# Interactive Data Visualization using Focusing and Linking

Andreas Buja
Bellcore
Morristown, NJ 07960

John Alan McDonald John Michalak Werner Stuetzle
Dept. of Statistics, U. of Washington
Seattle, WA 98195

**Abstract**

*This paper discusses two basic principles for interactive visualization of high dimensional data: focusing and linking. The paper and the accompanying video give examples of how graphical data analysis methods based on focusing and linking are used in applications including linguistics, geographic information systems, time series analysis, and the analysis of multi-channel images arising in radiology and remote sensing.*

## 1 Introduction

Most work in scientific visualization is concerned with high quality rendering of three-dimensional objects representing scalar or vector valued functions of $R^3$. Displays are typically created in batch mode, and interaction with the displays is limited.

In contrast, research in statistical graphics has concentrated on methods for visualizing high-dimensional data, using real time motion and interaction. For an overview, see Cleveland and McGill [11]. We have recently abstracted out from this research two concepts, which we refer to as *focusing* and *linking* [9], that seem to unify much of this diverse methodology.

To display complicated information, like a large program, an automobile and all its parts, or some multivariate statistical data, a common instinct is to draw a picture that is equally complicated, such as printing the program on the screen in a small font, rendering the automobile and its parts as transparent solids, or presenting the data as a tableau of Chernoff faces [10]. Attempts at such dense encoding are seldom successful. It is usually more effective to construct a number of simple, easy to understand displays, each focused clearly on a particular aspect of the underlying data.

Focusing techniques may involve selecting subsets, dimension reduction, or some more general manipulation of the layout of information on the page or screen. Examples of subset selection techniques are panning and zooming [8], and slicing [17]. Examples of dimension reduction techniques are projection and false coloring of multi-channel images [18,19].

Techniques for more general layout manipulation include univariate data transformations, logical zooming (demonstrated in the accompanying video) [9], and a variety of techniques from human-computer interface research for adapting to a user's "focus of attention" or "point of interest"[24], such as generalized fisheye views [16,13], Rooms [20], and Cone Trees [32].

Methods for focusing can be automatic, or interactive, or some combination of the two. An example of automatic focusing is exploratory projection pursuit as originally proposed by Friedman and Tukey [15]. The Prim-9 [14] and Orion [28,25,26] systems provided interactive versions of projection pursuit.

A consequence of focusing is that each view will only convey partial information about the data. We can compensate for this fact by displaying multiple views. Multiple views, however, should not be regarded in isolation. They need to be linked so that the information contained in individual views can be integrated into a coherent image of the data as a whole.

How views can be linked depends on whether they are displayed in sequence over time, or in parallel, simultaneously.

The principal mechanism for linking views over time is through smooth change in the position of objects on the screen. Examples for this kind of linking are rotating three-dimensional point clouds and a higher dimensional generalization, the Grand Tour [3,6]. More generally, any smooth animation can be considered as a set of linked multiple views, spread out over time.

There are a number of techniques for displaying multidimensional point data through simultaneous multiple views that are linked by drawing lines connecting the points in the different views corresponding to the same observation. Examples are Andrews' plots [2], parallel coordinate plots [23,34], and m-and-n plots [12]. The window "wiring diagrams" in Rooms [20] are an example of linking corresponding objects by lines in a very different context.



CH3046-0/91/0000/0156/$01.00 © 1991 IEEE

*Painting multiple views* is an alternative linking technique that is at least as effective for multidimensional point sets and is more easily generalized to other types of data. Painting multiple views integrates scattered information by marking corresponding parts of multiple displays with color (or some other form of highlighting). It is a generalization of "scatterplot brushing", which dates to the late 70's [30,28,26,29,4].

In this paper and the associated video we demonstrate a number of graphical methods that are examples of focusing and linking in applications, including linguistics, geographic information systems, time series analysis, and the analysis of multi-channel images arising in radiology and remote sensing.

## 2 Scatterplot brushing in geographic information systems

Figure 1 shows two plots of data from the Places Rated Almanac [5], a map of the 300 "places" in the right pannel, and a scatterplot of climate versus housing cost in the left panel. There is a cluster of places with mild climate and expensive housing in the upper right part of the scatterplot. To find out where they are located, we "painted" the cluster in the scatterplot, changing the glyphs representing the corresponding places. Glyphs representing the same places in the map have changed as well; they all lie on the California coast.

The map and the scatterplot convey two data dimensions each; by linking the two plots, we perceive four-dimensional structure. Linking views is crucial since seeing separate two-dimensional views is not enough to reconstruct four-dimensional data.

It is natural for a user to think of painting operations as a direct manipulation of the underlying database [22,29]. With this model in mind, the user expects painting in one view to change other views.

## 3 Painting discrete data

Viewing painting as direct manipulation suggests generalizing scatterplot painting to other kinds of data displays, for example tables of discrete data.

Figure 9 shows a collection of interactively painted plots produced by the Antelope system [27]. The data displayed were obtained from Mieko Ogura and Bill Wang of the Project on Linguistic Analysis at UC Berkeley [31]. In the upper left hand corner of the screen is a scatterplot of the latitudes and longitudes of 311 sites in England (and the Isle of Man). The plot on the lower left shows the villages classified according to the pronounciation of the word **great**. Similarly, the plot in the lower middle shows the villages classified according to the pronounciation of the word **beans**. Each cell in a table represents the subset of villages using a particular pronounciation. In the two-dimensional table on the right, a cell represents those villages with a particular pair of pronounciations for the two words. The size of the colored rectangle in each cell is proportional to the number of villages which use the corresponding pronounciation or pair of pronounciations.

Painting a cell in a table is equivalent to changing the color attribute of all the villages in the cell's subset. Cells display the colors of their subset's cases (when they are not all the same color) using a divided color bar. The percentage of a divided color bar that is, e.g., green is the same as the percentage of green cases in the subset.

In this example, painting reveals spatial coherence of the pronunciation — in other words, regional dialects.

## 4 Focusing and linking in time series analysis

We next illustrate how focusing and linking can be used in time series analysis. The data consist of 900 measurements of tidal levels, taken at 45 minute intervals, covering about 27 days.

First we examine the data using panning and zooming, which are simple examples of interactive focusing.

In the top panel of figure 2, the tidal levels are plotted against time. The plot reveals patterns among low and high tides. There are two kinds of high tides, and also two kinds of low tides, a low one of each and a high one of each. The low high tides and the high high tides approach each other and ultimately cross and change roles; the same is true for the low low and high low tides.

The middle panel shows the same data, after rescaling by stretching the horizontal axis and shrinking the vertical axis. It makes the quasi-periodicity and the smoothness of the series apparent. It is also easier to see that the usual sequence of tides is "high high", "low low", "low high", "high low". However, it is harder to see the crossing pattern.

Next, in the bottom panel, we expand both axes to zoom in on a 20 hour section. Panning horizontally, we discover a segment of the series containing an outlier

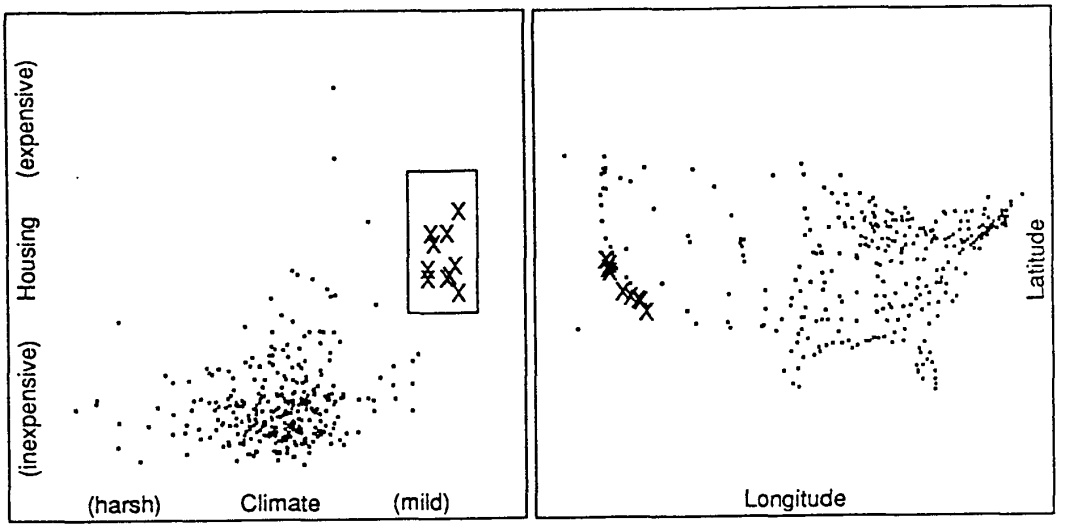


Figure 1: Plots of Places Rated data: metropolitan areas with mild climate and expensive housing highlighted.

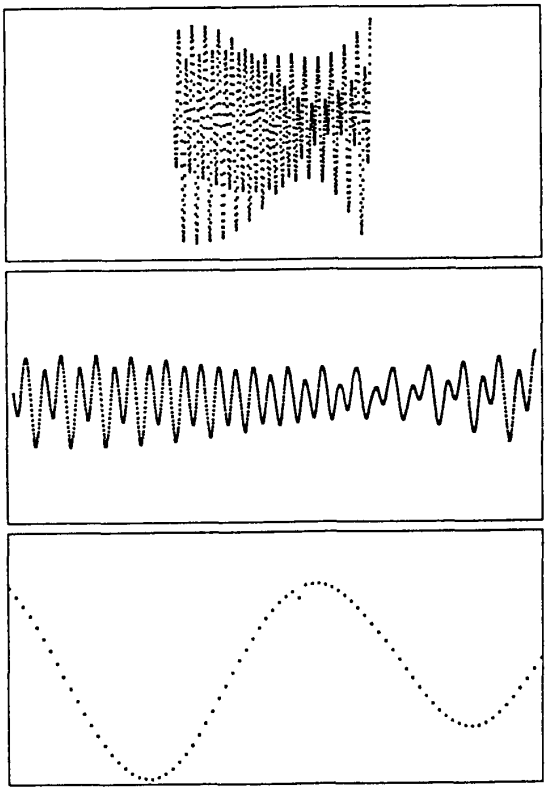

Figure 2: Panning and zooming tidal data.

that is invisible in the other two plots.

Next we use painting to have a closer look at the periodicities in tidal levels. The two left panels of figure 3 show tidal level plotted against time, whereas the right panels display a *lag plot* of the present tidal level versus the level three hours earlier.

The lag plot features a peculiar double loop pattern, and a natural question is to ask how it relates to time. We can answer this question by linking the time series plot with the lag plot; We paint time intervals in the time series plot and observe the corresponding points in the lag plot. Painting a time interval towards the beginning of the series (top panels of figure 3) shows that those observations lie on a double loop running along the periphery of the lag plot, while observations towards the end (bottom panels of figure 3) lie along a double loop in the interior. The properties of the lag plot revealed by painting give qualitative indications of facts which can otherwise only be detected with spectral analysis: the double loop indicates that there are two basic frequencies, one double the other. The shift in orbits is due to the presence of a very low frequency component.

## 5 Painting multi-channel image data

Our next example is from medical imaging. We examine a magnetic resonance image of a human head (obtained from David Haynor, Dept. of Radiology, U. of Washington). The image, shown in figure 10,

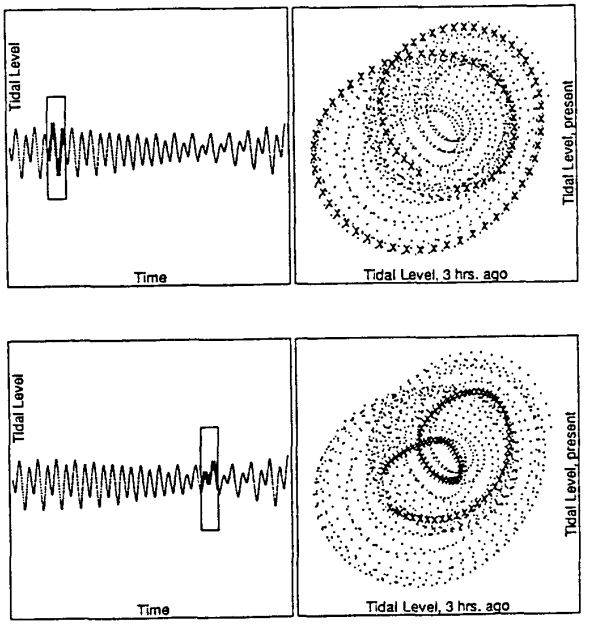


Figure 3: Lag plot of tidal data.

consists of 256x256 pixels. For each pixel we have 2 spatial coordinates (the pixel's location in the image), and the values of three variables, proton density, t1, and t2, that depend on the response of the tissue at this location to the imaging process.

On the left side of figure 10 is a grey scale image of the proton density channel. On the right is a scatterplot of t1 versus t2. The scatterplot is highly structured, having a number of well defined clusters. Linking the scatterplot to the image by painting makes it possible to see how the clusters in the scatterplot are related to the anatomy displayed in the image. Figure 11 shows the result of painting three of the clusters. We see that the red cluster corresponds to grey matter in the brain, the green cluster to white matter, and the blue cluster to corneal and spinal fluid. Alternatively, we could paint anatomical structures in the image and see where the corresponding pixels fall in the scatterplot.

## 6 Grand Tour and painting in the analysis of remote sensing data

Our final example shows how the Grand Tour, an interactive focusing method, can be used together with painting in the analysis of multispectral image data.

Figure 12 shows a Landsat MSS image of the confluence of the Rio Solimões and the Rio Negro near Manaus, Brazil, taken by Landsat 2 on July 31, 1977. The plate was produced by the WISP image processing system [33]. WISP has an interface that allows image pixels to be selected for analysis with the Data Viewer [7,8,21], a system for displaying multivariate data using a variety of techniques based on real time motion and interaction, including painting.

Among other things, WISP supports analysis of multispectral images using mixing models [1]. The basic idea of a mixing model is to decompose the observed spectrum of a pixel (which, in Landsat MSS, covers approximately an 80 meter square) into a combination of a small number of known (laboratory-measured) spectra of "pure" substances. One issue addressed with the Data Viewer and painting is whether a (convex) linear combination of spectra is sufficient or a more complex, non-linear mixing model is required.

In our example, we selected pixels from 5 rectangular areas in the image for examination with the Data Viewer: the small square in the upper part of the image covers pixels that are forest vegetation; the square in the lower right is grasslands; the square in the lower left is muddy water of the Rio Solimões; the square in the middle left is dark water of the Rio Negro; and the long narrow rectangle in the middle is a transect across the confluence (mixing) of the muddy and dark water.

Each selected pixel becomes a case in a Data Viewer data set, with six attributes, namely the $x$ and $y$ coordinates in the image, and the values for the four Landsat MSS spectral bands. We are specifically interested in whether the pixels in the transect can be represented as a convex combination of muddy and dark water spectra. This is true if the transect pixels lie along a line between the muddy and dark water pixels, in the four dimensional spectral space.

The selected pixels are displayed in two Data Viewer windows, shown in Figure 4: The left window, labeled `Brushing` shows a scatterplot of the original x and y pixel coordinates. The right plot shows a still from a Grand Tour [3,6] through the four-dimensional spectral space.

A Grand Tour surveys structure in multidimensional spaces by rapidly (10 frames per second) displaying a *smooth* sequence of projections of the data onto two-dimensional subspaces. Because the projection plane varies smoothly over time, we can follow individual points or clusters of points. Three-dimensional rotation is a special case of a Grand Tour.

Watching a dynamic Grand Tour, as shown in the

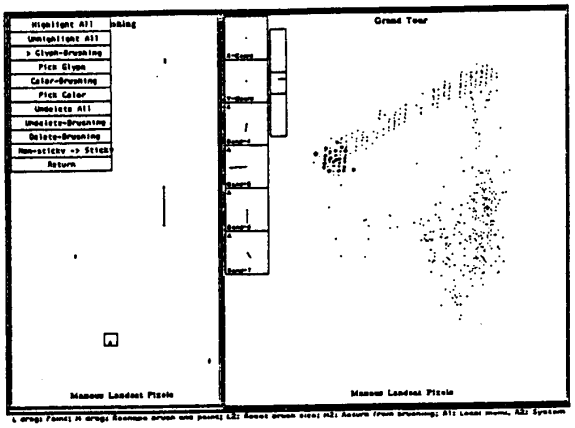


Figure 4: Highlighting muddy water.

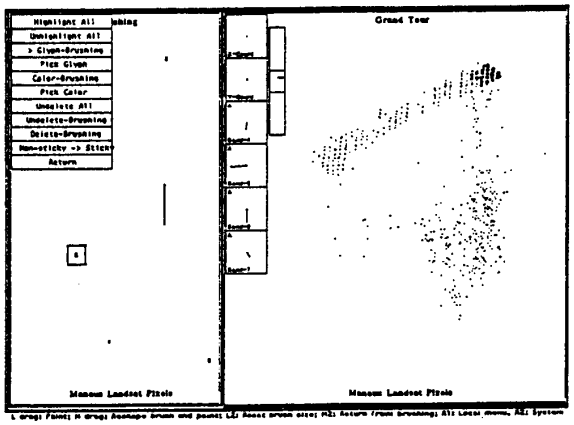


Figure 5: Highlighting clear water.

accompanying video [9], we can verify that the rod (the linear structure in the bottom left of the Grand Tour still) is in fact linear in four-dimensional space, because it remains linear in all projections.

In figures 4 and 5, known muddy and clear water pixels are highlighted, to show where they fall in this projection of the four dimensional spectral space. Figures 6 through 8 show a sweep across the transect of the confluence, confirming that these pixels are in fact, convex linear combinations of dark and muddy water pixels.

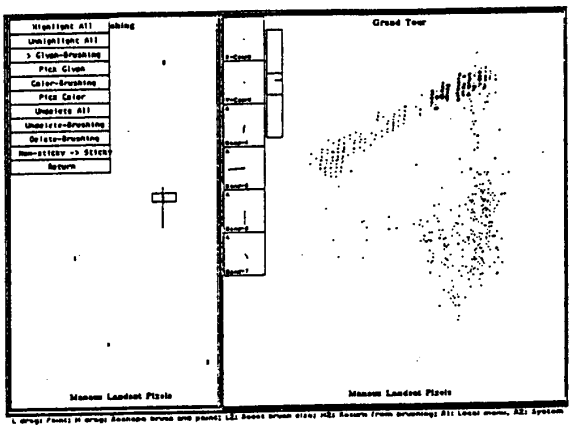


Figure 6: The top of the confluence.

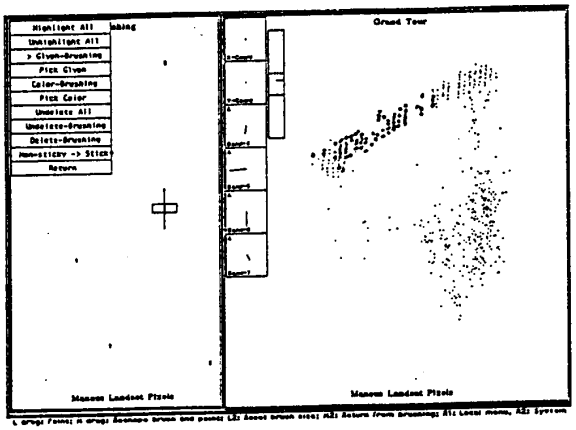


Figure 7: The middle of the confluence.

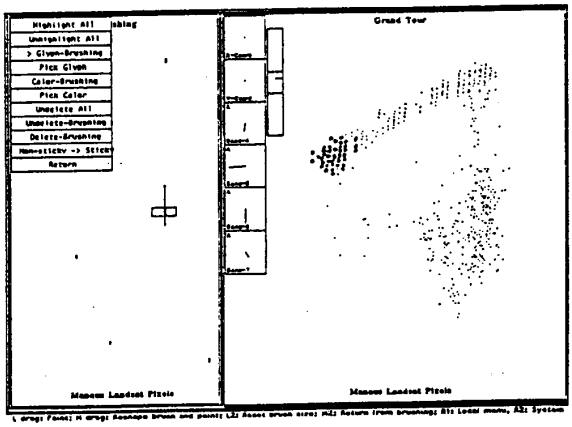


Figure 8: The bottom of the confluence.

## 7 Conclusions

Focusing and linking are principles that offer a solution to the problem of visual overload. Instead of maximizing the information in a single view, it is better to provide tools for quickly generating multiple views, each focussed on a different aspect of the data. Multiple views, however, should not be regarded in isolation. Linking makes it possible to integrate partial information contained in individual views into a coherent image of the data as a whole.

## 8 Acknowledgements

This work was partially supported by Bellcore, the Department of Energy under grant DE-FG06-85-ER25006, the Office of Naval Research under grant N00014-86-K-0069, and the National Science Foundation under grant DMS-8504359. We thank Steve Willis for providing the data and for his help in producing the figures in section 6.

## References

[1] John B. Adams, Milton O. Smith, and Alan R. Gillespie. Simple models for complex natural surfaces: A strategy for the hyperspectral era of remote sensing. In *IGARRS'89, the IEEE International Geoscience and Remote Sensing symposium*, 1989.

[2] David F. Andrews. Plots of high-dimensional data. *Biometrics*, 28:125–136, 1972.

[3] D. Asimov. The grand tour: a tool for viewing multidimensional data. *SIAM Journal of Scientific and Statistical Computing*, 6:128–143, 1985.

[4] R.A. Becker and W.S. Cleveland. Brushing scatterplots. *Technometrics*, 29:127–142, 1987.

[5] Richard Boyer and David Savageau. *Places Rated Almanac*. Rand MacNally, Chicago, 1985.

[6] Andreas Buja and Daniel Asimov. Grand Tour methods: an outline. In *Computer Science and Statistics: Proc. 17th Symp. on the Interface*, pages 63–67, Amsterdam, 1986. Elsevier.

[7] Andreas Buja, Catherine Hurley, and John Alan McDonald. A Data Viewer for multivariate data. In *Computer Science and Statistics: Proc. 18th Symp. on the Interface*, pages 171–174, Washington, D.C., 1987. ASA.

[8] Andreas Buja, Catherine Hurley, and John Alan McDonald. Elements of a viewing pipeline for data analysis. In W.S. Cleveland and M.E. McGill, editors, *Dynamic Graphics for Statistics*. Wadsworth and Brooks/Cole, Belmont, Ca., 1988.

[9] Andreas Buja, John A. McDonald, John Michalak, Werner Stuetzle, and Steve Willis. Visualization of quantitative data, 1990. A 27 minute video tape; Dept of Statistics, U. of Washington.

[10] Herman Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68:361–368, 1973.

[11] W.S. Cleveland and M.E. McGill. *Dynamic Graphics for Statistics*. Wadsworth and Brooks/Cole, Belmont, Ca., 1988.

[12] Persi Diaconis and Jerome H. Friedman. M and N plots. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday*, pages 425–447. Academic Press, New York, 1983.

[13] Kim M. Fairchild, Stephen E. Poltrock, and George W. Furnas. SemNet: three-dimensional graphic representations of large knowledge bases. In Raymonde Guindon, editor, *Cognitive Science and its applications for human-computer interaction*, chapter 5, pages 201–233. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

[14] Mary Anne Fisherkeller, Jerome H. Friedman, and John W. Tukey. Prim-9, an interactive multidimensional data display and analysis system. In *Proc. of Pacific 75, ACM Regional Conference.*, 1974.

[15] Jerome H. Friedman and John W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.*, C-23, 1974.

[16] George W. Furnas. Generalized fisheye views. In *Proceedings of CHI'86, Human Factors in Computing Systems*, New York, 1986. ACM.

[17] George W. Furnas. Dimensionality constraints on projection and section views of high dimensional loci. In *Computer Science and Statistics: Proc. of the 20 Symp. on the Interface*, Washington, D.C., 1988. ASA.

[20] D. Austin Henderson, Jr. and Stuart K. Card. Rooms: The use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics*, 5(3):211–243, July 1986.

[21] Catherine Hurley. *The Data Viewer: a program for graphical data analysis.* PhD thesis, Dept. of Statistics, U. of Washington, 1987.

[22] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. Direct manipulation interfaces. In D. A. Norman and S. W. Draper, editors, *User Centered System Design: New Perspectives in Human-Computer Interaction.* Erlbaum, Hillsdale NJ, 1986.

[23] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1:69–91, 1985.

[24] Jock D. MacKinlay, Stuart K. Card, and George G. Robertson. Rapid controlled movement through a virtual 3d workspace. *Computer Graphics*, 24(4):171–176, 1990.

[25] John Alan McDonald. Exploring data with the Orion I workstation, 1982. A 25 minute, 16mm sound film, which demonstrates programs described in McDonald (1982) It is available for loan from: Jerome H. Friedman, Computation Research Group, Bin # 88, SLAC, P.O. Box 4349, Stanford, California 94305.

[26] John Alan McDonald. Projection pursuit regression with the Orion I workstation, 1982. a 20 minute, 16mm color sound film, which demonstrates programs described in McDonald (1982) It is available for loan from: Jerome H. Friedman, Computation Research Group, Bin # 88, SLAC, P.O. Box 4349, Stanford, California 94305.

[27] John Alan McDonald. Antelope: data analysis with object-oriented programming and constraints. In *Proc. of the 1986 Joint Statistical Meetings, Stat. Comp. Sect.*, 1986.

[28] John Alan McDonald. Interactive graphics for data analysis. In W.S. Cleveland and M.E. McGill, editors, *Dynamic Graphics for Statistics*, pages 247–275. Wadsworth and Brooks/Cole, Belmont, CA, 1988. Ph.D. thesis, Dept. of Statistics, Stanford, June 1982.

[29] John Alan McDonald, Werner Stuetzle, and Andreas Buja. Painting multiple views of complex objects. *SIGPLAN Notices*, 25(10):245–257, 1990. (Proceedings OOPSLA/ECOOP'90).

[30] C. M. Newton. Graphics: from alpha to omega in data analysis. In P.C.C. Wang, editor, *Graphical Representation of Multivariate Data.* Academic Press, New York, 1978. Proceedings of the Symp. on Graphical Representation of Multivariate Data, Naval Postgraduate School, Monterey, Ca., Feb 24, 1978.

[31] Mieko Ogura and William Wang. Spatial distribution of the Great Vowel Shift in England. In *Language Transmission and Change.* Blackwell, London, England, 1989.

[32] George G. Robertson, Jock D. Mackinlay, and Stuart K. Card. Cone Trees: animated 3d visualizations of hierarchical information. Technical Report SSL-90-79, Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304, 1990.

[33] P. Shippert, G. Bradshaw, and S. Willis. *Washington Image and Spectral Package (WISP) Preliminary Documentation.* Remote Sensing Laboratory, Dept. of Geological Sciences, U. of Washington, Seattle, WA 98195, May 11, 1989.

[34] Edward J. Wegman. Hyperdimensional data analysis using parallel coordinates. *JASA*, 85(411):664–675, 1990.

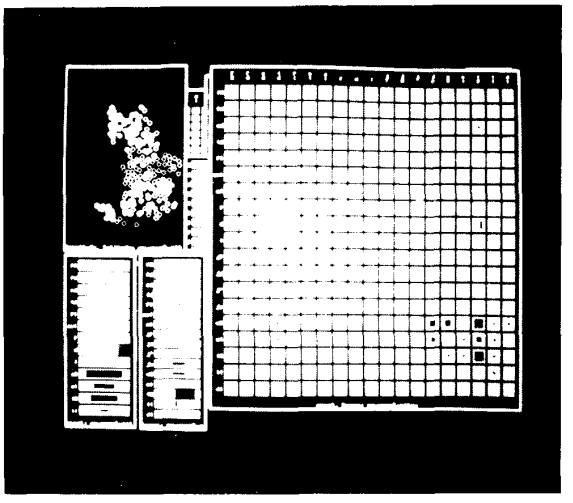

Figure 9: A map and contingency tables of English vowel data.

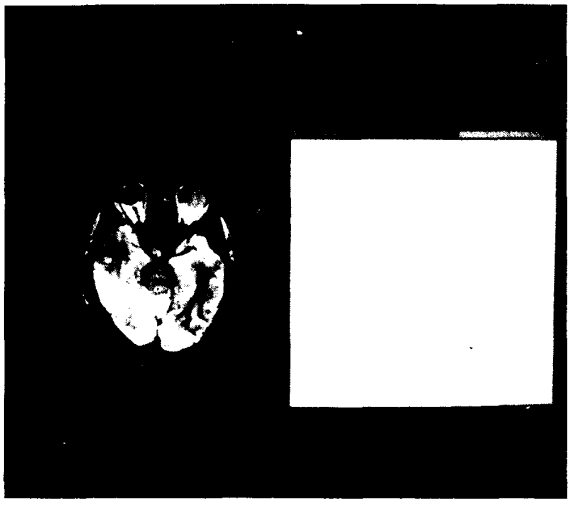

Figure 10: A proton density image and a scatterplot of t1 vs t2.

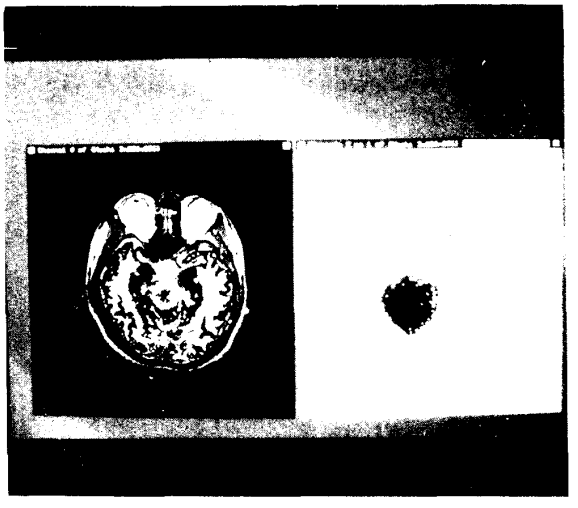

Figure 11: The proton density image and scatterplot of t1 vs t2 after painting.

Figure 12: Landsat MSS image of the confluence of the Rio Solimões and the Rio Negro near Manaus, Brazil.

***(See color plates, page 419.)***