a specialization view. By *further brushing* over one name in the 'From' field of the selection-table all the emails in this subset that have the same value for the From field are subsequently highlighted; Figure 2 B shows the result after brushing one name (depicting the emails more to the left), and Figure 2 C shows the result after brushing another name (more emails are on the right hand side). These result coincide with one secretary leaving and another arriving.

By doing this series of interactions the user is able to test the observations that they have made. The system as well allows user to see the results in all views to observe how these interactions relate to the visualization as a whole.

4.2. Example 2

In the second example we load in an email archive that contains all emails received over the course of a year. After exploring this data we discover that there are a whole series of emails with the subject line 'lab meetings'. It is interesting to investigate how many emails have been received with this Subject, sent to a specific mailing list, and whether that list is still in use. We can select one of these points and by brushing over the subject line we can highlight all elements with this subject line. Furthermore, we can 'click' on this variable and zoom in to this subset (see, Figure 3A). In the new view we find a data point that is to a mailing list, which is no longer in use, and decide to view the quantity of emails that were sent to this mailing list about this subject. Brushing over this variable we can see the number of emails to this mailing group (Figure 3B). Again this selection is mirrored in all views enabling users to see the effect across all of the email archive.

From these series of interactions we demonstrate the accumulative nature of the 'Click and Brush' technique. We are able to click on a subset of the visualization and zoom the view. We are also able to brush over a variable in the data table view and zoom in the same way. Using this set of interactions we are able to explore and find relationships in the data. This can then be used to drive further explorations, and as a result it is possible to see that the lab meetings subject is also sent to multiple mailing lists.

5. Summary and Conclusions

In this paper we have proposed a new brushing technique called 'Click and Brush' which allows user to find correlations and intersections in a visualization by brushing over elements in a visualization and then variables displayed in a selection-table. This type of interaction is most commonly associated with indirect manipulation techniques such as dynamic query slider however, our technique provides this interaction with a direct manipulation technique: brushing.

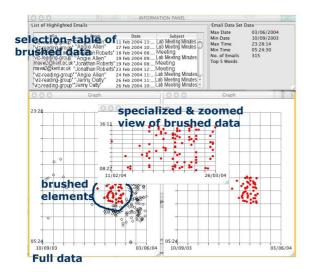


Figure 2 A.

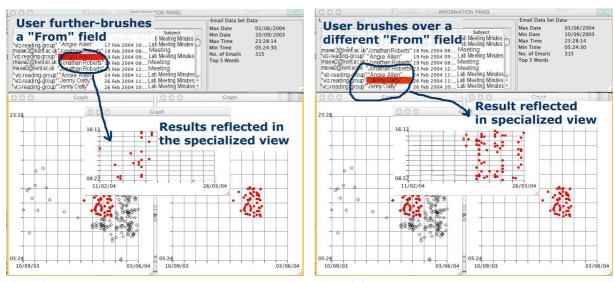


Figure 2 B. Figure 2 C.

Figure 2: The user can brush and select a group of elements; these are displayed in a separate view, as shown in 2 A. These points can be further explored through additional brushes. For example, screen shot 2 B demonstrates the results when the user brushes over a From field on the selection-table, and screen shot 2 C shows a subsequent brush on another field of the same selection-table.

The advantages to this direct compound brushing methodology is that users are provided with the tools to interact directly with the visualization; by providing users with a way of finding intersections based on variables of the elements displayed we are combining the advantages of indirect manipulation with a direct manipulation technique. On the other hand, a disadvantage of this method (and more generally with any direct manipulation techniques) is that the

interface is transparent and the user may not perceive the operation fully. However, the use of the history list (a non-linear history of previous brushing operations) helps the user overcome this problem. We believe that this technique could be expanded and developed further, perhaps generating a directly manipulated compound chain of other operations.

We have implemented click and brush as part of an email

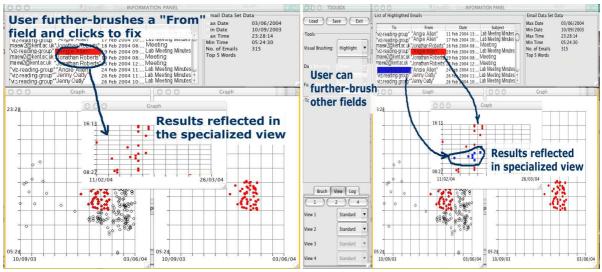


Figure 3 A Figure 3 B

Figure 3: Part 3A shows the result after the user has brushed over an area. In this case the elements are also displayed in a separate zoomed window. The user can subsequently brush over the selection-table (like the previous example) to select additional items. This subsequent data specialization can be further clicked-and-brushed, the result of which is shown in 3B. Each additional brushed elements, of a compound brush, are displayed in a different colour.

visualization tool which has been developed using the Java programming language. We are planning to further test this tool and the ideas presented in this paper, in order to ascertain its benefits given a set of different tasks, in comparison to indirect manipulation techniques. However, it is our opinion that click and brush would give users an advantage in finding correlating data and relationships in the data when compared to (i) traditional brushing techniques and (ii) indirect manipulation techniques such as dynamic query sliders.

References

- [AS94] AHLBERG C., SHNEIDERMAN B.: Visual information seeking using the filmfinder. In *CHI '94: Conference companion on Human factors in computing systems* (1994), ACM Press, pp. 433–434.
- [BC87] BECKER R. A., CLEVELAND W. S.: Brushing Scatterplots. *Technometrics* 29, 2 (1987), 127–142.
- [Ber81] BERTIN J.: Graphics and Graphic Information Processing. Walter de Gruyter Inc, Jan. 1981. Trans. by W.J.Berg and P.Scott.
- [Che04] CHEN H.: Compound brushing explained. *Information Visualization 3*, 2 (2004), 96–108.
- [CRMK95] CHUAH M. C., ROTH S. F., MATTIS J., KOLOJEJCHICK J.: SDM: Selective dynamic manipulation of visualizations. In *ACM Symposium on User Interface Software and Technology* (1995), pp. 61–70.
- [DE98] DIX A., ELLIS G.: Starting simple adding

- value to static visualisation through simple interaction. In *AVI98*, *ACM Press* (1998), T. Catarci M. F. Costabile G. S., L.Tarantino, (Eds.), pp. 124–134.
- [FRB05] FRAU S., ROBERTS J. C., BOUKHELIFA N.: Dynamic coordinated email visualization. In WSCG05 13th International Conference on Computer Graphics, Visualization and Computer Vision' 2005 (Plzen, Czech Republic, January 2005), Skala V., (Ed.).
- [HLD02] HAUSER H., LEDERMANN F., DOLEISCH H.: Angular brushing for extended parallel coordinates. In *In Proc. of the IEEE Symposium on Information Visualization* (2002), pp. 127–130.
- [Kli93] KLINKER G. J.: An environment for telecollaborative data exploration. In *VIS '93: Proceedings Visualization '93* (1993), pp. 110–117.
- [LBS*03] LI Q., BAO X., SONG C., ZHANG J., NORTH C.: Dynamic query sliders vs. brushing histograms. In CHI '03: CHI '03 extended abstracts on Human factors in computing systems (2003), ACM Press, pp. 834–835.
- [LRY02] LOWEKAMP B., RHEINGANS P., YOO T. S.: Exploring surface characteristics with interactive gaussian images (a case study). In VIS '02: Proceedings Visualization '02 (2002), IEEE Computer Society.
- [MAFT75] MARY A. FISHERKELLER J. H. F., TUKEY J. W.: Prim-9: An interactive multidimensional data display and analysis system. *Dynamic Graphics for Statistics* (1975), 91–109.

- [MW95] MARTIN A. R., WARD M. O.: High dimensional brushing for interactive exploration of multivariate data. In *VIS '95: Proceedings Visualization '95* (1995), IEEE Computer Society, p. 271.
- [Rob05] ROBERTS J. C.: Exploratory visualization with multiple linked views. Amsterdam: Elseviers, 2005. A.MacEachren and M-J.Kraak and J.Dykes, Eds, Chapter 8
- [Shn83] SHNEIDERMAN B.: Direct manipulation. a step beyond programming languages. *IEEE Transactions on Computers 16*, 8 (August 1983), 57–69.
- [Sii03] SIIRTOLA H.: Combining parallel coordinates with the reorderable matrix. In *Proceedings Coordinated and Multiple Views in Exploratory Visualization (CMV03)* (July 2003), J.C.Roberts, (Ed.), pp. 63–74.
- [vWvL93] VAN WIJK J. J., VAN LIERE R.: Hyperslice: visualization of scalar functions of many variables. In *VIS* '93: Proceedings Visualization '93 (1993), pp. 119–125.

- [War94] WARD M. O.: Xmdvtool: integrating multiple methods for visualizing multivariate data. In VIS '94: Proceedings Visualization '94 (1994), IEEE Computer Society Press, pp. 326–333.
- [War97] WARD M. O.: Creating and manipulating ndimensional brushes. In *Proceedings of Joint Statistical Meeting* (1997), pp. 6–14.
- [WB96] WONG P. C., BERGERON R. D.: Multiresolution multidimensional wavelet brushing. In VIS '96: Proceedings Visualization '96 (1996), IEEE Computer Society Press, pp. 141–148.
- [WB97] WONG P. C., BERGERON R. D.: Brushing techniques for exploring volume datasets. In *IEEE Visualization* '97 (1997), Yagel R., Hagen H., (Eds.), pp. 429–432.