Interactively Exploring Geotemporal Relationships in Demographic Data via Stretch Projections

Paul Murray
University of Illinois at Chicago
pmurra5@gmail.uic.edu

Angus Graeme Forbes
University of Illinois at Chicago
aforbes@uic.edu

ABSTRACT

This paper investigates the interactive projection of multivariate space-time data. Specifically, it investigates how complex datasets containing elements situated in space and time that include additional variables can be interactively explored to support effective multivariate geotemporal analyses. Given that substantive analyses of space-time trajectory data will likely be concerned with additional variables beyond space and time, we propose a novel technique, *stretch projections*, that allows trajectory positions to be determined by an arbitrary number of variables in a multidimensional space-time dataset. We present StretchPlot, an interactive prototype that allows multidimensional space-time trajectory data to be projected into a two-dimensional space according to multiple user-defined coordinate vectors. The StretchPlot tool is demonstrated using a large multidimensional space-time dataset that combines the trajectories of touring musicians with demographic data related to the locations of their performances. Examples are shown which identify relationships between multiple dimensions of space, time, and additional demographic variables, indicating that the use of of stretch projections may be useful for the exploratory analysis of multidimensional geotemporal data.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Spatial databases and GIS*

Keywords

Space-time data, Trajectories, Culture, Sociological data, Stretch projections

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. Including temporal data on a traditional map can be misleading, as data will often occlude itself (e.g., when many events occur in the same location but at different times), and in many cases the sequence of events will be difficult or impossible to discern. At the same time, a typical analysis of space-time trajectory data will likely contain variables beyond space and time. Here we present a novel interaction technique, *stretch projections*, that allows trajectory positions to be determined by an arbitrary number of variables in a multidimensional dataset. Trajectories are thus rendered as sequences of events that exist within a multidimensional space. Our technique includes an important element of interaction, allowing an analyst to observe relative differences between multiple dimensions. With this technique, the three dimensions of latitude, longitude, and time can be combined with any number of additional dimensions in a shared space. Our technique frames trajectories as the paths of entities that move through some high-dimensional space.

In this paper, we introduce our *stretch projection* technique and describe an interactive prototype, StretchPlot, used to explore trajectories within a geotemporal dataset. We describe in detail a specific use case scenario involving a multidimensional space-time dataset that combines the trajectories of touring musicians with demographic data related to the locations of their performances.

2. SPACE-TIME VISUALIZATIONS AS PROJECTIONS

In examining space-time data, researchers may wish to map a number of variables (including time, latitude, longitude, and others) into positions in a shared space. Mapping many variables into a two-dimensional space is routinely accomplished with the use of various dimension reduction and projection algorithms. Dimension reduction techniques are useful in that they produce high-level views of multidimensional data, grouping cases by similarity, and potentially revealing clusters of highly-associated data points. However, dimension reduction algorithms tend to be somewhat of a "black box," in that the viewer may have little intuitive understanding of why a given point was mapped to a particular space, or the extent to which different variables contributed to the position of each data point [16]. Rather than pass all data through a dimension reduction algorithm, an analyst may want to assign one variable to a particular orientation, or, for example, orient two variables orthogonally to compare their relative importance.

Given the definition of "projection" as a mapping of some set of variables into a lower-dimensional space, any two-dimensional visualization of three-dimensional space-time data can be thought of as a projection. Unlike traditional projections, however, visualizations of space-time data must deal with trajectories, rather than points. Visualizations must take trajectories that exist in a multidimensional space and map them – as connected paths – into a two dimensional space in some way. For example, StoryFlow [11] maps spatial positions to the vertical axis using a hierarchical ordering of spatial distances, while time is mapped linearly to the horizontal axis. Storygraph [17] arranges the dimensions of latitude and longitude as two vertical parallel coordinates, with time mapped to a horizontal axis between them. These examples limit themselves to visualizations of time and space; here, we consider trajectories that move through a multidimensional space, including variables of interest beyond location.

3. THE IMPORTANCE OF TRAJECTORIES IN SPACE-TIME VISUALIZATIONS

Space-time data is nested, with multiple events belonging to any one entity. When visualizing data related to movement, displaying connections between events is essential. Connections between events (i.e. with lines) serve two purposes. Connections define groups of events that are related to a given entity, and, with space-time data in particular, they define a temporal ordering, with each event connected only to the two events before and after it, reflecting the continuity of time. Groups of points connected by a continuous line represent events affiliated with an individual entity – in other words, cases are represented as collections of (ordered) events. Sequential rajectories – collections of ordered events belonging to an entity – are the essential unit of any visualization of space-time data. The importance of connections is particularly essential when displaying movement, and even moreso when many individual cases are represented on a single chart. Without the connectivity of a line chart, the movement of a case from one event to another is lost.

It is important that visualizations of trajectory data include paths of events connected by lines in order to differentiate the trajectory (i.e. the series of events) belonging to one case from the trajectories of other cases. In this context, the connections represent a *nominal* variable that distinguishes each case (each collection of events) from other cases. As a visual encoding, "connection" is included in the hierarchy presented by Cleveland and McGill [3] as the fourth-most effective encoding for nominal variables, behind position, color hue, and texture. Despite its lower ranking, it seems to be an intuitive choice for time-series representations, as the continuity of time is reflected in the continuity of the path.

Given the accuracy of positional encodings, together with the necessity of connected paths in trajectory data, an effective visualization of space-time data will encode trajectories as connected lines, while encoding events into positions in a shared space. Traditionally, trajectories are positioned according to some arrangement of latitude, longitude, and time. In our work, we allow the placement of trajectories on a two-dimensional projection plane to be determined by an arbitrary set of variables.

4. RELATED WORK

Our technique builds off of concepts introduced in prior work, Star Coordinates [7, 8], that allows multidimensional data to be positioned according to user-defined axis vectors. This 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.

The Star Coordinates technique allows for flexible, user-driven mappings of multidimensional 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 multidimensional 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 entities in a multi-factor analysis.

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 a technique similar to Star Coordinates that is geared specifically towards dynamic, spatio-temporal trajectory data, represented as connected linear sequences.

One common solution to the problem of rendering three-dimensional space-time data is to display a two-dimensional projection of trajectories embedded in a three-dimensional cuboid, called a Space-Time Cube (STC) [1, 2]. 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 [1], 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. Different views of an STC are akin to more traditional graphical representations. If an STC is rotated such that one of the two axes of geography is orthogonal to the viewer, it will show a typical time series chart, with time on the horizontal axis and either latitude or longitude on the vertical axis. When the axis of time is orthogonal to the viewer, the view of the STC is a simple map, with trajectories shown as connected paths.

Drawing on work by MacEachren [12], Andrienko and Andrienko [1, 2] have done considerable work examining STCs. 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

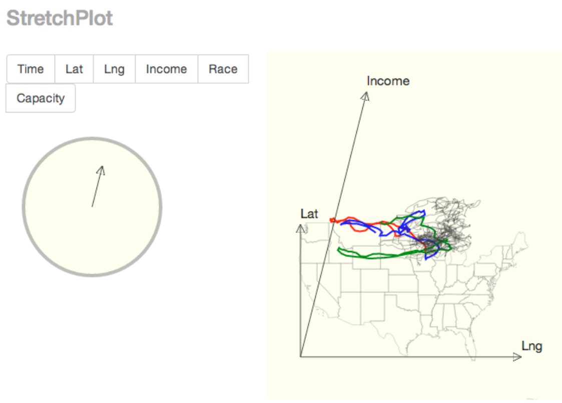

Figure 1: The StretchPlot interface. Users are presented with a list of variables available within the dataset, as well as a vector manipulation box that allows the user to manipulate the size and direction of a selected coordinate axis.

revealing clusters in the data. 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.

Despite a variety of creative examples, 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 line thickness [1, p. 6]. In our framework, trajectories can be mapped into a space defined by any number of variables in a multidimensional 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.

5. STRETCH PROJECTIONS

Our work extends the Star Coordinates technique and focuses specifically on space-time trajectory data. As with Star Coordinates, *stretch projections* allow users to interactively define the size and direction of a number of vectors, one for each variable in the dataset (including, but not limited to, space and time).

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

The StretchPlot application projects trajectories onto a two-dimensional projection plane by performing a linear combination of several coordinate vectors for each event in the trajectory. Initially, all vectors are of length zero. To manipulate one of the vectors, users choose from a list of all variables in the dataset. The user can then click and drag within a "vector manipulation area" to manipulate the given vector.

Figure 1 shows the basic StretchPlot interface, with buttons corresponding to each variable within the dataset, and the vector manipulation area allowing the user to determine the size and direction of each coordinate vector. As the vector is manipulated, the positions of trajectories within the projection space are updated, resulting in an interactive manipulation of trajectory data.

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 and data are episodic rather than continuous, 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 multidimensional, 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.

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 [8].

The interactive manipulation of multiple dimensions is a key compontent of our application. Aligning two or more dimensions in similar directions results in the conflation of variables, and one will not be able to determine the extent to which each variable is contributing to the position of a trajectory. However, through interaction, these relative contributions can be revealed visually through their movement. Interactively manipulating the coordinate axes of a multidimensional space changes the *relative* distance between cases (and between trajectories), revealing the relative contributions of a variable.

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, 13] has suggested that visual encodings exist in a hierarchy, where different encodings are interpreted with varying degrees of accuracy. Position, above other encodings such as size or color, is consistently found to lead to the most accurate interpretation of interval and ratio data. 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.

The aim here is to generate a better understanding of the multidimensional nature of the data through the detection and observation of clustered trajectories. Stretch Plot thus serves as a means to visually explore clustered trajectories that exist within a high-dimensional space of geographic, economic, and social data that is evolving through time.

6. CASE STUDY

StretchPlot was used to explore a large dataset related to traveling musicians. Data include 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.

Exploratory analyses were based on a line of sociological research into the evolution of genre by Jennifer Lena, who theorizes that musical genres exist as one of four different "genre types" [10]. Each genre type is defined by certain cultural and economic contexts. For example, industry-based genres are driven and marketed by large, wealthy corporations, with extensive ties to the production and entertainment industries. On the other hand, "avant-garde" genres consist of artists who perform in small, un-commercialized venues to meager audiences with little-to-no financial backing.

Importantly, Lena's framework is also dynamic, positing that genres evolve through time – genres follow trajectories through a multidimensional space that includes geographic, economic, and social variables. Genres themselves exist as trajectories. Questions that arise from Lena's framework are based on combinations of these variables: Which economic contexts are associated with high popularity? Do certain genres prosper more in certain geographical areas? Or with certain racial distributions?

Prior research has examined the evolution of certain genres over time, such as the spread of surf music out of southern California in the 1960's [14], and the movement of Jazz musicians between major US cities [15]. These examples are essentially case studies, targeting specific genres, and they aim to either provide support for a-priori hypotheses or to simply document historical trends. Stretch projections, on the other hand, are geared towards a much more exploratory approach.

6.1 Discovering multidimensional Trajectory Positions

Figure 2 illustrates five views of StretchPlot using event data from thirty musicians and bands over the span of six years, with a total of over 2,300 events. In Figure 2(A), raw trajectory data is shown, resulting in a high amount of overlap and occlusion. Figure 2(B) shows a simplified view of the data, achieved by calculating 30-day moving averages of each artists' trajectory. Despite the use of this smoothing technique, trajectories still greatly occlude each other, as they are "clumped" around an overall average event location. However, this view reveals that three of the trajectories stretch much further west than the others. These three trajectories have been highlighted for the purposes of illustration.

Figure 2(C) shows the result of manipulating the "time" coordinate vector so that it extends downward and to the right. Although StretchPlot only places data in two dimensions, Figure 2(C) is akin to an orthogonal projection of a three-dimensional STC. Comparing Figure 2(B) to 1(C) reveals the relative temporal order of events. In terms of average performance locations, the blue artist moved west over time before moving back towards the majority of trajectories in the sample; the red artist followed a similar pattern, yet their overall trajectory occurred later in time; the green artist first moved east before returning west, and finally moving back east. Without the added dimension of time, the relative temporal location of each trajectory would be impossible to discern.

Figure 2(D) shows the result of manipulating the "income" coordinate vector so that it extends upward. In this case, the "income" variable represents the median household income of the census block group in which each event occurred [18]. By comparing Figure 2(B) to 2(D), it is seen that when they performed in the west, the artists highlighted in red and blue performed, on average, at venues located in relatively high-income neighborhoods as compared to the artist highlighted in green. Finally, Figure 2(E) includes the four variables of Latitude, Longitude, Time, and Income. Note the increased distance between the red and green trajectories between Figure 2(C) and Figure 2(E), resulting from the fact that the green trajectory is situated in a relatively lower "income space."

Thus, a multidimensional story of three traveling artists can be told. All three artists, at some point in their careers, performed in locations that were located much more to the west relative to other artists in this sample. First, the blue artist "moved" west before returning back east (note that here the concept of "motion" is indirect, and refers to an artist's average performance location). The red artist followed a similar pattern – moving west and performing in relatively affluent neighborhoods – albeit after the blue artist. The green artist has the most dynamic trajectory. They began performing in relatively western locations – and in affluent neighborhoods – before moving east and performing in neighborhoods with relatively lower income. They later moved back west, but they performed in neighborhoods with relatively low household incomes.

In coming to these conclusions, it is critical to note the effect of movement produced by user interaction, which serves as a means of comparison across dimensions in the dataset. With only one of the the plots in Figure 2 alone, one would not be able to discern important multidimensional relationships. For example Figure 2(E) conflates many variables, and ex-

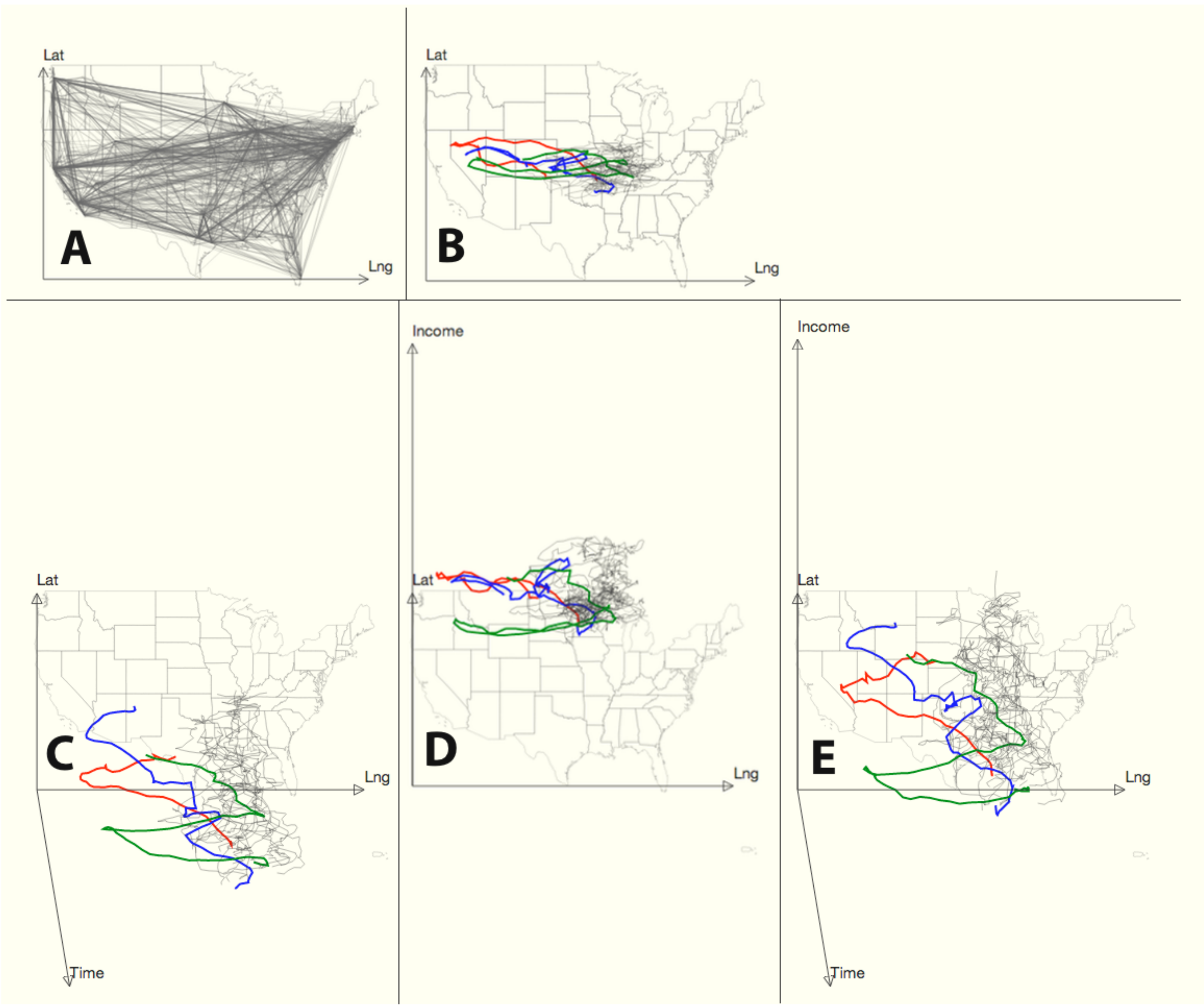

Figure 2: Five views of Stretch Plot, revealing multivariate trends amongst a number of trajectories. Figure A shows raw trajectory data. Figure B shows 30-day moving averages, with three trajectories highlighted. Figure C is the result of incorporating time as a coordinate vector stretching downwards, akin to an STC. Figure D adds a coordinate vector for income parallel to latitude. Figure E is the result of incorporating all four coordinate vectors of time, space, and income into a shared trajectory space.

actly which variables are contributing to a given trajectory's position cannot be determined. However, by comparing Figure 2(E) to Figure 2(C) – which occurs through user interaction – one can observe *relative* differences in position that occur as a result the addition of a variable.

7. DISCUSSION

Our case study provides an example of how complex multidimensional relationships can be discovered with stretch projections via interaction with our StretchPlot application. multidimensional relationships are revealed in two ways: First, each variable in a multidimensional dataset is encoded by position (i.e. by the position of a trajectory), rather than other attributes such as color or size. Second, the interactive nature of StretchPlot reveals the relative contribution of each variable to trajectory positions.

7.1 Visual Encodings of Multi-Dimensional Space-Time Data

It is common for data visualizations to assign variables to various visual encodings, such as position, size, or color. However, visual encodings can differ greatly in terms of how accurately they are perceived. Cleveland and McGill [3] propose an ordered hierarchy of visual encodings, based on how accurately viewers can determine the quantitative relationships that the encodings represent. A number of experiments confirm their hypothesis, suggesting that different graphical presentations exist in a ranked hierarchy, where different visual encodings are interpreted with varying degrees of accuracy [13]. In particular, they find that mapping data to position – rather than other attributes such as length or angle – is one of the most perceptually accurate methods of visualizing quantitative data. After confirming that visual mappings exist in an ordered hierarchy, Cleveland and McGill [3] suggest that effective graphs will utilize visual encodings as high up in the hierarchy as possible.

Given such a hierarchy of visual encodings, choosing to map two variables into two different visual encodings implies a hierarchy amongst variables. In an exploratory study, however, an analyst may wish to be agnostic to the importance of variables, and may not want to imply that any one variable is more important than another. In this sense, an unbiased mapping of variables to visual attributes is achieved by mapping all variables into the same visual encoding simultaneously.

8. FUTURE WORK

Similar to Star Coordinates [8], our technique is highly flexible. However, future implementations could allow for more structured configurations to be chosen by the user. As seen in our case study, the importance of interaction is in revealing relative differences across several different configurations of multiple dimensions. Instead of allowing the user to place coordinate axes in any arbitrary configuration, a more structured approach might simply allow the user to interactively transition between two pre-determined configurations. Future work might also incorporate algorithmic techniques related to related to the detection of similar event sequences [9, 19, 20], rather than relying on subjective judgements alone.

A well-designed user study will be crucial in order to verify the utility of our design. An important question is whether typical users will be able to independently reach similar conclusions about multidimensional trajectory data. A more detailed analysis would aim to identify which interaction techniques are most suited to revealing certain relationships for structures in the data. For instance, the use of motion might be well-suited to demonstrate the effect of one variable in a multidimensional dataset [4, 5, 6].

9. CONCLUSION

While a majority of space-time visualizations are concerned only with locations in physical space and time, the work presented here represents a first attempt at a flexible visualization of trajectories that exist in a multidimensional space. Our technique reveals relative differences amongst groups of trajectories in a high-dimensional space through the interactive manipulation of coordinate axes. As we continue to refine and analyze our technique, we hope to develop a comprehensive application for the exploratory analysis of multidimensional trajectory data.

10. 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*, pages 10–16, Durban, South Africa, 2003. International Cartographic Association.

[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] R. Etemadpour, P. Murray, and A. G. Forbes. Evaluating density-based motion for Big Data visual analytics. In *Proceedings of the IEEE Conference on Big Data*, Washington, DC, October 2014.

[5] A. G. Forbes, T. Höllerer, and G. Legrady. Behaviorism: A framework for dynamic data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1164–1171, November-December 2010.

[6] A. G. Forbes, C. Jette, and A. Predoehl. Analyzing intrinsic motion textures created from naturalistic video captures. In *Proceedings of the International Conference on Information Visualization Theory and Applications (IVAPP)*, pages 107–113, Lisbon, Portugal, January 2014.

[7] 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.

[8] 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.

[9] P. Köthur, M. Sips, A. Unger, J. Kuhlmann, and

D. Dransch. Interactive visual summaries for detection and assessment of spatiotemporal patterns in geospatial time series. *Information Visualization*, 13:283–298, 2014.

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

[11] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu. Storyflow: Tracking the evolution of stories. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2436–2445, 2013.

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

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

[14] D. F. McCarter. *Spatial analysis of surf music: 1961-1966*. PhD thesis, California State University, Northridge, 2012.

[15] D. J. Phillips. Jazz and the disconnected: City structural disconnectedness and the emergence of a jazz canon, 1897–1933. *American Journal of Sociology*, 117(2):420–483, 2011.

[16] M. Schaefer, L. Zhang, T. Schreck, A. Tatu, J. A. Lee, M. Verleysen, and D. A. Keim. Improving projection-based data analysis by feature space transformations. In *IS&T/SPIE Electronic Imaging*, pages 86540H–86540H. International Society for Optics and Photonics, 2013.

[17] A. Shrestha, Y. Zhu, B. Miller, and Y. Zhao. Storygraph: Telling stories from spatio-temporal data. In *Advances in Visual Computing*, pages 693–702. Springer, 2013.

[18] U.S. Census Bureau. *2012 American Community Survey, 5-Year Estimates*, 2012.

[19] T. von Landesberger, S. Bremm, T. Schreck, and D. W. Fellner. Feature-based automatic identification of interesting data segments in group movement data. *Information Visualization*, 13:190–212, 2014.

[20] K. Vrotsou, A. Ynnerman, and M. Cooper. Are we what we do? exploring group behaviour through user-defined event-sequence similarity. *Information Visualization*, 13:232–247, 2014.