VisFiles
Presentation Techniques for Time-series Data

T. Todd Elvins
San Diego Supercomputer Center
UC San Diego

Introduction

This is the first in a series of columns on the subject of data visualization. I'm excited to have this opportunity because it will give me an excuse to finally learn about some subjects that I've wanted to learn about but never gave myself the time. Gordon Cameron, *Computer Graphics* Editor, has given me some latitude on topics so future visualization columns will deal with color selection, data visualization systems and APIs, data visualization research, information visualization, user interface issues, tips and hints, etc. If you have comments on this column or suggestions for future columns, please send me email *(todd@acm.org)*.

In this issue I will give an overview of some familiar, and some probably unfamiliar, time-series data presentation techniques. Time-series data visualization is becoming an increasingly important topic as data archives grow exponentially and computing becomes ubiquitous. Some researchers claim that time-series data can be imaged using techniques similar to those applied to spatial data (since time is most often linear), and some claim that temporal data is a much different beast. For example, temporal data can have multiple strands and events, features not present in spatial data. There are also many interesting unsolved problems and data handling issues associated with temporal data:

- Boolean and other mathematical operations
- Data storage and temporal databases
- Data query and retrieval
- Data generalization (handling data at vastly different resolutions)
- Data models and representation
- Data presentation

I have lately been working on designing a user interface for a simple geographic information system capable of manipulating and compositing time-series geo-referenced data. This project started me thinking about the different ways of visually representing temporal data. Without going into much detail on any one method, following are all the methods I could find within arms length in my office. Most of these methods allow the scientist or visualizer to present the time-series in a single image, while the final method uses an animation technique. Good techniques for visually presenting time-series data will:

- Reveal spatial and temporal relationships
- Uncover patterns
- Show variability

Static Image Techniques

Probably the best known example of static temporal data visualization is the image created by Charles Joseph Minard in 1869 and reprinted in Figure 1. This flow image shows the diminishing size of Napoleon's army as it advanced toward Moscow in 1812, and then froze during a winter retreat. The thickness of the trails indicates the size of the army at any point in time and the westward retreating trail is linked to a temperature scale. This image does a fine job of communicating the spatial nature of the march, the dwindling size of the army over time and the effect of freezing temperature on troop movement.

In 1981, Bertin [1] identified seven primitives for displaying temporal data in static images: size, value, texture, color, orientation, shape and location in space, all of which can be found in Minard's classic work from 100 years prior.

A number of more contemporary presentation methods are described by Monmonier [7]. Figure 2a shows the familiar line plot and Figure 2b shows a bar graph. Additional techniques include chess maps (Figure 2c) where sequential maps are juxtaposed so that the viewer can compare the patterns at sequential times. In dance maps, a series of map symbols are used to describe events over a period of time and show the transitions from place to place. The isochronic map of California in Figure 2d can be considered a kind of dance map. Change maps are also a single map, but symbols vary in value, size or some other appropriate visual variable to represent the direction, rate or absolute amount of change. The four presentation techniques in Figure 2 each use a subset of Bertin's primitives to reveal temporal relationships, patterns and show variability.

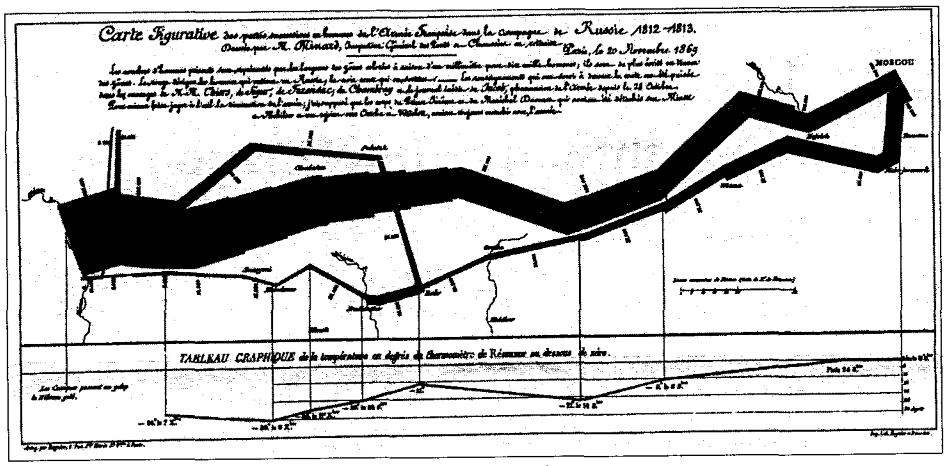

Figure 1: Napoleon's march on Moscow in 1812 by Charles Joseph Minard.

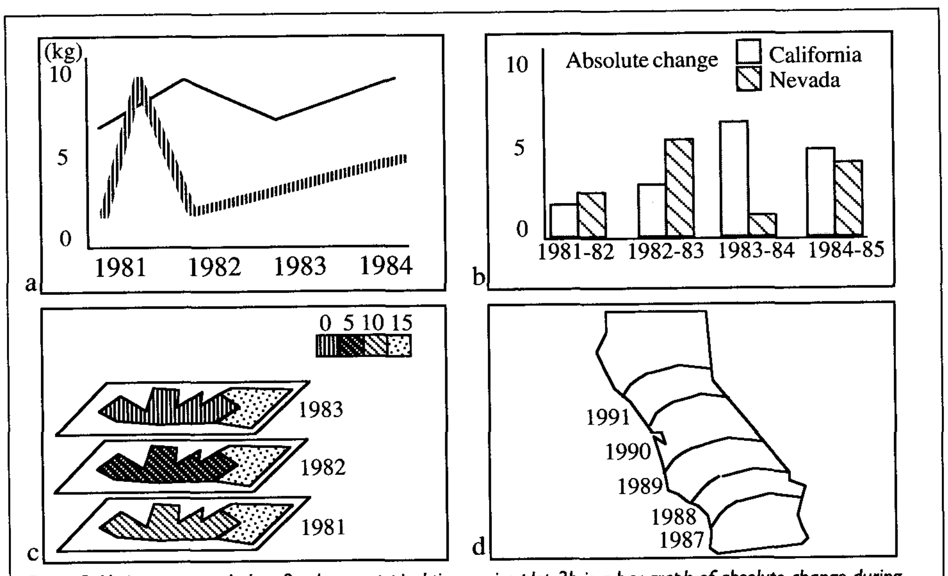

Figure 2: Various temporal plots. 2a shows a typical time-series plot. 2b is a bar graph of absolute change during a time period. 2c is a chess plot and should have a figure legend, and 2d shows isochronic lines in a sort of dance map (over the past decade fish taco restaurants have expanded north through California up into Oregon).

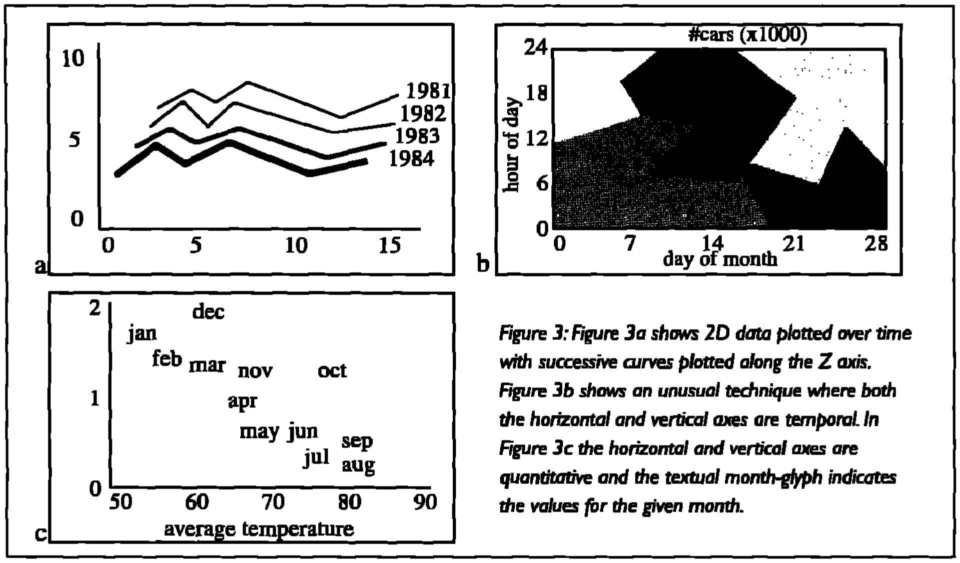

Figure 3: Figure 3a shows 2D data plotted over time with successive curves plotted along the Z axis. Figure 3b shows an unusual technique where both the horizontal and vertical axes are temporal. In Figure 3c the horizontal and vertical axes are quantitative and the textual month-glyph indicates the values for the given month.

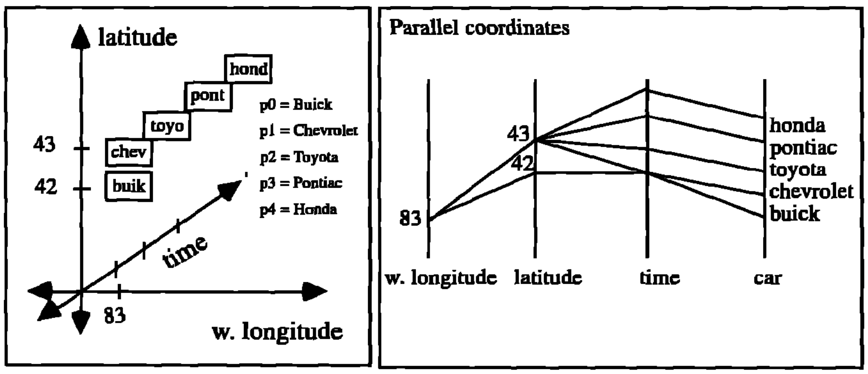

Figure 4: In Figure 4a on the left, points p0 - p4 are all at the same longitude and points p1 - p4 are all at the same latitude. Points p0 and p1 occur at the same time while p1 - p4 all occur at different times. All of this information is also shown in Figure 4b on the right using parallel coordinates.

Figure 3 shows some additional time series presentation techniques. Figure 3a shows an extension to the familiar line plot where curves are juxtaposed along the Z axis (into the paper) to show two measurements over a multiyear period. This is a sort of 3D chess map and could be further extended by coloring and/or surfacing the curves. Figure 3b shows a technique where dense data fill a raster. Both the horizontal and vertical axes are temporal in this diagram which might be used to count cars on the freeway. Figure 3c shows a scatterplot technique where the horizontal and vertical axes are quantitative and a textual month-glyph is located at the intersection of the temperature and precipitation for the month indicated. Plots such as those in Figures 3b and 3c may exhibit linear patterns indicating a possible pair-wise relation in the data.

Parallel coordinates [4] can be both a static and an animation technique. In the static case, each individual data dimension is plotted on a vertical axis as in Figure 4b. For example, west longitude might be on one vertical axis, latitude on another, time of day on another and car type on a fourth. Polylines are then used to connect corresponding values on the axes. For example, a 7 a.m. Chevrolet car theft in Detroit would appear as a polyline from 7 a.m. on the time axis to 83 degrees on the west longitude axis, to 42 degrees on the latitude axis, to Chevrolet on the car type axis. Using parallel coordinates, relationships between adjacent dimensions appear as similarly oriented line segments. For example in our car theft example, a number of nearly parallel line segments between time of day and latitude might indicate a direct correlation. Adjacent dimensions with negative correlation cause crossed connecting line segments between the dimensions. Interesting patterns can be revealed using this technique, however, determining an optimal horizontal ordering of the vertical axes is considered a non-trivial problem.

To make parallel coordinates an animation technique, one of the dimension values can be varied to show the effect on the other dimensions. For example, the west longitude value in Figure 4b could be varied to find related latitudes, times and car types.

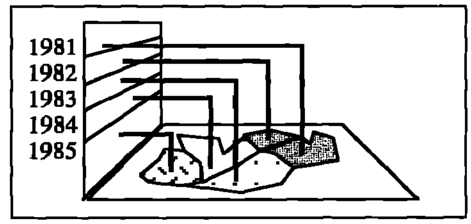

Figure 5: A 3D plot where locations are tied to dates and/or times on a timeline. Putting time on one of the spatial axes is sometimes called dimension shifting.

Returning to more traditional Cartesian plots, a static or dynamic timeline can alternatively be included within a spatial data plot and then spatial data points, lines, regions or rasters can be associated with a timestamp on the timeline. One way to associate spatial data with a timestamp is to draw a line between them. Although limited by the number of association lines that can be drawn, this technique is an interesting foray into the world of 3D plots. This technique, pictured in Figure 5, may also be useful for plotting spatial data height-fields such as topography, contours and for other geographic applications.

A more sophisticated method that may be more difficult to interpret is the use of biplots [7]. In Figure 6 we see a number of statistical curves plotted over a period of time. Principle-component analysis is used to collapse the census division and year variables to points in a new space. In this space the eigenvector coefficients serve as coordinates for plotting the years, and the component scores serve as coordinates for plotting the census divisions. The origin is near the center, and the horizontal and vertical axes are not drawn. Both adapted from [7], Figure 6b is computed from the data shown in Figure 6a. Monmonier comments on Figure 6b:

"The configuration of place-points demonstrates the temporal polarity caused by a temporary labor shortage in New England

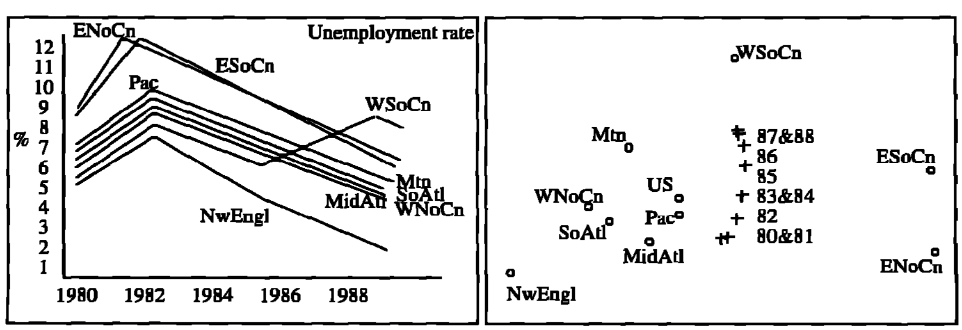

Figure 6: Figure 6a on the left is a traditional statistical plot, and Figure 6b on the right is a biplot of the same information.

and severe economic distress in the East South Central and East North Central divisions. The biplot also shows the effect of slowed recovery in the West South Central and Mountain divisions and suggests that the Pacific division was most similar to the nation as a whole. ... That 1988 is farthest removed from 1980 suggests a significant alteration in the geographic pattern of unemployment, not a simple cyclic return to earlier conditions."

An Animation Technique

"Animation provides the illusion of movement to show time variability with time-dependent multivariate symbols and icons such as graduated symbols, histograms and time-lines." [2]. MacEachren [5] suggests that presenting data in attribute value order may present additional insights not apparent in chronologically ordered data visualizations. Varying the duration and rate of change should also be considered. For example, blinking a subset of the data on and off can allow a viewer to isolate the subset without removing the visual context. Human vision is especially sensitive to change making animation an effective means of communicating temporal information. Since there are a large number of animation methods, I will just mention one animation technique for presenting data, and will save a discussion of data interaction and exploration for a future article.

Here at the San Diego Supercomputer Center, John Helly and I have developed an effective technique for visualizing San Diego Bay sediment sample data which are sparse in time and space [3]. Test sites for a particular day are highlighted with a white disk, station locations are indicated by a black dot and sample values are displayed as an opaque colored disk behind the black dot. Non-detects are displayed as blue diamonds. In the animation, the white disks show the sampling pattern while the sample values and sample locations remain fixed. An ancillary line plot is incorporated in the lower left showing the precipitation for the year. A yellow vertical marker moves from left to right during the animation indicating the rainfall in the temporal vicinity of each sediment sampling date.

Summary

More information on each of the techniques described above can be found in the references. Time-series data are an interesting challenge. The visualizer can attempt to convey the patterns and relationships in the data in a single image or, if this doesn't work, can create an animation. I enjoy doing both. Successfully authoring a single image that clearly gets the message across is a satisfying experience. Single images are easier to publish and hang on the wall than animations, and as Minard showed, a single image can be stunning.

Authoring an informative animation is also satisfying. I'm always amazed at the fluidity achieved by simply stringing together what I know to be static rendered snapshots. Animations engage the human eye's sensitivity to movement and leverage time-based media to convey time-based information.

Sound can also be used as part of an animated time-series. An additional data attribute or parameter can be played in the soundtrack for example. One final note on animations is a word of caution. Be aware that rendering only every *nth* frame can hide potentially important data points and poorly chosen colormaps can be worse than no colormap at all.

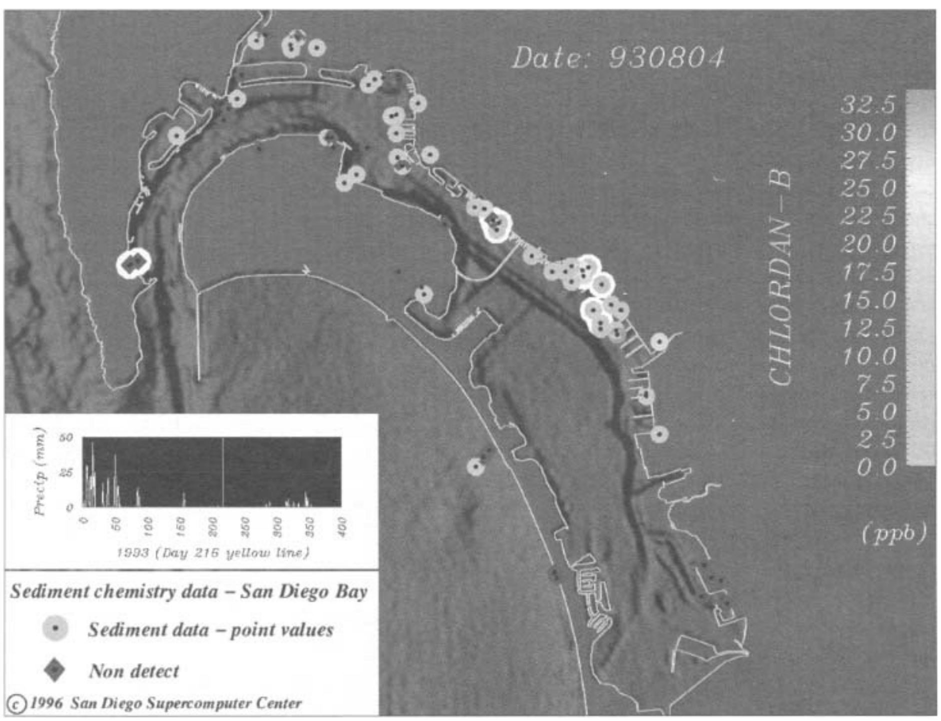

Figure 7: Sediment chemistry in San Diego Bay. One frame from an animation (MPEG available at http://www.sdsc.edu/sdbay).

References

1. Bertin, Jacques. *Graphics and Graphic Information Processing*, translated by William J. Berg and Paul Scott. Berlin; New York: Walter de Gryter & Co. Publisher, 1981.
2. Egenhofer, Max and Reginald G. Golledge. "Time in Geographic Space: Report on the Specialist Meeting of Research Initiative 10," *National Center for Geographic Information and Analysis, Report 94-9, 1994. ftp://ncgia.ucsb.edu/pub/Publications/tech_reports/94/94-9/*
3. Elvins, T. Todd, Stuart Gage and John Helly. Integration, visualization, and analysis of GIS Data. In Proceedings of the IBM Data Explorer User's Group (October 1996, San Jose, CA). *http://www.sdsc.edu/~todd/dx*
4. Inselberg, Alfred and Bernard Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In Proceedings of the IEEE Visualization 90 Conference (San Jose, CA) pp. 361-378, Oct., 1990.
5. MacEachren, A. "Time as a cartographic variable," *Visualization in Geographic Information Systems*, D. Unwin and H. Hearnshaw (eds.), Belhaven Press, pp. 115-130, 1993.
6. Minard, Charles Joseph. From the collection: Tableauz Graphiques et Cartes Figuratives de M. Minard, 1845-1869, a portfolio of his work held by the Biblioteque de l'ecole Nationale des Ponts et Chaussees, Paris.
7. Monmonier, Mark. "Strategies for the visualization geographic time-series data," *Cartographica*, 27(1), pp 30-45. 1990.

T. Todd Elvins is a Staff Scientist at San Diego Supercomputer Center, and a Computer Engineering Ph.D. candidate at the University of California, San Diego. His research interests include spatio-temporal user interfaces, wayfinding in virtual environments, volume visualization and learning to surf (in the ocean).

T. Todd Elvins
San Diego Supercomputer Center
University of California, San Diego
MC 0505
La Jolla, CA 92093-0505, U.S.A.
Email: *todd@acm.org*
http://www.sdsc.edu/~todd/