StretchPlot: Interactive Visualization of Multi-Dimensional Trajectory Data

Paul Murray*
University of Illinois at Chicago

Angus Forbes†
University of Illinois at Chicago

Abstract

The visualization of space-time trajectory data is challenging due to the fact that, in most cases, the data inherently exists in more than two dimensions. In many space-time visualizations, additional variables must be mapped into other visual attributes such as color or size. Given that substantive analyses of trajectory data will likely be concerned with variables beyond space and time, we propose a method that allows trajectory positions to be determined by an arbitrary number of variables in a multi-dimensional space-time dataset. Our method is based on Star Coordinates [5], a prior technique for interactively positioning multi-dimensional data. Our technique is catered specifically to multi-dimensional space-time trajectory data. The technique is demonstrated using a large multi-dimensional space-time dataset related to touring musicians.

1 Introduction

The visualization of space-time trajectory data is challenging due to the fact that, in most cases, the data inherently exists in more than two dimensions. One common solution to this problem is to display a two-dimensional projection of trajectories embedded in a three-dimensional cuboid, called a Space-Time Cube (STC) [2, 1]. While STC techniques effectively display three dimensions of space and time, additional variables must be mapped into other visual attributes, such as color or size. Here we present a work-in-progress technique that allows trajectory positions to be determined by an arbitrary number of variables in a multi-dimensional dataset. The technique is an extension of prior work, Star Coordinates [4, 5], that allows multi-dimensional data to be positioned according to user-defined axis vectors. Our technique allows a user to map trajectories into a high-dimensional space that includes the positional variables of space and time in addition to other variables of interest to an analyst.

2 Related Work

The Star Coordinates [4, 5] technique allows for flexible, user-driven mappings of multi-dimensional data into a two-dimensional space. In this technique, multiple axes – each representing one dimension in a dataset – are arranged in a radial pattern, with each axis having a distinct length and angle. The length and direction of each axis can be manipulated by the user. Each data point in a multi-dimensional dataset is then mapped into a position in two-dimensional space by treating each axis as a unit vector – the position of one data point is calculated through a linear combination of each vector with each data point. The result is an arbitrary mapping of high-dimensional data into a two-dimensional space.

Through the interactive arrangement of multiple dimensions, Star Coordinates have been shown to help users discover clusters in hierarchically-defined datasets. Additionally, they are effective in revealing high-dimensional associations between variables and data points in a multi-factor analysis.

*e-mail: pmurra5@gmail.uic.edu
†e-mail: angus.forbes@gmail.com

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

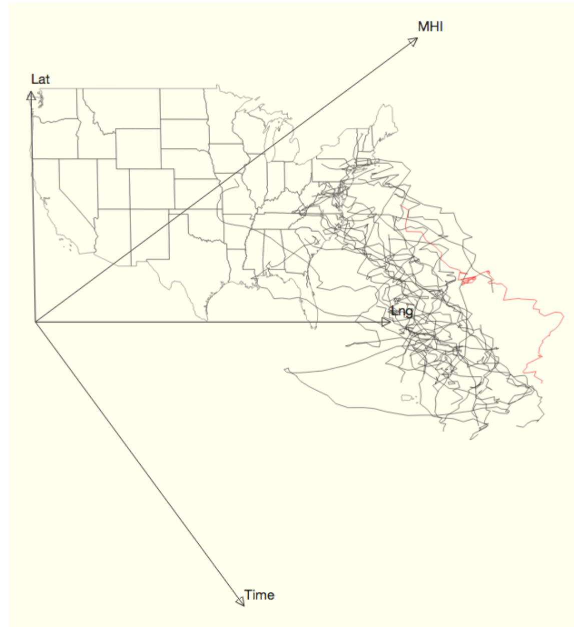

Figure 1: StretchPlot showing 30-day moving average trajectories, positioned according to the variables of latitude, longitude, time, and the median household income of event locations.

Star Coordinates has primarily been used with static high-dimensional data, where the entire dataset is visualized as a cloud of points. Here, we consider an extension to the Star Coordinates technique that is geared specifically towards dynamic, spatio-temporal trajectory data, represented as connected linear sequences.

Typical space-time datasets involve at least the three dimensions of latitude, longitude, and time, and visualization techniques must therefore grapple with the problem of displaying multi-dimensional data in a two-dimensional space. Drawing on work by MacEachren [7], Andrienko and Andrienko [1, 2] have done considerable work examining one such technique called the space-time cube (STC). An STC is a representation of space-time data that places events within a hollow rectangular cuboid in three dimensions. Two dimensions of the cube represent latitude and longitude, while the third represents time. In most implementations, the STC can be interactively rotated and scaled. The display of a space-time cube in two dimensions is the result of a projection of the three-dimensional cuboid as viewed from some angle.

STCs are typically used to display the trajectories of a number of entities across some geographical space and over time. Many creative implementations of STCs exist, and various methods of clustering or distorting spatial data are explored, often with the aim of revealing trends, reducing occlusion, or revealing clusters in the data.

Despite this variety, implementations of STCs are almost always rooted in three dimensions – two dimensions of space and one dimension of time. Additional variables beyond from space and time are typically mapped into other visual attributes, such as color or size. In our framework, trajectories can be mapped into a space defined by any variable in a multi-dimensional dataset – including, but not limited to, space and time. This builds on the assumption that a substantive analysis of space-time data will likely include further variables that may be related to spatiotemporal position.

3 StretchPlot

Our work extends the Star Coordinates technique to work specifically with space-time trajectory data. Just as with Star Coordinates, users can interactively define the size and direction of a number of vectors, one for each variable in the dataset (including space and time).

Our system is designed to work with trajectory data, in which sequential events (and the variables associated with that event) are nested within entities. The position of each event is determined through a linear combination of several coordinate vectors with each vector's associated value for the event. Events, in turn are connected according to their temporal order, forming trajectories.

Users are able to select a variable from the dataset and interactively define the size and direction of that variable's coordinate vector. Positioning two coordinate vectors orthogonally, for example, will create a classic time-series plot. Users are able to continue to add and manipulate coordinate vectors for any number of variables in the dataset. When latitude and longitude are used as coordinate vectors, an option is provided to display a geographic map, for visual reference.

In addition to raw trajectory data, an option is provided to display n-moving average trend lines (where n is chosen interactively by the user). Using moving averages helps to avoid clutter and occlusion. This is especially true of human mobility data, where transit (via airplane, for example) is often rapid, resulting in many criss-crossing trajectories that occlude each other. The option to include n-moving average trend lines also stems from the assumption that an analyst would likely be interested in relative trends – how certain entities or groups *tend* to move in relation to each other – rather than the precise location of an entity at a given time.

The design of this technique was driven by the fact that space-time data is inherently multi-dimensional, and it is assumed that empirical analyses that utilize trajectory data will be concerned with some set of variables beyond spatio-temporal position. Along with space and time, additional variables (as long as they are interval or ratio scale) are used to determine the positions of trajectories, and the relative contribution and direction of any one variable is interactively controlled by the user. By design, specific positions and geographic patterns may be misinterpreted. The goal of the technique is to reveal high-dimensional relationships, rather than simply display geo-temporal positions.

The flexibility of this technique allows for a variety of configurations. Notably, when only the three vectors corresponding to latitude, longitude, and time are displayed (i.e. when all other vectors are of length zero), the display is akin to a projection of an STC. Likewise, classic time-series plots can be created by simply placing two vectors orthogonally. Aside from allowing for such flexibility, the user interaction demanded by StrechPlot likely aids in the perception of certain high-dimensional structures, as suggested by prior research related to Star Coordinates [5].

Mapping variables into spatial positions, rather than other visual attributes, has two key advantages. First, beyond a small number of variables, one will "run out" of attributes (e.g. size, color) available to be mapped into. Second, prior research [3] [8] has suggested that visual encodings exist in a hierarchy, where different encodings are interpreted with varying degrees of accuracy. Position is consistently found to lead to the most accurate interpretations of interval and ratio data, above other encodings such as size or color. Thus, mapping additional variables into other visual attributes aside from position implies a hierarchy of variables, implicitly suggesting that certain dimensions are more salient than others. In an exploratory analysis, one may wish to avoid implying such a hierarchy of variables. Finally, allowing all variables to map into position may more effectively reveal high-dimensional clustering.

4 Case Study

StretchPlot was used to explore a large dataset related to traveling musicians. The essential space-time data included the date and geographic location of performances given by over 3000 musicians over the span of four years. In addition, social and demographic data – such as median household income and racial distributions – was collected based on the geographic coordinates of each performance location. At the musician level (rather than the event level), popularity data was collected via weekly publications of national college radio "top 200" charts.

A number of multi-dimensional relationships between genre, time, popularity, and geography were explored. These exploratory analyses were based on a line of sociological research into the evolution of genre [6], which suggests several different stages, or "genre types", through which a musical genre can evolve. StretchPlot served as a means to visually explore the interactions of theory, history and time. The dynamic nature and multi-dimensionality of such a large space-time dataset demands a flexible interface, capable of displaying potentially high-level interactions. In exploring this data, important distinctions and clusters were discovered. For instance, while certain groups of musicians might travel to similar cities, when the relative prosperity of their destinations were considered, their trajectories diverged. In this context, a trajectory can be seen as moving not only through geographic space and time, but also through the abstract "space" of a number of demographic and social variables.

5 Conclusion

Future work will apply this technique to other data sets, and enable additional interactive tools for geospatial analysis. Additionally, we will carry out user studies to evaluate and asses the effectiveness of our technique.

References

[1] N. Andrienko and G. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.

[2] N. Andrienko, G. Andrienko, and P. Gatalsky. Visual data exploration using space-time cube. In *Proceedings of the 21st international cartographic conference, International Cartographic Association, Durban*, pages 10–16, 2003.

[3] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[4] E. Kandogan. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *Proceedings of the IEEE Information Visualization Symposium*, volume 650, page 22. Citeseer, 2000.

[5] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 107–116. ACM, 2001.

[6] J. C. Lena and R. A. Peterson. Classification as culture: Types and trajectories of music genres. *American Sociological Review*, 73(5):697–718, 2008.

[7] A. M. MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 2004.

[8] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.