Christo Allegra

3D Data Visualization / Casey Reas

Large scale data visualization offers the ability to see many data points at once. By providing more of the raw data for the viewer to consume, visualization hopes to surface trends and enable qualitative analysis. This visualization research focuses on financial data, specifically the daily trading data for all of the companies in the S&P500 for the year 2008. The specific daily data points are company name, open and close price and trading volume. The daily open and close data is used to calculate day change, year-to-date change, and new year-to-date highs and lows. Trading volume shows how the companies shares are moving the stock market. Additionally, a subset of companies is used to demonstrate the ability to highlight user specific interests.

Target Audience

The design tension between displaying succinct summaries or massive blasts of data for users in visualization is best mitigated by the target audience. In the case of this visualization, the target audience is financial professionals who are familiar with the flashing terminal displays that are currently the standard of trading floors and professional financial information systems. In addition this visualization also targets younger users who have exposure to video games and other time-based interactive mediums where 3D space and abstract representations of environments are expected of a data display.

Platform

The technology solution for this project was a local MySQL database connected to Processing which rendered 3D objects using OpenGL. The initial application design relied heavily on loading all data upfront but as the visualization changed from one which displayed all data at once to a 4d approach using time, requests were restructured to be dynamic to the database. This greatly diminished the upfront load time for the tradeoff of a 200ms response time.

Representational Forms

Visualization of abstract data requires an analysis of the appropriate form. While most visualization methods can encompass any data set, the decision to choose a tree-map, pie chart, or time series histogram can significantly impact your visual results, privileging certain data interpretations over others or worse obscuring important aspects of the information.

Three different forms were explored in order to achieve the final results. All three forms used X, Y, Z, color, and size to visually model the data on screen. The final result (4D) applied time to the display and broke away from the design goal of attempting to show all performance at once. By synchronizing and animating time along the time axis, the daily changes and their shifts became more pronounced and helped to amplify the overall effect of the 3D visualization.

2D and 3D oftentimes are put at odds in visualization with 2D winning in an evaluation of merits as it can simplify the data display. However, when adding a time component to X, Y, and Z the relationship of objects in the 3D space aids comprehension via the motion over time, enhancing the simulation and amplifying data comprehension. Oftentimes the solution to information overload is simplification, but in this research additional harmonious complexity clarifies the data instead of obfuscating it.

An underlying design rule of this research was to show massive amounts of data for pattern recognition. But when dealing with large data sets, the number of axes you can display at one time reduces the amount of variables that can be displayed without confusion. Relying on time as an axis to animate data, allowed data behaviors to emerge through repetition while multivariate data enhanced the point in time information collection via multiple active dimensions, color, X, Y, Z, relative size, and other highlighting effects.

Topographical

In the "Topographical" display, the companies ran alphabetically down the Y-axis against each days performance on the X-axis. This design creates a large landscape that allows for trend tracking in two dimensions. Company volume on the a Z-axis allows the user to focus on where the daily bulk of trading is happening. The concept of thresholding becomes valuable for focusing the user's attention. With so much data to be shown at a time (3 data points x 264 trading days x 470 companies = 372,240 data points), managing data density is crucial and overlays are necessary for finding what matters. The threshold that was used in this display was percent change and the user could control the variable's range. When this overlay is turned on, the company name appears next to its visual representation, allowing users to quickly orient and scan across the highlighted companies for reference.

The overall volume of data was too high for comprehension in this display. Additionally, the large landscape was cumbersome to navigate. Design directions that were distilled from this exercise were to group companies into their relevant sectors or industries in order to breakdown the density of information. A additional interaction concept was annotate with a cross highlight that would allow users to view either a single day for all companies and/or a single company for all days.

The target audience review found this display valuable in that it focused the user on the high volume stocks that affect the market and the ability to scan large volumes of data over time to see granular trend data. A smaller data-set was suggested of the Dow 30 and the use of market capitalization on the Z-axis.

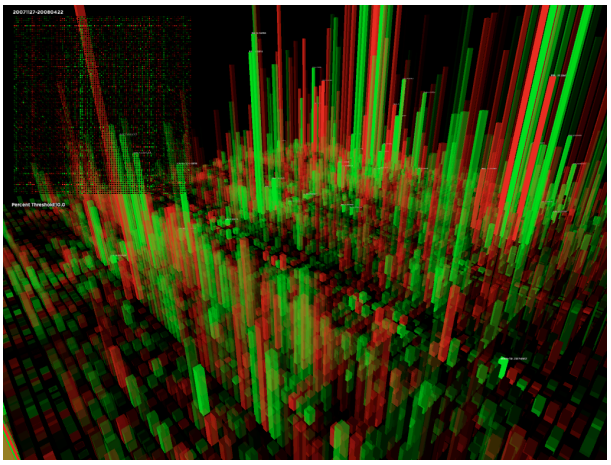

Fig 1. View of Landscape

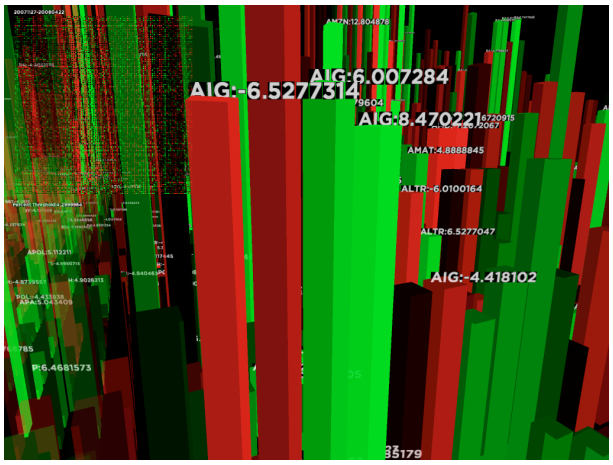

Fig 2. Dynamic Labeling

Dimensional

Financial data is oftentimes categorized and traversed in a tree structure with the S&P500 breaking down into sectors which further categorize into industries and subIndustries. Lisa Strausfeld's "Financial Explorer" offers a design construct that allows for traversing planes of data while allowing users to pivot tables into view for comparison in a pure representation of the database as a floating dimensional space. “Dimensional” takes cues from “Financial Explorer” and stacks time series data on the z-axis, with the most recent date closest to the user. Sectors group companies which are listed on the y-axis alphabetically with their respective performance. Annotation pulls out the year highs and lows for companies allowing users to see contextual data as they "swim" through this display. One design discovery on this display was the use of color shifting to highlight outliers. Generally in information visualization saturation and brightness are the primary color cues for range-finding. In this display, colors shift towards yellow or purple from the base orange and blue in order to further annotate focus points for the user.

The primary positive of this display is the ability to create a feeling of immersion. Once the user zooms in, they are in the midst of a data-scape to explore by spinning and moving. However, there is also a strong sense of disorientation and the overall amount of data comprehension is very low. Tracking

trends this way becomes difficult and other overlay methods are needed as scanning in three dimensions in this display is ineffective. Looking further at "Financial Explorer" and the controls that allow table intersections and analysis could enhance the possibility of this display. This display also suffers from heavy data overload.

The target audience found this display overwhelming and disorienting and had difficulty discerning trends.

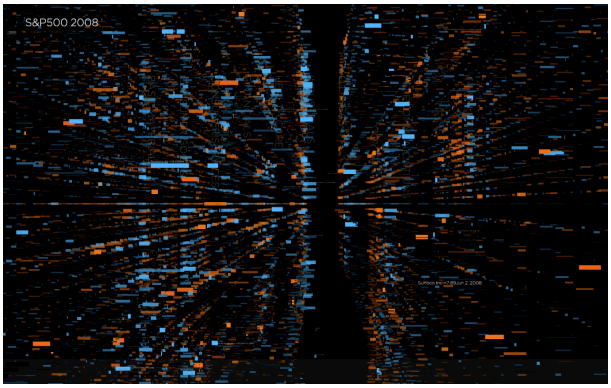

Fig 1. Sector Groupings

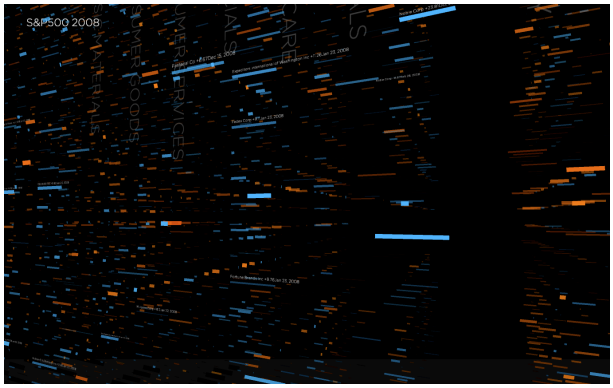

Fig 2. Traversing Dataspace

4D

3D taps into our visual sense of perspective, scale, and depth. These visual skills become more active when there is a relative context for dimensional attributes. Motion will establish an objects relative context. When an object is moving it provides additional distance, size, scale data which defines its physical relationships to its environment and the user. In the "4D" system, time is animated and each day is show in succession so that the visualization simulates the rhythmic change that the market produces with its expansion and contraction. Typography plays an important role in this display. As companies reach user-defined thresholds, type appears, locating the company and providing both ticker symbol and data point. The Y-axis is year-to-date percent change, the X-axis is daily change, and the Z-axis is volume. The size of each cube represents volume and its color represents a combination of both year-to-date change and day change. Color hues for day positive + year-to-date positive shift from green to blue, while day positive + year-to-date negative shifts from red to yellow. Companies that reach new year-to-date highs and lows glow in the appropriate color (green or red respectively.) Additionally, by moving the mouse the axis can shift so that one of the dimensions is flattened allowing the user to focus on the movement of one variable. Every interval (default:2 seconds) another 2000 data points are presented.

Time is a critical factor in making this display work. By using motion to describe the trend display, it becomes easier to understand the rhythmics aspects of market dynamics. The markets move. The back and forth see-saw of an emotional market that goes up one day and down the next becomes apparent quickly when watching "4D." Tracking companies and their aggregate movement allows for a more complex comprehension of overall market shifts. Additionally the interweaving of visual attributes to multiply the effect of a data value (color and X and Y-position represent change, scale and Z-position represent volume) allows for reinforcement of meaning. Additional comprehension cues like sounds could create further depth in order to translate more data to the user and build anticipation around the behavior of the markets.

In looking at this display, the target audience found that the motion helped to describe the behavior in a more visceral way even commenting on the vertiginous effect of the animation when the heavy swings of the October/November crash were played. The primary suggestions were to use streaming intra-day tick data to model the immediate market behaviors and add sector averages into the display with connecting lines to the companies that comprise them. This latter suggestion would show how the aggregate behavior was affected by its individual constituencies.

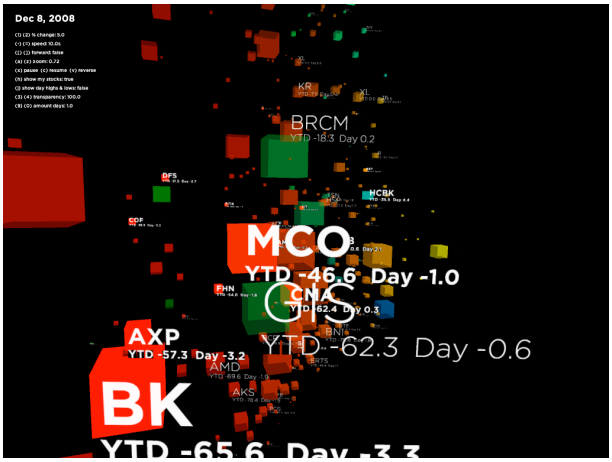

Fig 1. Full View

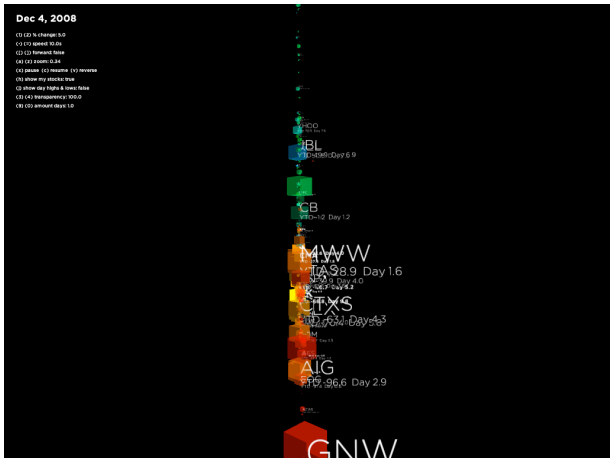

Fig 2. Side View

Conclusion

3D visualization of market data becomes more valuable when modeled in time. The visualization begins to simulate the behaviors it describes which aids in comprehension of the systems overall dynamics. Two of the displays shown have treated the data as a solid mass that is navigated, analyzed, zoomed and rotated, while this offers the ability to have more "surface area" for drawing vs. a 2D plane, it does not add significant comprehension value for this dataset. The form does not reinforce the shape. Over time, the shape of the data begins to emerge and it establishes a baseline for expected behaviors.

In 3D there are different rules than 2D. 2D information visualization oftentimes associates each attribute (color, position, size) with unique data-points. Blending attributes reinforces the behaviors of the 3D visualization and aids in comprehension and trend analysis, specifically when high volumes of data are being shown. This visual cueing helps to build anticipation and expectations around "what the data will do next" which helps to identify the behavioral shifts of the data.

Looking at current trading desktops where motion is confined to blinking points on the screen these new 3D displays start to condense the activity that is being cognitively assembled by traders instantly. The opportunity to replay, simulate, annotate and model behaviors can move comprehension into a visual space that is reactive to data, interactive to the trader's needs, and thick with the overall volume of information.