Visualizing Multi-dimensional Clusters, Trends, and Outliers using Star Coordinates

Eser Kandogan

IBM Almaden Research Center
650 Harry Rd.
San Jose, CA 95120, USA
+1 408 927 1949

eser@us.ibm.com

ABSTRACT

Interactive visualizations are effective tools in mining scientific, engineering, and business data to support decision-making activities. Star Coordinates is proposed as a new multi-dimensional visualization technique, which supports various interactions to stimulate visual thinking in early stages of knowledge discovery process. In Star Coordinates, coordinate axes are arranged on a two-dimensional surface, where each axis shares the same origin point. Each multi-dimensional data element is represented by a point, where each attribute of the data contributes to its location through uniform encoding. Interaction features of Star Coordinates provide users the ability to apply various transformations dynamically, integrate and separate dimensions, analyze correlations of multiple dimensions, view clusters, trends, and outliers in the distribution of data, and query points based on data ranges. Our experience with Star Coordinates shows that it is particularly useful for the discovery of hierarchical clusters, and analysis of multiple factors providing insight in various real datasets including telecommunications churn.

Keywords

Multi-dimensional visualization, knowledge discovery.

1. INTRODUCTION

Visualization techniques are powerful sense-making tools that support knowledge workers in their decision-making activities by stimulating visual thinking. Fayyad et al. [6] recognizes the primary importance of user interaction and graphical representations for data as part of the whole knowledge discovery process to make patterns more understandable by users.

While most of the scientific, engineering, and business data is multi-dimensional; i.e. datasets contain typically more than three attributes of data, representing and making sense of multi-dimensional data has been challenging, partially due to the three-dimensional space we live in. The motivation for this work is to relieve the dimensionality curse on knowledge discovery through simple data representations that are derived from familiar and easy to understand lower dimensional representations.

One of the goals of visualization systems is to support knowledge workers in early stages of their data-understanding tasks. At this stage users' purpose is to gain insight into the distribution of the data, to explore interesting trends and clusters, and to discover outliers. Thus, the work done is primarily qualitative not quantitative analysis. It is assumed that once the users have a better overall understanding of the data, they will know where to look for numerical details for further analysis in the remaining stages of the knowledge discovery process. Then, through dimensionality reduction and the use of other means they can numerically support the hypotheses they formed earlier. Thus, it is expected that users are likely to tolerate loss of information in the initial steps and trade insight for certainty.

Use of simple visual representations uniform across all dimensions is crucial in the early stages, as users need to explore and compare a number of different options rapidly. It is too early in the process to expect users to map data attributes to visual cues such as color, size, direction, and position since users do not yet have a good understanding of their data. Besides, use of multiple encodings makes it difficult to compare trends and clusterings and to understand data distribution in high dimensional space.

The inspiration for the present work come from Bertin's Permutation Matrices [2], which allows users to rearrange rows and columns to discover patterns and clusters from coarse graphical depictions of data. In Permutation Matrices, data in each cell are represented using simple visuals such as black or white colored cells. Using this simple but coarse representation, it is possible to grasp the distribution of data and clusters without the need for exact data values. Users can tolerate loss of information for the sake of gaining insight into the data. The strength of the Permutation Matrices also lies in the interactivity that enables users to integrate and separate dimensions in the visual representation.

Another source of inspiration is Parallel Coordinates [11], which lays out coordinates in parallel, encoding each dimension uniformly through the same visual cue (i.e. position). In Parallel Coordinates each data element is represented as a line passing through the coordinate axes at the value of the element for that dimension. While Parallel Coordinates is a very powerful technique -especially for modeling relations- these visualizations require user expertise and knowledge of mathematical methods.

In summary, desired properties of a multi-dimensional visualization system for the early stages of exploration are:

- Encoding all data dimensions using a single visual cue,
- Minimal, simple visual representations of data, and
- Interactive techniques to integrate and separate dimensions.

This paper is organized as follows. First, the basic intuition and mathematics behind Star Coordinates is introduced, followed by descriptions of its interaction features through the use of various examples on a sample dataset. Next, telecommunications churn data is used to demonstrate how Star Coordinates can be used in the knowledge discovery process for understanding customer patterns and distributions. Then, a short review of existing multi-dimensional techniques is provided along with a comparison of Star Coordinates based on the characteristics of the visual transformation processes. The paper concludes with a sketch of future work.

2. STAR COORDINATES

The intuition behind Star Coordinates lies in the definition of dimension and coordinate systems. The n^{th} dimensional space can simply be defined as the space that is constructed through a 'sweep' of the $(n-1)^{st}$ dimensional space (Figure 1). While sweeping is done through a new orthogonal direction for each dimension in the Cartesian coordinate system, it can as well be an angular sweep as in the cylindrical and spherical coordinate systems. However, as Inselberg points out very appropriately orthogonality uses up space rapidly [12] due to the fact that every point in space is uniquely identified through projections back onto the coordinate axes. In this work we are going to relax this constraint a little bit to be able to build on users existing perception of low dimensional space and extend it to higher dimensions.

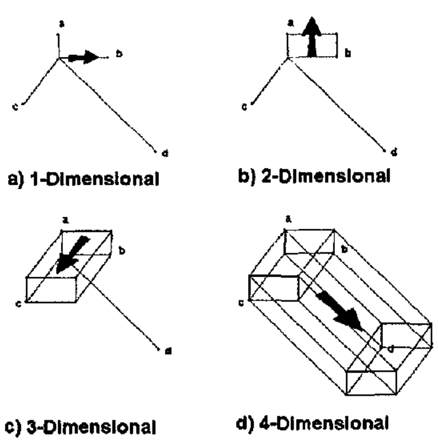

Figure 1. n^{th} dimensional space is simply a *sweep* of the n-1 dimensional space.

The fundamental idea of Star Coordinates is to arrange the coordinate axes on a two-dimensional plane, where the coordinate axes are not necessarily orthogonal to each other (Figure 2). The minimum data value on each dimension is mapped to the origin, and the maximum value is mapped to the other end of the coordinate axis. Thus, unit vectors on each coordinate axis are calculated accordingly to allow scaling of data values to the length of the coordinate axes.

Each point P_j basically represents an n-dimensional data element D_j, where its location $P_j(x,y)$ is determined by the vector sum of all unit vectors $(\vec{u}_{xi},\vec{u}_{yi})$ on each coordinate C_i multiplied by the value of the data element for that coordinate, as shown below:

$$P_j(x,y)=\left(\sum_{i=1}^{n}\vec{u}_{xi}\cdot(d_{ji}-\min_i),\sum_{i=1}^{n}\vec{u}_{yi}\cdot(d_{ji}-\min_i)\right),$$

where,

$$D_j=\left(d_{j0},d_{j1},\ldots,d_{ji},\ldots,d_{jn}\right),\ \vec{u}_i=\frac{\vec{C}_i}{\max_i-\min_i},$$

$$\min_i=\min\{\,d_{ji},0\le j<number\_of\_elements\,\},$$

$$\max_i=\max\{\,d_{ji},0\le j<number\_of\_elements\,\}.$$

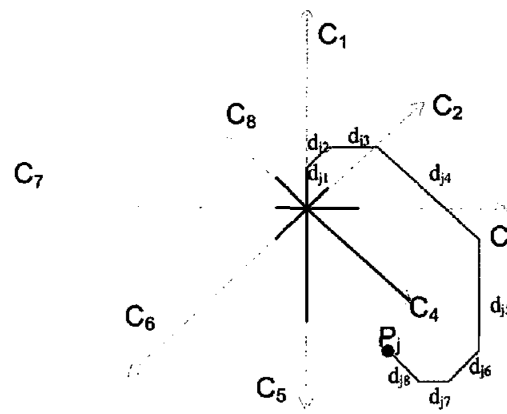

Figure 2. The location of an 8-dimensional point is the vector sum of its value on each dimension.

Clearly, Star Coordinates builds on simple two- and three-dimensional scatter plots but extends them by allowing higher dimensional spaces to nest or overlap through non-orthogonal coordinate axes. Overlapping however introduces ambiguities, as the location of a point in the space no longer uniquely identifies its values on each coordinate axis. Clearly, the location of a data point on a 2d projection of a 3d scatter plot does not uniquely identify its data values either. Only with the aid of interactive dynamic transformations such as rotations and translations can users make sense of data distribution. Star Coordinates basically attempts to extend this idea to higher dimensions.

The approach taken is to provide interactions on the visualization that will help users resolve ambiguities as it is done in most 3-dimensional visualization techniques. Real datasets with large number of data elements also decrease this ambiguity problem by the way data is clustered naturally. Clusters, trends and outliers in the datasets are preserved in the projected multi-dimensional visualization and interactions help to confirm this. In the following section such interaction techniques are described through examples on a sample dataset.

3. INTERACTION TECHNIQUES

Current implementation of the Star Coordinates visualization system provides a number of interaction features, which users can utilize to gain an improved understanding of their datasets.

These features are demonstrated using an example dataset that contains basic auto specs (e.g. mpg, cylinders, weight, acceleration, displacement, origin, horsepower, year, etc.) on approximately 400 car makes manufactured worldwide. Figure 3 shows the initial visualization of this dataset along with the histogram and settings panels. According to this view, distribution of the data points suggest that the data contains a number of clusters, which are likely to be revealed more clearly when transformations and other interaction features are applied, as described below.

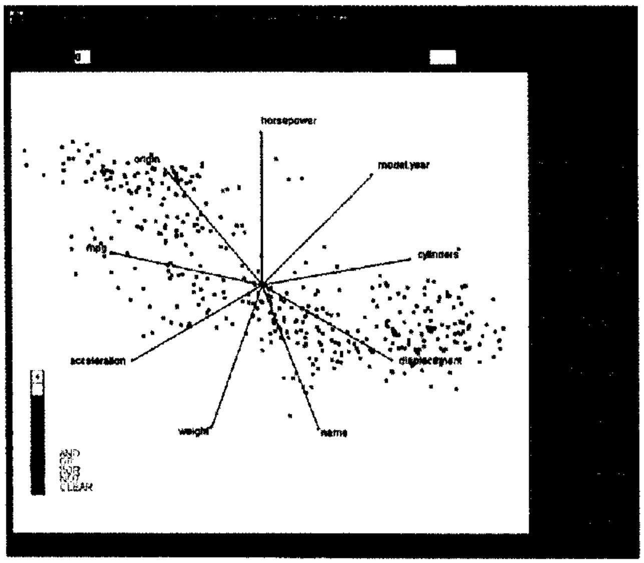

Figure 3. Star Coordinates Interface: a) Visualization of auto specs of about 400 cars b) Histograms and c) Settings panel.

3.1 Scaling

Scaling transformation allows users to change the length of an axis, thus increasing or decreasing the contribution of a particular data attribute on the resultant visualization. In Figure 4, the user scales down the 'name' attribute from the initial layout to examine the distribution of the data points independent of the car producer. As a consequence, five clusters emerge in the new arrangement of the data points.

Scaling transformation is also a natural way to interact with hierarchically expanding and collapsing clusters. In Figure 5, a subsequent scaling of the 'origin' attribute (i.e. continent of the car producer: America, Europe, Asia) from Figure 4 results in merging of all the three clusters (A, B, and C) into one (N). In this visualization points representing cars are distributed solely on the engine specifications independent of the car producer. Thus, it suggests that all cars produced in these two continents, (clusters A and B) and some cars in the other continent (C) are similar with respect to engine specs. However, the remaining continent also produces cars of quite different specs (D and E).

To apply a scaling transformation users pick the end point of an axis and pull or push with respect to the origin. When scaled the data points are remapped according to the new scaling factor based on the length of the axis. Scaling can be also applied on multiple selected axes at once to examine the combined effects of multiple dimensions and the correlations among them.

Alternatively, coordinate axes can be turned off and on effectively setting the scaling factor to 0 and back to its previous value.

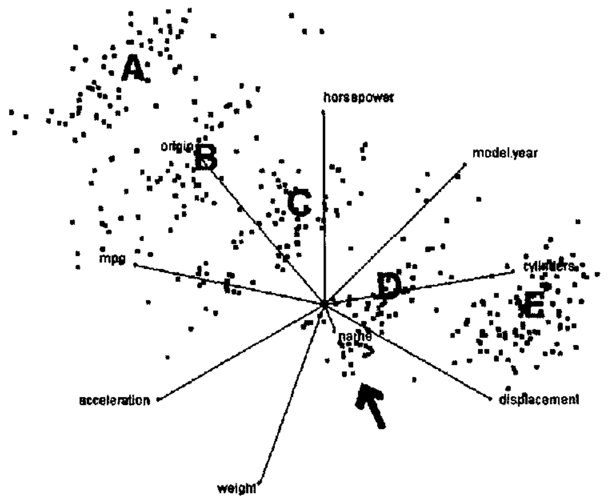

Figure 4. Scaling of the 'name' attribute from the initial layout (Figure 3) reduces the effect of car producers on the distribution of the data points in the visualization, thus revealing five clusters in the dataset (e.g. A, B, C, D, and E).

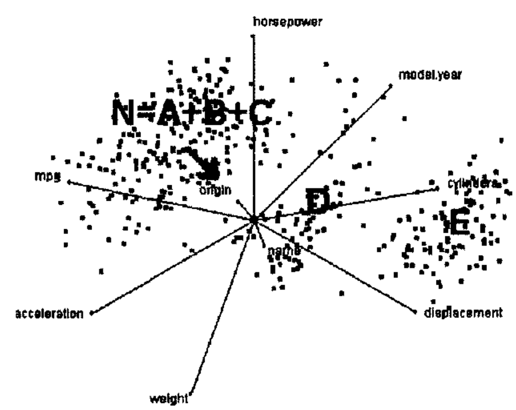

Figure 5. A subsequent scaling of the 'origin' attribute (continent of car producer) reveals the hierarchical clustering in the dataset, where three clusters (A, B, and C) form one larger cluster (N).

3.2 Rotation

Rotation transformation modifies the direction of the unit vector of an axis, thus making a particular data attribute more or less correlated with other attributes. When multiple axes are rotated to point in about the same direction, their contributions are effectively aggregated in the visualization. Similarly, any other axis pointing in the reverse direction has the opposite effect. This way, users can easily perform a multi-factor analysis where attributes representing desirable properties (e.g. 'horsepower', 'acceleration' and 'mpg') point in the same direction and those attributes representing undesirable properties (e.g. weight) point in the reverse direction (Figure 6). As a result, these transformations provide a restructuring of the data points in the visualization based on the criteria chosen.

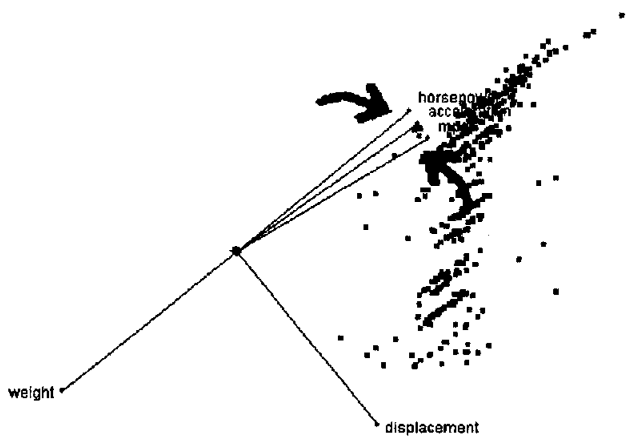

Figure 6. Rotation transformation allows restructuring of the data points based on desirable (e.g. 'horsepower', 'acceleration', and 'mpg') and undesirable factors (e.g. 'weight'). 'displacement' is made perpendicular to the other attributes to examine its effects more explicitly.

In Figure 6, the 'displacement' attribute is rotated to be perpendicular to the other axes. This allows users to examine the tradeoff between 'weight' and other three attributes (e.g. 'horsepower', 'acceleration', and 'mpg') more explicitly as it relates to the 'displacement' of the car.

Rotation also helps considerably in resolving ambiguities in the visualization. It allows users to separate overlapping clusters that seem to be only one cluster.

To rotate, users pick the axis from any point on it and drag to set the new direction to be the vector from the origin to the drag point. Rotation transformation can also be applied on multiple dimensions at once.

3.3 Marking

Users can mark data points by either selecting individual points or by selecting all points within a rectangular area. Data points are then painted with the selected color, thus making them easier to follow in the subsequent transformations. This is useful for examining how clusters are redistributed under new parameters as a result of transformations. When new clusters are observed after applying a number of transformations, having some interesting points marked in advance is likely to help users determine the origin or content of these clusters.

Additionally, when users move the mouse over a data point all attribute values of that particular data element are displayed over the visualization.

3.4 Range Selection

Users can also select data value ranges on one or more axes and mark the corresponding data points in the visualization by the selected color. This operation allows users to understand how particular data value ranges are distributed in the current layout of the data points. Once a range is selected, it can be dynamically expanded or moved. For example, in Figure 7 the user makes an initial range selection on the 'weight' attribute and corresponding data points are marked with the selected green (lighter) color. As the user moves the selection interactively, the visualization is updated dynamically to reflect the updated range. This, in effect allows the user to see contour regions dynamically and observe how data points are distributed as the values of a particular attribute change.

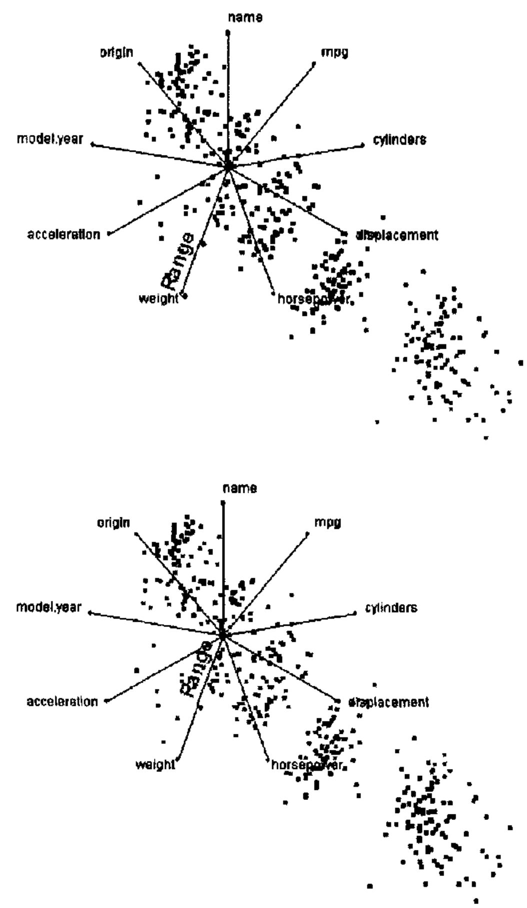

Figure 7. Points corresponding to the selected range are marked with the selected color. Ranges can be moved as well as expanded interactively and the visualization is updated dynamically corresponding to the new range.

When multiple selections (range or rectangle) are made consecutively, users can choose to apply a logical operation such AND, OR, NOT, XOR between the selection sets. These logical operations become especially useful in slicing the data points using specific queries (e.g. mpg > 30 AND horsepower > 130).

3.5 Histogram

Histograms are especially useful when a group of interesting data points is identified and the user is seeking the distinguishing features of this group of points. Histograms provide the data distribution of all points in the dataset and of this group over each individual attribute. Marking individual data points or groups of points by rectangular or range selections update the histogram view by using the selection color to indicate the selected points in the distribution. In Figure 8, the user having identified five interesting clusters (A, B, C, D, and E) selected each individual cluster and marked each with a different color. The histogram view clearly indicates that cars in cluster A and B are of two different origins and clusters C, D, and E have the same origin. Furthermore, cars in cluster A and B have similar engine specifications. Both clusters have cars with high mpg and low weight. On the other hand, cars in cluster E are heaviest among all cars. Cars in C, though from the same origin as cars in cluster E, they too have high mpg and low weight. Cars in cluster D on the other hand have medium range engine specs compared to cars in C and E.

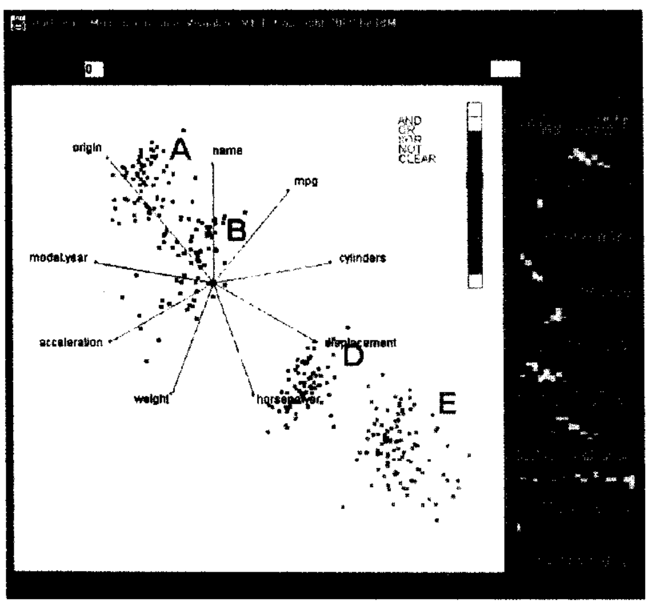

Figure 8. Histograms reveal data distributions of individually marked clusters over a number of dimensions. Once users discover interesting clusters of cars and mark them, histogram views inform the user about the distinguishing origins and engine specifications of these cars.

3.6 Footprints

When users apply transformations such as scaling and rotation they are applied to the visualization dynamically, thus repositioning the data points. However, occasionally it becomes necessary to examine the path data points took to travel when a series of consecutive transformations are applied. Footprints help users in such situations by leaving marks of data points on the trail for the last specified number of transformations when they are being applied continuously.

When used in conjunction with multi-dimensional scaling and rotation, footprints help significantly in determining the correlations between multiple attributes. For example, Figure 9 shows footprints on multi-dimensional scaling of 'mpg' and 'weight' attributes for the three clusters extracted. For cluster A, although variations exist, most of the data points moved along the axis of the 'mpg' attribute as a result of the scaling transformation. Thus, these points represent cars with high 'mpg' and low 'weight'. On the other hand, most of the data points in cluster C moved along the 'weight' attribute indicating that these are rather heavy and low mpg cars. Finally, cars in the middle cluster have a mix of the above specifications.

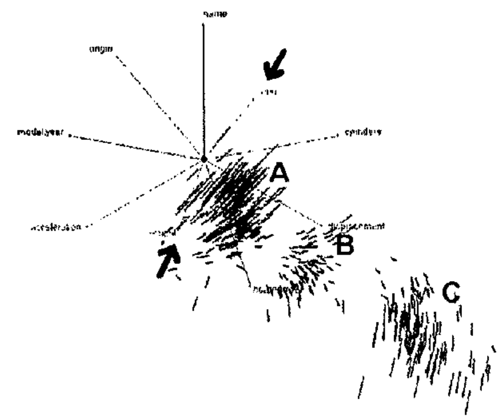

Figure 9. Footprints on multi-dimensional scaling of 'mpg' and 'weight' attributes reveal the correlation between these two attributes for the three clusters in the dataset.

Figure 10 shows how to interpret the footprints on a two-dimensional scaling transformation. Scaling transformation moves points along the vector sum of the axes involved in the transformation. Let say point P^- indicates the point location without the contribution of the dimensions x and y and let P indicate the location when all dimensions contribute. As a result of a two-dimensional scaling point P will move to P', thus the direction of the movement gives an indication of the ratio of the two attribute values and the length of the footprint gives an indication of an aggregate (vector sum) of both attributes.

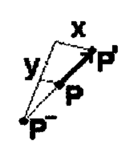

Figure 10. Footprints on a two-dimensional scaling: While the slope of the footprint indicates the ratio between the values of the two attributes (x and y), its length indicates an aggregation (vector sum) of both attributes.

3.7 Sticks

Star Coordinates technique integrates a number of attributes of a data element into a single visual representation, a point. However, there are occasions where an examination of the individual attribute values could be useful. One way to separate a dimension is to use a stick representation. When an attribute is mapped to a stick, the attribute value is mapped to the length of a stick centered at the data point in the direction of the corresponding axis for that attribute. Multiple attributes can be mapped to sticks as well. This way users can examine a number of dimensions at once. In Figure 11, 'cylinders', and 'weight' attributes are mapped to the stick's length in southward and eastward directions, respectively. The location of the sticks is determined in the normal way where every dimension other than 'cylinders' and 'weight' attributes contribute.

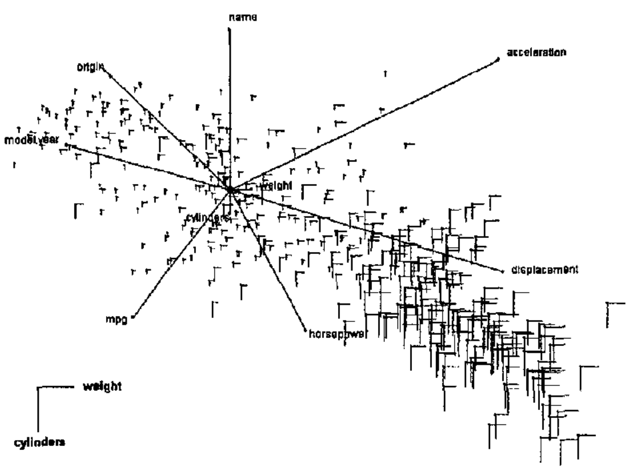

Figure 11. Sticks can encode multiple dimensions on the same data representation. Here, 'cylinders' and 'weight' attributes are mapped to the stick's length in the southward and eastward directions, respectively.

4. CHURN ANALYSIS

It is of primary importance to telecommunications companies to understand why their customers "churn" (i.e. cancel services) in order to identify customers that exhibit similar patterns in advance so that new services, promotions, etc. can be offered to retain their business. Telecommunication companies collect data from various sources such as billing, customer service, call information, etc. to improve their services and to retain and expand their customer base. The 'churn' dataset that is examined here is one such dataset, which contains information such as account length, phone number, total day, evening, night and international minutes, calls, and charges, voice mail and international service plan membership, number of customer service calls, etc.

Figure 12 shows an overview of the dataset, where churned customers are marked with a blue (dark) color and the 'churn' attribute is turned off to remove its contribution to the location of data points. Immediately, there appears to be two dominant clusters of customers. Marking one of the clusters and cross examining with the histograms (not shown in figure) reveals that this clustering is based on the voice mail plan membership and number of voice mail messages. An important question to ask at this point then is whether the customers with voice mail plan churn significantly more compared to those who don't use this service. Scaling both of these axes reveals that this is not the case. In both clusters there appears to be about equal percent of churned customers.

Then, it is worthwhile to examine the significant factors that lead to churn. At this point, it is wise to turn off some of the attributes such as state, area code, and phone number since these attributes are not likely to be of much interest. Also, for the sake of simplicity, let's examine only the charge amounts (i.e. total day, evening, night, and international charges) and turn off total minutes and calls for day, evening, night, and international calls. In fact, an examination by foot-printing a multi-dimensional scaling of total day, minute, and charge attributes reveals that there is no discrepancy between these attributes that leads to churn, as indicated by the footprints that are parallel to each other as in Figure 13. This is also true for evening, night, and international calls. In fact, this clearly shows that the telecommunication company in this dataset charges directly proportional to the duration of the calls made.

Once these attributes are removed data point distribution reveals interesting patterns (Figure 14). First of all, there appears to be four clusters, two of them heavily populated below the other two. This becomes more clear when the international plan attribute is scaled as shown in Figure 15. According to this figure, clusters C and D have a higher proportion of blue (dark) points, indicating that customers with an international plan are more likely to churn. Interestingly however, customers in cluster B (customers with voice plan) are less likely to churn, a situation not observed in C and D. Thus, only for customers without an international plan, voice plan membership makes them less likely to churn. This may indicate a good voice mail plan deal that makes customers more loyal. An examination of the distribution of blue (dark) points in cluster A reveals that the blue (dark) points tend to be denser in the northern and southern parts of the cluster. Rotating and scaling total day charge and number of customer service calls accumulates the blue (dark) points in the southern part of the cluster, thus indicating clearly that these attributes play the most significant role in churn for customers without an international and nor a voice mail plan (Cluster A) as shown in Figure 16.

In conclusion our analysis of the churn dataset using Star Coordinates reveals the following findings:

1. Customers are clustered into two main groups based on voice mail service plan membership, but both groups exhibit similar churn probability.
2. Telecommunication company makes charges that are directly proportional to the call duration, thus there is no discrepancy between number of calls, total charge, and total minutes of a call with regards to churn.
3. Customers with an international plan are significantly more likely to churn.
4. Customers without an international plan but with voice mail plan are less likely to churn compared to those that do not have voice mail plan.
5. Total day charge and number of customer service calls play the most significant role in churn for customers without an international plan and voice mail plan.

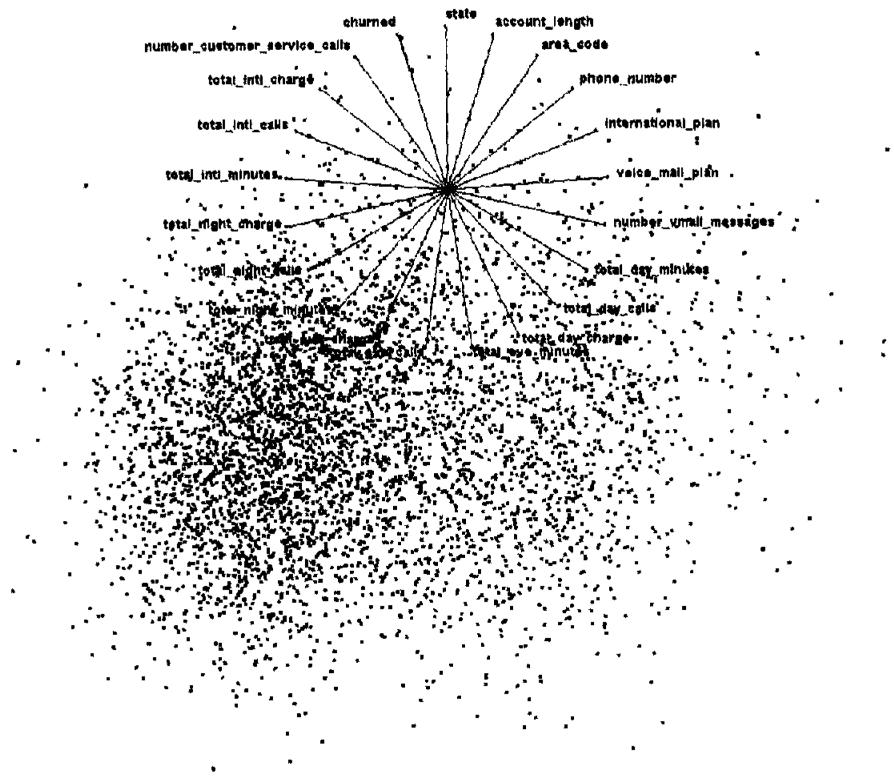

Figure 12. Overview of 'churn' dataset, where churned customers are marked with blue (dark) color.

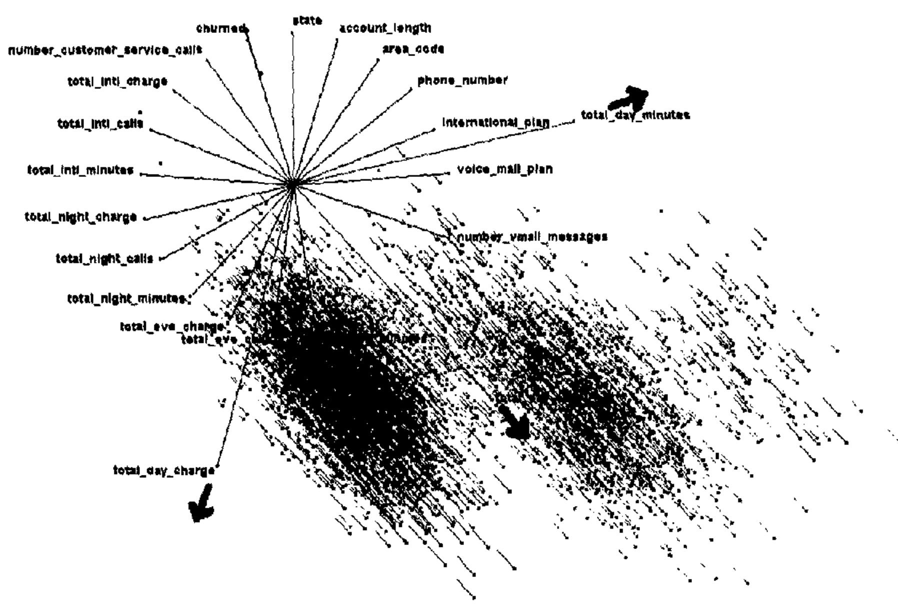

Figure 13. Total day, minute, and charge attributes are in agreement with regards to their effect on churn.

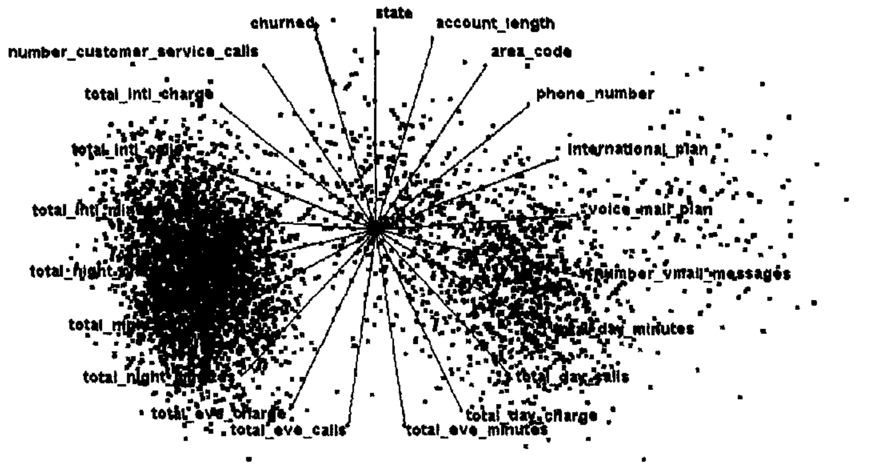

Figure 14. Data point distribution after removing state, area code, phone number, and total minute and calls for day, evening, night, and international calls.

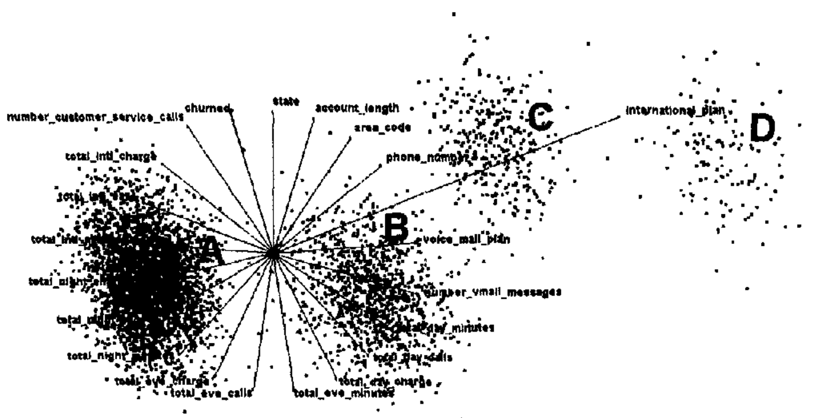

Figure 15. Data points partitioned into four clusters based on international service plan and voice plan membership.

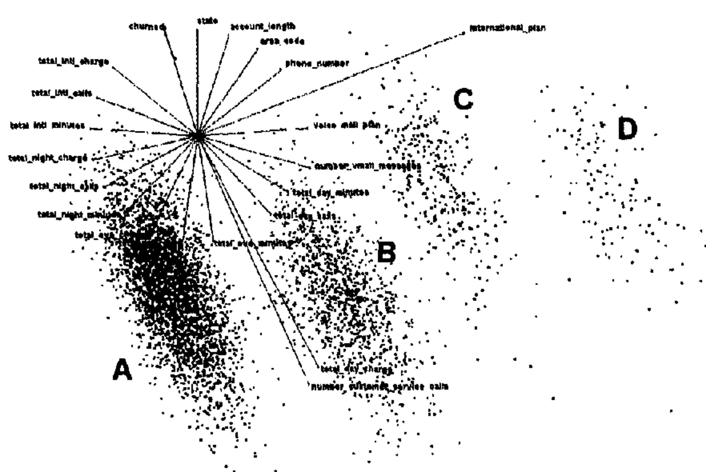

Figure 16. Total day charge and number of customer service calls play the most significant role in churn for customers without international and voice mail service plans.

5. RELATED WORK

Researchers proposed a number of approaches to visualizing multi-dimensional datasets. These approaches differ in their methods to transform and map data into visual representations as well as in the richness of interaction features. While some approaches employ algorithmic techniques to transform/map data, others rely more on the users' interaction to reveal interesting patterns, clusters, trends, etc. Thus, these visualizations differ significantly in their use in the knowledge discovery spectrum from pure exploration to validation.

Among the pioneering work is Bertin's Permutation Matrices [2] in which data are visualized in rows and columns of cells containing simple graphical depictions. By rearranging rows and columns, users try to form clusters –typically on the diagonal of the matrix– to understand distribution of data. Chernoff's [4] use of faces to represent multi-dimensional data is also among the most well known work in the area. Tufte [24] also provides many compelling historical examples of multi-dimensional visualizations, including the famous Jacques Pickard's 6-dimensional visualization of the Napoleon's march to Moscow. While Inselberg's Parallel Coordinates [12,13] uses parallel arrangement of coordinates to accommodate multiple dimensions on a two-dimensional surface, Worlds within Worlds [7] uses a nested coordinate system model to visualize multi-dimensional functions. Star plots [9] arrange coordinates axes on a circle with equally spaced radii extending from the center of a circle and links data points on each axis by lines to form a star. More recent pure visualization work include Table Lens [21], VisDB [15], Dynamic Queries [22], Visage [5] and [1,11,23,25,26]. Self-organizing maps [16,17] on the other hand, utilize automatic arrangement of high-dimensional data using statistical algorithms. Other techniques that couple algorithmic techniques to visualization include [3,8,18,19,20].

It is worthwhile to examine multi-dimensional visualization techniques in more detail since each has different pros and cons, which make them suitable for different kinds of applications and datasets. In order to do an effective comparison however, we need to understand the transformation process that maps data to visual elements in the visualization.

Visual transformation process basically takes in n data values and maps them into k attributes (e.g. location, color, size, etc.) of a visual element, representing a row of the n-dimensional dataset. Effectively, there are k number of n-ary transformation functions. Visualization techniques differ in the number of parameters of the transformation functions as well as in the number of attributes of the visual elements. For example, Parallel Coordinates technique represents a row of data using $n-1$ connected lines passing through n coordinates axes arranged in parallel at the corresponding values for each dimension. Thus, there are n attributes of the visual element representing one row of the data table and each transformation function takes only one parameter. In Star Coordinates, on the other hand, a row of data is represented by a point (i.e. two attributes), but the transformation function takes n parameters. Thus, we can say that while Parallel Coordinates reveals the value at each dimension directly in the visual representation of the data, it lacks to provide an integrated effect, which Star Coordinates provides through a visual representation where each dimension contributes. Besides, a point is a much simpler and space efficient representation compared to a set of connected lines, thus avoiding screen clutter considerably. Besides, the complexity of the visual elements may also have a significant impact on the understandability of a representation. Scatterplot matrices also represent each data element as a point but the transformation function takes only two parameters (i.e. two attributes at a time) thus it too lacks to show an integrated effect. Most three-dimensional visualization techniques however do better in terms of integration since they map three attributes to the position of the visual element and additionally map other attributes to size, color, etc. Thus, a single visual element integrates a number of attributes, however multiple visual encodings (size, location, color, etc.) are used for a visual element, which makes it difficult to interpret appropriately. Thus the uniform encoding (i.e. location) in Star and Parallel Coordinates may be better for cross-comparison and cluster discovery tasks.

6. CONCLUSION

As Grinstein [10] points out visualization techniques harness the perceptual and cognitive capabilities of the users. Star Coordinates, as an exploratory visualization technique, aims to excel users' such capabilities through a representation of the higher dimensional space built on the well-known simple representations and also through dynamic interactions that allow users to discover trends, outliers, and clusters easily.

Current experience with Star Coordinates on a number of datasets indicates that it is a viable approach to gain insight into multi-dimensional data. However, in order to scale this approach to orders of magnitude higher dimensions (>>10) coupling with statistical data mining techniques seems inevitable. Such an alliance could yield interesting features such as animations guided by supervised as well as unsupervised data mining techniques.

In the broader context of knowledge discovery process providing a history of the transformations to the user with the ability to backtrack, or switch to alternative paths is likely to improve the efficiency of the knowledge worker as well as highlight not only the final result obtained but also the whole process, which is equally important.

7. ACKNOWLEDGMENTS

I would like to thank Ismail Haritaoglu and Paul Maglio for their careful review of this paper.

8. REFERENCES

[1] Ankerst, M., Keim, D. A., Kriegel, H.-P., Circle Segments: A Technique for Visually Exploring Large Multidimensional Data Sets. Proc. IEEE Visualization '96, Hot Topics, 1996.

[2] Bertin, J., *Graphics and Graphic Information Processing*, Walter de Gruyer & Co., Berlin, 24-31, 1981.

[3] Buja, A., Swayne, D. F., Littman, M., Dean, N., XGvis: Interactive Data Visualization with

Multidimensional Scaling, to appear in *Journal of Computational and Graphical Statistics*.

[4] Chernoff, H., The Use of Faces to Represent Points in k-Dimensional Space Graphically, *Journal of American Statistical Association*, 68, 361-368.

[5] Derthick, M., Kolojejchick, J., Roth, S. F., An interactive visualization environment for data exploration. Proc. of ACM SIGKDD '97, pp. 2-9, 1997.

[6] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM 39*, 11.

[7] Feiner, S., Beshers, C., Worlds within Worlds: Metaphors for Exploring n-Dimensional Virtual Worlds. Proc. UIST '90, pp. 76-83, 1990.

[8] Feldman, R., Kloesgen, W., Zilberstein, A., Visualization Techniques to Explore Data Mining Results for Document Collections. Proc of ACM SIGKDD '97, pp. 16-23, 1997.

[9] Fienberg, S. E., Graphical methods in statistics., *American Statisticians*, *33*, 165-178, 1979.

[10] Grinstein, G. G., Harnessing the Human in Knowledge Discovery, Proc. of ACM SIGKDD '96, pp. 384-385, 1996.

[11] Johnson, R.R., Visualization of Multi-Dimensional Data with Vector-fusion. Proc of IEEE Visualization '00, pp. 297-302, 2000.

[12] Inselberg, A., Parallel Coordinates: A guide for the Perplexed, Proc. of IEEE Conference on Visualization, Hot Topics, pp. 35-38, 1996.

[13] Inselberg, A., Multidimensional Detective. Proc. of IEEE Information Visualization '97, pp. 100-107, 1997.

[14] Kandogan, E., Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions. Proc. of IEEE Information Visualization, Hot Topics, pp. 4-8, 2000.

[15] Keim, D. A., Kriegel, H.-P., VisDB: Database Explorations Using Multidimensional Visualization. *IEEE Computer Graphics and Applications*, 40-49, 1994.

[16] Kohonen, T., Self-organized formation of topologically correct feature maps. *Biological Cybernetics 43*, 59—69.

[17] Lagus, K., Honkela, T., Kaski, S., Kohonen, T., Self-Organizing Maps of Document Collections: A New Approach to Interactive Exploration. Proc. of ACM SIGKDD '96, pp. 238-243, 1996.

[18] Lee H-Y., Ong, H-L, Quek, L-H., Exploiting Visualization in Knowledge Discovery. Proc. of SIGKDD, pp. 198-203, 1995.

[19] Mihalisin, T., Timlin, J., Schwegler, J., Visualizing Multivariate Functions, Data, and Distributions. *IEEE Computer Graphics and Applications*, 11(13), 28-35, 1991.

[20] Mihalisin, T., Timlin, J., Schwegler, J., Visualization and Analysis of Multi-Variate Data: A Technique for All Fields. Proc. of IEEE Visualization '91, pp. 171-178, 1991.

[21] Rao, R., and Card, S. K., The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus + Context Visualization for Tabular Information. Proc. CHI '94, pp. 318-322, 1994.

[22] Shneiderman, B., Dynamic Queries for Visual Information Seeking. *IEEE Software*, 11(6), 70-77, 1994.

[23] Spoerri, A., InfoCrystal: A Visual Tool for Information Retrieval. Proc. of IEEE Visualization '93, pp. 150-157, 1993.

[24] Tufte, E. R., *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut, 1983.

[25] van Wijk, J. J., van Liere, R. D., Hyperslice. Proc. Visualization '93, pp. 119-125, 1993.

[26] Wright, W., Information Animation Applications in the Capital Markets. Proc. of IEEE Information Visualization '95, pp. 136-137, 1995.