Interactive Visualization of Multivariate Financial Data

Mansha Mahtani

Advisor: Norman I. Badler  
University of Pennsylvania

# Abstract

*Data visualization techniques have been integral to almost all fields of study ever since the need to observe trends and analyze data qualitatively. This project focuses on analyzing multi-dimensional financial data by creating an interactive tool that visualizes information with 3D scatter plots and 2D geometrical methods. Because of the dilemma that exists between displaying succinct summaries of information and providing as many dimensions as possible, it is important to allow for interactivity to control how much of the data should be viewed. For this project, I will also research appropriate UI features to incorporate into the tool and thereby help the user reveal underlying trends in the data.*

*Project Blog: http://mansha.blogspot.com*

# INTRODUCTION

The explosive growth of data and the computerization of financial institutions have created a need for tools that reveal the underlying patterns of a data set. Financial information, for example stock data, is known to be large, complex, and multidimensional and therefore there have been considerable efforts to visualize such data using 3D plots or innovative geometric methods.

3D scatter plots typically include three variables although a few more dimensions are often added with the help of color and intensity. Yet, there is a limit to the amount of information that can be added to a 3D scatter plot while enabling strong comprehension of the data in question. To solve this problem, geometric methods like Parallel Coordinates and Andrew’s Curves can be used. These methods allow for analyzing trends in high dimensional data using 2D graphics. Although these graphs have been researched extensively in isolation, there has been little development on the possibility of combining different visualization techniques in one tool to arrive at a more holistic conclusion about the data set.

The challenge in creating visualizations of high dimensional data is ensuring that the complexity of the data does not reduce our understanding of it. Interactivity can help a user dynamically filter information for his or her needs. However, it is important to research the kind of interactivity that will not only provide the maximum utility but also create the best experience for the user.

This project makes the following contributions:

* Developing of a tool to allow for better comprehension of multivariate financial data
* Designing of UI features specific to financial information to enable the revealing of trends in data sets
* Incorporating 3D scatterplots as well as geometric methods such as, Parallel Coordinates for understanding correlations and Andrew’s curves for detecting outliers

## Design Goals

This project targets professionals in financial institutions that have to deal with analyzing high dimensional stock data. However, this project can be extrapolated for other kinds of multivariate data beyond financial data. In that case, any professional interested in understanding patterns of multivariate data could be part of the target audience as well. The users of this tool will be able to understand the underlying meaning of their data with the specialized interactivity included.

## Projects Proposed Features and Functionality

This project will incorporate features, including but not limited to:

* Loading data from csv files
* Viewing visualization from all angles
* Specialized UI features to filter the data
* Graph displays: volumetric scatter plots, parallel coordinate graphs, Andrew’s curves

# RELATED WORK

## Cartesian Scatter Plots

In a scatterplot, the frame for the data presentation is a Cartesian coordinate system in which the axes correspond to the three dimensions [Bar\*05]. Although, there are not many examples of three-dimensional stock data, one can hypothesis the possible variables on the three axes – from date to opening price to closing price. Because simply plotting points is not conducive to observing trends, a surface or volume can be created to connect the points. To add an additional dimension or variable, heat maps are often used in 3D graphs as different colors represent different values of a particular 4th dimension.

## Geometrical methods

Geometrical methods encompass a large range of techniques to exhibit high dimensional data. For this project I will be observing Parallel Coordinates and Andrew’s Curves.

### Parallel Coordinates

Parallel Coordinates was conceptualized by Alfred Inselberg and is based on the idea that dimensions are depicted as parallel axes. The axes are linearly spaced and each point is graphed as a polyline intersecting all the parallel dimensions at the coordinates of the point [Kum\*07]. For example, in the below graph the variables are date, opening price, high price, low price, closing price, and trade volume. The general trend of the lines can establish correlations and reveal trends that are difficult to observe with fewer than four dimensions.

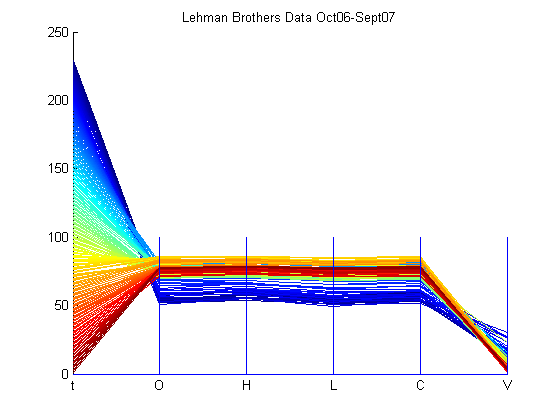


Figure 1: *Stock Data in Parallel Coordinates [Kum\*07]*

The idea of clustering is often used to show relationships between different groups of data in parallel coordinates. In the below graph, we can observe clusters of four different time periods. For example, the worst episode of subprime mortgage crisis hit the market around mid-July 2007 to early-September 2007 and one can notice that the cyan cluster exemplifies the impact of this crisis on the stock value of Lehman Brothers. The stock has lost value than what it was in October 2006, which is marked by the blue cluster. However, one can notice that the cyan cluster has exceptional volume in comparison to other clusters. This can be viewed as traders bargain hunting and loading up on the Lehman Brothers’ stock for future appreciation of their money. The red cluster signifies the first show of the subprime mortgage crisis’ head in February 2007, and it is clear that the stock value of Lehman Brother’s decreased. Then, the stock value rose a bit as the yellow cluster conveys before the crisis stormed through the market once again in mid-July 2007. Overall, clustering can be very useful in finding structures in datasets and explaining certain aspects of the dataset [Kum\*07].

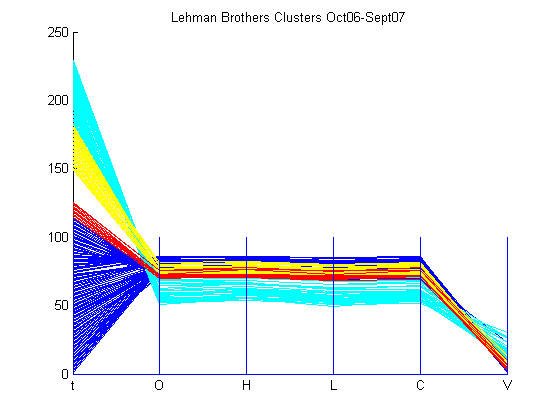
****

Figure 2: *Clustered Stock Data in Parallel Coordinates*

### Andrew’s Curves

This method plots each data point as a function of the data values using a speciﬁc equation. The data point curves are usually sketched in the interval −π < t < π.The function which draws these curves is shown as:

where and are the values of the data points for the particular dimension.

Andrew’s Curves are notable for detecting outliers in data. Their disadvantage is the long computational time as every data point requires the calculation of a trigonometric function [Bar\*05].

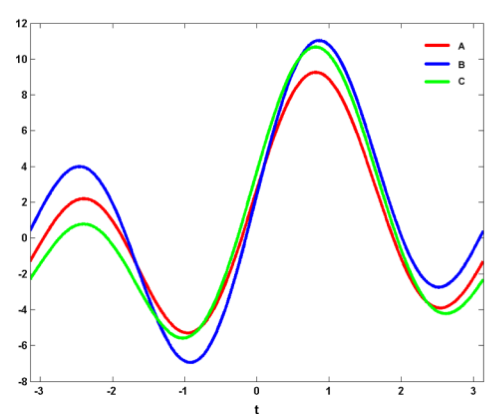


Figure 3: *Andrew’s curves with 3 simple coordinates: A = (1; 3; 2; 5), B = (2; 4; 1; 6) and C = (1; 4; 3; 5) [Bar\*05]*

# PROJECT PROPOSAL

The main focus of this project is to create a tool that depicts financial stock data in ways that reveal the underlying pattern between the variables. As humans we are limited by the ability to only view up to three dimensions. I would like to use this project as an opportunity to create a data-mining tool that is a solution for this limitation.

The project will require pulling data from CSV files and using the multi dimensional values to plot points on the interface. The 3D scatter plot will involve computer graphics class concepts like camera views and ray interactions. I plan to customize the way geometric graphs are generated by paying close attention to concepts like clustering that can abstract complex data. As the tool is essentially a user interface, I will consider UX techniques in helping the user filter data as required.

## Anticipated Approach

I plan to first study financial stock data and research common features that professionals find useful. In this stage I would like to familiarize myself with the necessary concepts that make visualizations more robust and comprehensive. After this background reading, I will pull data from CSV files and generate each of the visualizations. Initially my program may deal with only specific variables like date, opening price, closing price, high price, low price, and volume. However, once I can generate good visualizations I will extrapolate and use other variables. The visualizations will use concepts such as camera views, geometric transformations, interpolation, and voxels. Once I am able to get all the data on the interface I will work on the user experience and the interactivity. This will include the ability to filter values of specific variables.

## Target Platforms

I will be programming primarily in C++ in Visual Studio with OpenGL and QtDesigner.

## Evaluation Criteria

I will evaluate my project by comparing the interpretations I make using my software with interpretations made on isolated visualizations. If my visualizations can abstract more underlying information I will know that my software is more useful to analyze.

# RESEARCH TIMELINE

**Project Milestone Report (Alpha Version)**

* Completed all background reading
* Determine method of pulling data into OpenGL
* Preliminary visualization of specific points
* Collected all necessary data and images

**Project Final Deliverables**

List what you will deliver at the end of the semester

* Software with the visualizations finalized
* Interactivity in the interface
* Documentation

**Project Future Tasks**

* Use real time data
* Interactivity with motion capture

# Method

# RESULTS

# CONCLUSIONS and FUTURE WORK

**APPENDIX**

1. **Optional Appendix**

Some text here. Some text here. Some text here.

# References

|  |  |
| --- | --- |
| [Bar\*05] | BARTKE, K.: 2D, 3D and High-Dimensional Data and Information Visualization. *Seminar on Data and Information management*, 2005. |
| [Kum\*07] | KUMAR, A.: Parallel Coordinates, http://people.ece.cornell.edu, 2007. |
| [Lux\*] | LUX, M.: Visualization of Financial Information, *Fraunhofer-Institute for Computer Graphics* |
| [Rub\*03] | RUBY, M.K.: Information Visualization for Financial Analysis. *University of Durham*, September 2003. |
| [BP\*] | BIN L.S., PRAKASH E.C.: Hierarchal Volume Visualization for Financial Data. sydney.edu.au/engineering/it/~vip2000 /poster/bin\_vis.doc. *University of Sydney.* |
| [Lou90] | INSELBERG, A.: Parallel Coordinates: Visual Multidimensional Geometry and its Applications. *Springer,* 2009. |

|  |
| --- |
|  |

**Figure 2:** *For publications with color tables and figures that span two columns like your gant chart or results will.*