

Design Study to Visualize Stock Market Bubble Formations and Bursts

Sruthi G. Iyer

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science and Applications

Yong Cao, Chair
Christopher L. North
Steven D. Sheetz
Dane Webster

May 1, 2014
Blacksburg, Virginia

Keywords: Visualization, Stock Market, Multiple Coordinated Views
Copyright 2014, Sruthi G. Iyer

Design Study to Visualize Stock Market Bubble Formations and Bursts

Sruthi G. Iyer

(ABSTRACT)

The stock market is a very complex and continuously changing environment in which many varying factors shape its growth and decline. Studying interesting trends and analyzing the intricate movements of the market, while ignoring distracting and uninteresting patterns has the potential to save large amounts of money for individuals as well as corporations and governments. This thesis describes research that was conducted with the goal to visualize stock market data in such a way that it is able to show how behavior and movement of various market entities affects the condition of the market as a whole. Different visualizations have been proposed, some that improve on existing traditional methods used by the Finance community and others that are novel in their layout and representation of data and interactions. The proposed design, by the use of interactive multiple coordinated views showing overviews and details of the stock market data using animated bubble charts and statistics, aims to enable the user to visualize market conditions that lead to the formation of a bubble in the market, how they lead to a crash and how the market corrects itself after such a crash.

Dedication

This thesis is dedicated to my parents, who have been my biggest inspiration all these years and never failed to give me confidence and encouragement when I needed it the most. Their undying support and confidence in me have kept me strong and constantly enthused throughout my studies.

Acknowledgments

First of all I would like to thank Dr. Yong Cao for giving me the opportunity to work on this interdisciplinary project with the Pamplin College of Business at Virginia Tech. His constant motivation and guidance were invaluable to me throughout my graduate studies at Virginia Tech. I would also like to thank Dr. Chris North whose timely and knowledgeable comments have given concrete direction to this research. A constant source of support and expert knowledge has been Dr. Steven Sheetz who has been extremely patient with me. His sheer enthusiasm in this project, knowledge and motivation have pushed it to become what it is today. I would also like to extend my heartfelt gratitude to Dr. Raman Kumar who so kindly gave us his time, effort and whose expertise has led this project to become much more than what it initially set out to be. Dane Webster has helped me gain and maintain confidence to code and express my own creativity in this project. He has been my strongest support through this research and has made it such an enjoyable experience.

I would also like to thank my colleague and project team mate Joseph Rakestraw for his constant support and encouragement throughout this research. I also need to extend gratitude to my peers Avik Sengupta, Junpeng Wang, Peng Mi and Run Yu who have all helped me throughout my Masters program at Virginia Tech. They helped take the pressure off and made my studies a lot less stressful. I wish them all the very best for their doctoral studies and future careers.

I cannot thank enough, my father Mr. Ganesan, my mother Mrs. Dheepa Ganesan and my brother Shreyas Iyer for their undying love and encouragement throughout my graduate studies so far away from home. This project is dedicated to them. Last but not the least, I thank the Almighty for keeping me healthy and providing me with the best opportunities at every stage of life.

Contents

List of Figures	vii
List of Tables	ix
1 Introduction	1
1.1 Motivation	2
1.2 Research Contributions	3
1.2.1 Contributions in Information Visualization	3
1.2.2 Contributions to Financial market data analysis	4
1.3 Thesis Organization	4
2 Literature Review	5
2.1 Popularly Used Visualizations	5
2.2 Time Series Data Visualizations	8
2.2.1 Animations	8
2.2.2 Multiple Views and Details	9
3 Proposed Design	11
3.1 Bubble View	11
3.1.1 Attribute-Channel Mapping	11
3.1.2 Circle Packed Design and Interactions	13
3.1.3 Formative Study	16
3.2 Statistics View	20

4	Case Studies	24
4.1	Using Different Performance Measures	24
4.2	Details on Demand Level	26
4.3	Using Multiple Views	27
5	Pilot Test	33
5.1	Design of the Test	33
5.2	Hypothesis of the Study	35
5.3	Participant Task Results	36
5.4	Participant Opinions on Visualizations	40
6	Conclusion	41
6.1	Intellectual Merits	41
6.2	Broader Impact	41
6.3	Future Work	42
	Bibliography	43
A	Formative Study	47
B	Pilot Test	63
C	Design Iterations	79

List of Figures

2.1	Map of the Market showing (1) How a details level would only show that sector without its context in the overview of the market and (2) How some rectangles are very small and distorted	6
2.2	Yahoo Finance chart for the HP stock, the website found on http://finance.yahoo.com/ , 2014, Used under fair use, 2014.	7
3.1	The sector 'Computers' shows sudden drop in market capitalization value after a market crash. Other sectors rush in to fill the empty space left by this sudden shrinking.	14
3.2	User Study Statistics (a) Preference of users (b) Distance from correct answer (c) Answers regarding most useful feature	18
3.3	Details view for Telecommunications and Petroleum sectors before and after the Dot Com crash. Notice the number of IPOs has fallen drastically for the bubble sector Telecommunications and has remained almost the same for the non-bubbling sector Petroleum	21
3.4	Line Charts showing 48 sectors of the Market (a) Performance Chart (b) Market Capitalization Chart	22
4.1	GMR Mode: Effect on colors by Selecting different start dates	25
4.2	100Percent Mode: Flipping of colors trend after burst occurs	26
4.3	Enterprise view where larger sector is "Communications" and smaller sector is "Hotels, Rooming, Houses and Camps". (a) Geometric Mean Return (b) 100 Percent.	28
4.4	Histogram and Bubble Chart for the market, before and after the Dot Com crash of 2000	29
4.5	Line chart and Bubble chart highlighting the Petroleum sector for March 2001, which is one year after the dot com crash	30

4.6	This shows the Median and Strike Chart for a date one month before the crash. The strikes have been sorted and sectors to the far right show those that have contained bubbles for long periods.	31
4.7	The 1998 Market correction occurred in August 1998 because of finance based sectors like Trading, Banking and Insurance	32
5.1	Line Charts showing 48 sectors of the Market with Insurance and Trading sectors highlighted (a) Performance Chart (b) Market Capitalization Chart	34
5.2	Line Charts showing the performance of Fabricated Goods at 07/2000	39
C.1	Using circle packed bubbles to show stocks within sectors of the market	79
C.2	Overview with few sectors and details selected	80
C.3	Details view with circle packed sectors and circle packed stocks within sectors . . .	80

List of Tables

3.1	Participant responses to visualization statements	17
3.2	Regression Estimations	19
5.1	Participant Tasks for Each Visualization	36
5.2	Time to Perform Tasks (in seconds)	37
5.3	Average of Responses	38

Chapter 1

Introduction

Analyzing stock market data through visualizations has recently become popular in the field of finance where everyone from the personal investor to a market regulator prefers visualizations to reading stock prices off a market ticker for making informed decisions regarding investments and the state of the market. So visualizations can be used for anything from monitoring stocks that one has a vested interest in to monitoring sector or market wide trends which may hint at drastic events like bubbles or crashes. Visualizations for these different tasks have to be specifically tailored so that user can answer his question in the easiest and fastest way possible. We propose a design that addresses the latter issue of visualizing the stock market in such a way that market wide trends can be perceived and market bubbles and crashes can be visualized with minimum cognitive effort by the user.

Examples of some popular visualizations which are frequently accessed by finance professionals include Map of the Market [37] and Yahoo's Finance Charts [1]. The former is a treemap visualization [18] [29] which shows the hierarchical structure of the market, as sectors and stocks through their market capitalization values and the latter is a line chart whose time axis can be adjusted to show data for the required time range. Both visualizations are generally used by individuals to assess the state of the market with the primary aim of making investment decisions based on judging how a particular stock is going to perform in the immediate future or for the monitoring of daily performance of a stock that has already been invested in. Map of the Market only provides the user with the current state of the market while Yahoo Finance provides historical data for the individual stock of interest, without any representation of the state of the market as a whole. While the state of the entire market may not be of major concern to the average personal investor, an area which is of great interest to market regulators and professional portfolio managers is the analysis of market conditions and factors that lead to the formation of bubbles and crashes. Factors that are indicative of market crashes are still being studied in the finance community in which direction focuses on market capitalizations of sectors, their abnormal performance over specific time periods and how they affect the market as a whole. Thus the need is for a visualization that represents not only the market capitalizations of stocks and sectors, but also a measure of their performances and their

effect on the entire market, along with indications of abnormal conditions and exceptional performances which are strong indications of a potential market bubble or crash. We aim to overcome the shortcomings of the previous approaches using our new design while taking advantage of their fortes to strengthen the visualization and reduce cognitive strain for the user during analysis. The next few sections explain how this is achieved in greater detail.

1.1 Motivation

In order to better understand the needs for such a visualization, it is essential to identify certain key aspects of the stock market so that specific design decisions may be taken to tailor the visualization to achieve the best results. Bubbles in the stock market occur when stock prices rise at a very fast pace and begin exceeding by very large amounts, the true value of the stocks for the firms they represent. Stock bubbles can occur in the market in varying degrees ranging from being localized to one particular stock to being spread over an entire sector of the global market. When the market has reached a condition where a bubble has grown to a large extent, the two possible outcomes are either the gradual deflation of the market bubble or the more drastic bursting of the market bubble, which may have consequences depending on how large the bubble got during its peak. Once bubbles have reached a large size, they are generally given additional momentum as their popularity among investors increase. Due to such a perception of high gain that is given by the price of the stock, investors begin to neglect the stock's fundamental value and continue to invest in large quantities, adding to the bubble in the market. However, once the stock fails to provide the expected returns, or investors begin to suspect that their return expectations will not be met as desired, they begin to sell the stocks. If the selling takes place around the same time, the market experiences a bubble burst. Another reason for bubble growth momentum [2] is when some investors realize the presence of a bubble and the imminent danger it poses but still decide to invest in it anyway, while planning to sell the stock at its peak value just before the burst occurs and investment is lost. Such an activity is called riding-the-bubble and it adds to the bubble's rapid growth. Due to this, bubble peaks often witness record breaking prices, and this causes further disastrous repercussions for the market when the bubble eventually bursts.

Market regulators are interested in studying this phenomena by analyzing past manifestations of such bubbles. This kind of study can lead to insights regarding reasons for their formation such as trends that lead up to a sudden burst or casual deflation and flow of investments among different sectors of the market before, during and after the bubble. Since stock market data contains various attributes that have complex dependencies, studying such data manually by analyzing values of stocks may be a tedious operation. Also, a study of market-wide bubbles is possible only by comparing values over a period of time for different stocks so that market wide trends will be observable because bubbles do not form overnight. A visualization that is interactive in nature and that allows the regulator to visualize the state of the stock market rather than reading actual data values can be a more intuitive and faster way of observing market trends. It will also be of help not just to the regulator but also the personal investor who can make broader decisions regarding

the prudence of investing anywhere in the market during its present condition. We present a design for visualizing stock market dynamics of a bubble formation and burst through the interactive animation of market capitalization values and performances of market entities, with an overview first and details-on-demand approach. Our design makes use of multiple coordinated views where one view shows an interactive animated bubble chart for revealing trends of market capitalization values and performances and another view shows statistics of the bubble chart using a histogram and a line chart to show historic values. A third view is like a cross section of the first two views which reveals the comparative performance of the entities while giving an indication of the severity of the bubble formation. The bubble chart is the main visualization and while the other two are used to support the initial hypothesis derived from it. The bubble chart shows market entities which are market sectors in the overview and individual stocks of selected sectors in the details view. Following sections further expound on our design choices and explain how they are justified.

1.2 Research Contributions

This research was done in collaboration with the Pamplin College of Business. Motivations behind this research keeps evolving because of the exploratory nature of the visualization of the stock market data. Novel contributions therefore are beneficial to both fields; information visualization as well as financial market data analysis.

1.2.1 Contributions in Information Visualization

Among the different coordinated views, the line charts show the historical view of the data which facilitates tasks that involve identifying a time from the data set.

The main view is the bubble and strike chart which uses a bubble chart to depict the 2 attributes of stock market data, the market capitalization value of the market entity and its performance.

Level of Details are provided to the user for studying financial market data. The first level shows the overview of the market data which depicts the sectors of the market which are generally the cause for market wide bubble formations and bursts. The second view which shows details-on-demand show the individual stocks of the sectors which have been selected from the overview. Since there may be thousands of stocks within a sector, the details view enables a more focused view of the data without causing cognitive overload while still using the same metaphor as the overview level.

Context is also preserved during the focused detailed view, by enabling the user to see the state of the sector in the whole market while viewing its individual stocks. On the other hand when the user is looking at the overview, there is a clear view of what the entire market is performing like at the moment. This way context is maintained throughout the visualization.

Use of multiple coordinated views, where each one shows a part of the same data. 2 views show the same attributes of the data but for different times and with different marks. The third view does data processing to show a summary like view of the data which is beneficial to specific tasks.

1.2.2 Contributions to Financial market data analysis

The use of the bubble chart allows market data analysts to select and observe specific sectors and stocks individually and study their trends over time using the multiple views

The use of the same color scheme as market data analysts are accustomed to while they view a new type of visualization in the bubble chart, allows them to visualize data that fits their cognitive mental model.

The visualization facilitates tasks that involve identification of stocks, location of a specific time period of interest in the data, correlation among different market entities of the same level and that between the market entity and the whole market as well as prediction based tasks.

Finally the visualization also facilitates the development of different type of trading strategies in the market by allowing the user to identify periods that are the best time to buy and sell stocks in specific sectors.

1.3 Thesis Organization

Literature review about recent research done in the field of time series visualizations and financial data visualization is presented in Chapter 2 where the existing approaches in the two most popular visualizations of the stock market and their shortcomings when it comes to market-wide bubble analysis, and give a brief overview of other types of stock market data visualizations are described. The new design with detailed justifications for design decisions are explained in Chapter 3 along with a formative study with finance experts. Case studies are detailed where our design was put to use and interesting insights were obtained by market professionals, verifying the success of the design and giving it credibility for use in market analysis. Finally the conclusion is presented along with directions for future work in this application domain.

Chapter 2

Literature Review

2.1 Popularly Used Visualizations

TreeMaps [29] use a space-filling approach that divides the available space into rectangles that represent the hierarchical structure of the market data being visualized. This approach is extremely useful when most importance is given to the data in the leaf nodes of this hierarchical structure or when the intermediate nodes of the hierarchy have data values that can be derived from those on the leaf nodes. Many applications have used the treemap [3] for portfolio visualizations and for tasks that support diversification [7] and even to detect fraud in market trading by adding a third dimension to the entities in which heights of the rectangles represent the change in stock values [15]. In the case of stock market bubble analysis, though the value and performance of sectors is derived from those of individual stocks that make up the sector, analyzing the sector itself is what should be given the most importance. Even though the treemap does divide the available space by sector first and then by stocks, there is no sense of how the sector is performing other than the space it occupies because colors only represent the performance of the stocks. Also when a sector has been clicked and one can see details of all the stocks within that sector, the sector has now taken up the entire space and there is no indication of how it looked in the overview, which may result in the loss of context for the user (See Figure 2.1(1)). Thus there is a need to see sectors as an overview and stocks as details on demand while still preserving context and maintaining the hierarchy.

The treemap represents all market entities as rectangles, with areas representing the price and colors indicating the entity's performance compared to a previous value in its own price. It is however a static visualization that does not encode time and in order to compare market values over a period of time, individual treemaps will have to be compared against each other. Comparing values over time is of importance when studying the nuances of bubble formation because the key is in spotting abnormal behaviors or sudden changes in the trends of certain market entities. Assuming there is a way to animate the treemap so that colors and areas of the rectangles representing market entities

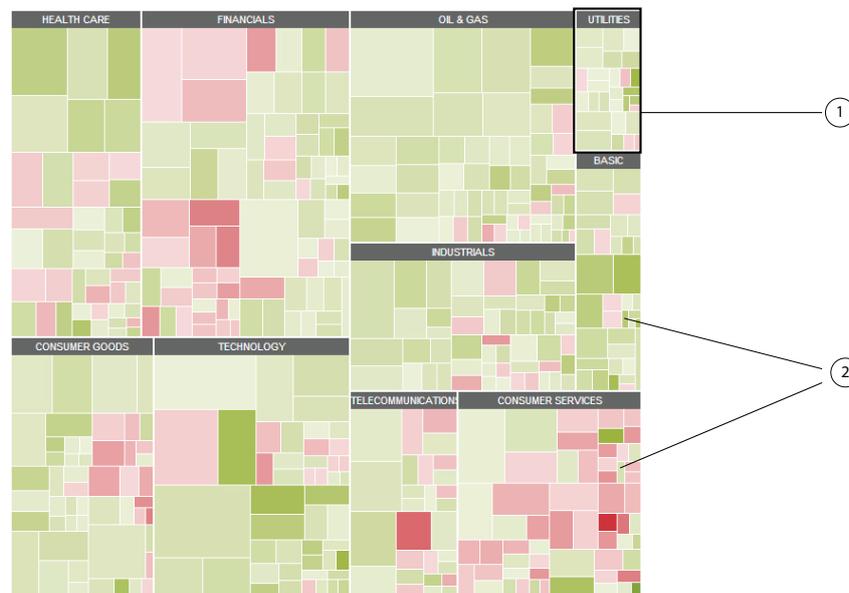


Figure 2.1: Map of the Market showing (1) How a details level would only show that sector without its context in the overview of the market and (2) How some rectangles are very small and distorted

keep changing over to time, the visualization would be very confusing to observe and abnormal behavior would be difficult to spot because the edges of the rectangles would modify to allow for suddenly growing entities. While entities that show a significant growth compared to other entities may be apparent because they seem to be consuming a large space in the current time, a considerable growth that has occurred suddenly compared to itself in the previous time stamp will not be easily observable. Hence there is a need for a visual cue to be given to the user so that attention can be drawn to entities which have shown an abnormal behavior over a period of time.

The treemap by nature works based on the division of a fixed area (representing the market) into smaller areas (representing individual market entities). Because of this, if there are more than one entities showing similar kinds of abnormal behavior, it will not be very obvious during a treemap animation. Also since the available space shows the whole market, changes to the market conditions due to constituting entities is not visible at all; because the market itself (which is the root of the hierarchy) does not seem to undergo any visual changes. It gives the impression of funds being redirected within the market among the entities, when in situations such as a market crash there is a drastic reduction in investments throughout the market where investors have sustained losses. Thus there is a need to not only show market entity performances but also their effect on the market as a whole.

The market capitalization of the entity in a treemap is represented by the area enclosed by a rectangle, dimensions of which can change while not greatly affecting the area. Because of this, just seeing the visualization change (without also observing the actual values simultaneously) may be distracting to the user because some entities may appear to have changed their area while they



Figure 2.2: Yahoo Finance chart for the HP stock, the website found on <http://finance.yahoo.com/>, 2014, Used under fair use, 2014.

were just reoriented to account for changes in adjacent (rectangular) entities. Also, as pointed out in [34] if some rectangles have become too thin, it is difficult to click on them due to this distorted shape as seen in Figure 2.1(2).

Yahoo's Finance Charts [1] use line charts to show single stock market data dimensions over time. As seen in Figure 2.2 a single stock can be selected by the user and an interactive line chart that allows the user to manipulate the time range is displayed. [14] also uses line charts and horizon charts which are also graphs that focus on the historical view of the data. One issue with line charts is that, since line are indicators of individual market entities, the overall state of the market is not visible and the user has to expend time and considerable cognitive effort [4] trying to form a mental model of the whole market's status. A similar situation arises when sound is used to represent one dimension of data as in the case of [22] where different audio cues are delivered to represent trade nature such as purchases and sales; because the user will have to remember a number of these cues if an overall situation of the market needs to be reckoned.

2.2 Time Series Data Visualizations

2.2.1 Animations

Because the need is to monitor the behavior of data over a number of years, the visualization must be one that can show the time-varying nature of the data. The main task for the user is to be able to visualize and study changes that market entities go through over the years. Thus the main challenge is in mapping temporal data to the spatial domain in such a way that cognitive load is decreased and pattern finding is facilitated. One recent visualization that gained much attention from the visualization community is the Gapminder Trendalyzer [25] which shows the wealth and health of nations by denoting the time axis as an animation. It was presented initially more as a presentation tool rather than one to be used for exploration, which is why many studies have been conducted since then to verify its rationality and use as an exploration or analysis tool [24]. Experimental findings by Robertson et al. [24] and arguments to support these findings [32] [33] say that users find it easier to recognize trends and find answers to specific questions on the visualization when it is presented as a set of static images (each showing one nation) or with traces [23] instead of an animation. This school of thought is based on the experimental findings [24] that users were getting distracted when certain entities in the animation showed counter trends because of which they were unable to focus on others (or rather the general trend).

In our proposed design we employ a Gapminder-like [25] animation of the stock market entities because the aim is to identify erratic or abnormal behavior of a market entity. This entity is not known before-hand so the visualization needs to allow the user to be able to pick it out among its peers. Another reason for using animation instead of static images is that, the number of entities which need to be observed and the time over which the observation is to be conducted is very large. Market wide bubbles take a long time to develop. They may even take many years to gain momentum and finally peak and burst. The objective is to identify a potential bubble-causing entity as soon as possible. So if static images of different time stamps have to be used the number of images would be too large and each image would contain too many entities. An animation of all these entities over time along with visual cues of anomalous behaviors, on the other hand will allow the user to spot the entity more easily and can enable further detailed exploration of the same. A drawback of using animation that was found by the experiments in [24] was that animation caused the users to take more time to find the answers because they had to keep playing back the animation before they decided on an answer. That may be the case where the user is not permitted to control the direction in which the animation occurs or cannot play-and-pause required. This is why our design allows for these interactions.

Further results by [24] show that animation is more visually engaging than are static line charts or data logs of stock market performance values. Participants found animations interesting and exciting to use, rather than analyzing static images of the visualization. Other studies [4] [26] show that using some kinds of motion to get the user's attention to indicate an alarm or warning condition has proven to be effective and successful in the past, because the human visual system is

highly sensitive to sudden changes [38]. Our visualization uses animation of the market data over time to show the transitions the market entities have gone through and motion through jitter to draw the user's attention toward abnormally behaving entities. [39] has emphasized the effectiveness of animation in visualizing dynamically changing data like that in the stock market. Dao et al. [8] have also used a similar design in FireStox, where only the changes of market velocity (volume of shares) and force (ratio of buy to sell) attributes are shown for those firms that are chosen by the user through an animation. Another approach explored by Schreck et al. [27] is to allow the user to choose the time range for which updated data points need to be shown. This is similar to giving the user the ability to control the speed of animation being played. Our design gives the user control over the speed in addition to the direction of time line navigation.

2.2.2 Multiple Views and Details

Because of the high dimensionality of stock market data, another challenge lies in being able to present important aspects of the data without cluttering the screen space which can be confusing to the user and cause cognitive overload. [30] uses candlestick charts and trend lines to show the different attributes of financial data to help the user make smart investment decisions. One way to approach this is by using different views which are synced or coordinated with each other and change uniformly [35]. Merino et al. [21] have conducted extensive user studies to test its effectiveness and provide the user with interactive zooming and multiple coordinated views to view time series data in their final interface. [12] shows the design of a visualization that uses bar charts and line charts in multiple views which are coordinated with each other through selection and the time line. It is for any kind of data that is time varying in nature. The main view in TreeVersity2 uses a special kind of bar chart which shows data for a specific time and the secondary smaller views show the line charts that show the data for the entire time for which the data is available. In our design, the multiple windows shows statistics of the main view which serves as stage for confirming hypothesis before conducting detailed examinations. The statistics change according to the animation in the bubble chart window. [5] also uses multiple views in the visualization of financial data where the secondary smaller views are used to show snapshots of historical searches that were done on the main view.

Being able to select the market entity is an important functionality for a visualization that needs to facilitate exploration when the user does not know the exact time or entity that they are looking to study. A drawback of using line charts is in the selection of specific entities because of the large number of entities that may be seen in the visualization at the same time. [20] shows how using glyphs makes selection based tasks easier for the user. [9] uses columns and worm-like structures which are easy to select because of their distinct visual representation. Our proposed novel visualization uses the visual metaphor of circular bubbles to represent the market entities for easy selection. [16] stresses on the effectiveness of using a visual metaphor so that it can be intuitive to the users who interact with it. Another way of dealing with the high dimensionality of data to be visualized is by providing different amounts of detail based on the user's choice [19]. Shneiderman's Information Seeking Mantra [28]: overview first, zoom and filter, and then details-

on-demand, has been followed extensively for successful visual data exploration. Since the purpose here is to find market bubbles that may have far reaching effects in the event of a burst, which are generally spread across whole sectors, allowing the user to have an overview of the entire market is crucial. Research which support this principle include [31] and [17] where a high level overview of the market is first given to users as a means to facilitate exploration. Our design also gives the user an overview of the market with its sectors and details can be viewed when specific sectors of interest are chosen.

Chapter 3

Proposed Design

3.1 Bubble View

3.1.1 Attribute-Channel Mapping

We propose the design of a visualization system using multiple coordinated views which are synced to the same time variable. In the main view which we called the Bubble View, the overview of the market can be seen wherein market sectors are represented by circles of different sizes and colors. The sizes of the circles indicated the market capitalization of the sector at that time (which is the aggregated market capitalization values of all stocks in that sector) and colors of the circle indicate how the sector has performed over a period of time. Market capitalization of a stock can be computed by taking the product of the stock's price and the number of outstanding shares at a particular time. This attribute of market data has been chosen for the visualization since it shows a very high value for entities that contain or contribute to a bubble in the market by revealing continued high return levels. In the visualization, when trends show stocks of a certain sector being sold in the market, capital can be seen being funneled into other sectors while one sector starts losing capital. A derived attribute which can also be used to study the formation of a bubble in the market is the performance of different stocks. This is a measure of how the market capitalization value has changed over time. Two ways to calculate and map the performance of the stock to the bubble view visualization are described below.

Geometric Mean Return Calculation The color of the stock is the deviation from the Geometric Mean of its return. Geometric Mean of the stock is a better indication of the performance of the stock than just its market capitalization value [6] [40] because, the actual values of the investments are hidden and a more fair comparison of the efficiencies of stock performances can be made. It shows the compounded growth of the market capitalization of a stock over a specific time period as selected by the user. When the stocks do not pay any dividends, It the geometric mean monthly

return (R) is calculated as follows:

$$R = \left(\frac{V_n}{V_0} \right)^{(1/n)} - 1 \quad (3.1)$$

Where V_n is the current market capitalization of the stock, V_0 is the market capitalization at the beginning of the first month selected by the user, and n is the number of months that have passed between the first month and the current one. When stocks pay dividends, the formula used to calculate the geometric mean return is modified to:

$$R = ((1 + R_1)(1 + R_2)(1 + R_3) \dots (1 + R_n))^{(1/n)} - 1 \quad (3.2)$$

where n is the number of months that have passed since the beginning of the visualization, R_1 is the initial return in the first month and R_n is the current return in the last month. We come up with the following stages of (increasing) risk as a stock begins to earn unexpectedly large returns which is an indication creation of the bubble, and thereby the potential of the stock bubble to burst.

Stage of Decline: $R < 0.0$

Stage 1: $0.0 \leq R < 0.015$

Stage 2: $0.015 \leq R \leq 0.025$

Stage 3: $R > 0.025$

Stocks in decline are green in color as they are in no current danger of creating a bubble. However the other stages turn deeper shades of red as they increase in size indicating high levels of danger of an imminent bubble burst. The rationale behind color assignments in this approach is that red is a color generally used to represent a dangerous situation or a warning. Users realize that a stock market crash or bubble is an unwanted scenario that should be avoided, so a bubble condition is represented by red and deeper shades of red representing increasing danger. This color scheme however does not fit the average mental model of a person with finance or accounting expertise. Users that study stock market data in any form are accustomed to a certain color scheme for representations of stocks that they tend to retain. This includes the one that is used in Map of the Market [37] where red is used to represent stocks whose prices have fallen and green for those whose prices have risen compared to some previous initial value. Investors also usually follow this scheme wherein they would rather invest in a stock which is performing better than before, and therefore see red as an indication of bad stock performance trends and green as good performance. Their mental model is therefore tuned to spot greens in the visualization. The GMR mode in StockBubbler, however uses red to indicate better performance and green for badly performing stocks because the objective here is not to let the user decide where to invest but rather to let them detect the onset of a market wide crash and its origins.

100 Percent Visualization Calculation Performance of the entity (say sector S1) is calculated by a one year moving average as follows:

$$A_c(S_1) = \sum_{n=0}^{t-11} \frac{M_{(c-n)}}{t} \quad (3.3)$$

$$P_c(S_1) = \frac{M_c - A_c}{A_c} \times 100 \quad (3.4)$$

Where A is the moving average, c is the current date, t is the number of data points that is needed for one whole year (12 in case of monthly data), M is the Market Capitalization value and P is the current performance of the entity. Stages of bubble are defined in the following way so as to assign specific colors to each of the stages.

Negative performance if Lesser than -50%, -50% to -20%; -20% to -10%, -10% to 0

Positive if performance is between 0 to 10%, 10% to 20%, 20% to 50% and Greater than 50%

Reds indicate a negative performance, and greens indicate positive performance where deeper shades are used for increasing values in each direction. Thus the deepest red is performance worse than -50% and the deepest green is a performance better than 50% over the past one year. In the details view, blue shades are used to indicate IPOs which have just recently gone public and do not have one year of data to depict a one-year performance.

As mentioned in the previous section, since the visualization is an animation, at any given time, data for the current date is being displayed. From this overview, interesting sectors can be selected and viewed in the details view where a similar visualization (same attribute-channel mapping) is shown with individual stocks belonging to those selected sectors.

3.1.2 Circle Packed Design and Interactions

The entities themselves, which as mentioned before, are represented as circles arranged around a central point on the screen using a circle packing method [34] which ensures a minimum distance around each entity so that there is no overlap among them (See Figure 3.1). Thus at any given time the market entities are radially arranged around the center of the screen in a cluster like structure. However unlike Wang et al.'s [34] approach, because the entities will change sizes as the animation plays or when the time line is traversed, the layout for the packing is not fixed throughout the time period. This is why we include two interactions between the entities, one

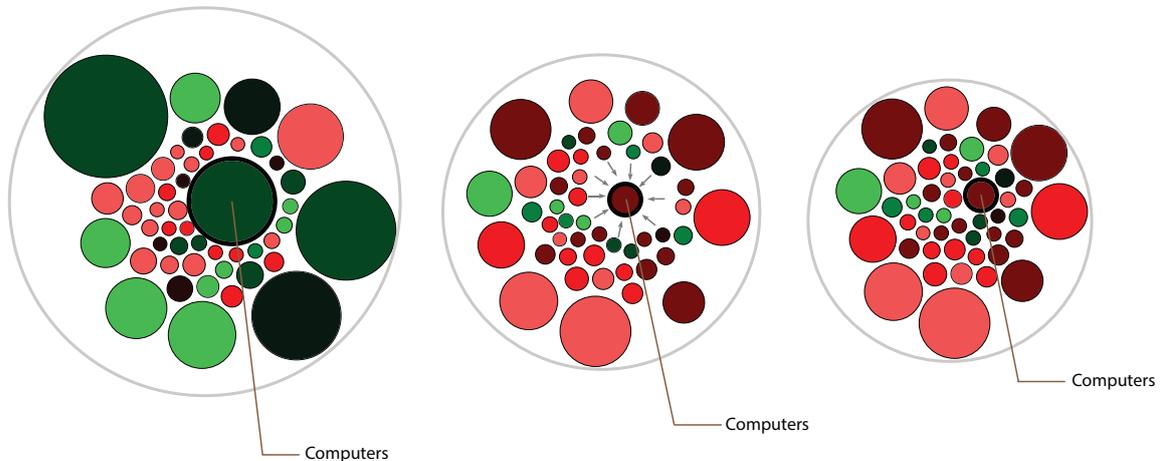


Figure 3.1: The sector 'Computers' shows sudden drop in market capitalization value after a market crash. Other sectors rush in to fill the empty space left by this sudden shrinking.

that ties to pack them all together by coalescing them toward the center and another that keeps them apart by ensuring a minimum space around them so that they don't overlap when their sizes change. Due to these interactions between them, the entities which are larger in size will tend to go toward the periphery of the cluster while smaller ones tend to move toward the center in order to achieve the most stable positions with the minimum space around all of them preserved. The entities are enclosed by one large circle which in the case of the sector view, represents the size of the market (aggregated market capitalization values of sectors) and in the details view of the stocks themselves, the enclosing circles will represent the sectors that were chosen in the overview. Thus the same metaphor that is used in the overview level is continued in the details level. New entities will always originate at the center of the cluster, though such a condition only arises in the details level when firms go public and become IPOs. The enclosing circle always has a very faded inner glow from the center which gives an indication of the overall performance of all the entities within it. Thus in the overview level, the market entity will have an inner glow of the color of its overall performance in that time and in the details level the sectors will show an inner glow of the same color that they have in the overview level at that same time. This makes sure that the user does not lose context when observing the details view of how the sector is performing in the market. When the visualization is played or the time line is being traversed by the user, the entities change color according to their performance over the past year and change sizes according to their current market capitalizations. Due to the force model employed to keep the entities clustered but apart, the changes in the sizes of the entities affects their layout to account for either the lack of space or availability of additional space which may be created. For instance, when an entity suddenly

shrinks in size (drastically so), this creates a lot of additional space around it which is immediately filled in by the adjacent entities which try to keep the cluster-form intact. This effect is seen all over the market and the overall market size will subsequently adjust according to the placement of all the entities within it. Such a situation will be effectively visualized by this design when the market has crashed and investors have sustained losses (i.e., capital has not been redirected into other sectors) due to which the whole market takes a hit. Conversely when a sector has suddenly shown rapid growth, it will grow in size very fast because of which adjacent sectors will be pushed outward toward the periphery of the cluster. Thus the market size will be seen to increase if this has happened in a situation where new investment has been funneled into a particular sector rather than being redirected from another one. The presence of many such sectors is generally a sure indication of an imminent bubble formation in the market. Hence this is a situation which should be easily detectible by the user.

Due to the dynamics of the force model, as explained before, large bubbles tend to move outward from the center of the cluster. If a sector has suddenly grown, it will become unstable and begin to jitter as it tries to find its way out of the cluster to a location where it can become stable again, and while doing so, disturb the sectors around it while they adjust their location so as to avoid overlaps. This movement is key to spotting the abnormal or unstable nature of the market which could be the indication of a bubble formation that may result in a crash. The same metaphor is preserved while looking for details within sectors at the individual stocks. The stocks are now enclosed within the sectors and it can be noticed which firms are causing the behavior of the sector that was seen in the overview level. The sector in this details level (the circle enclosing the stocks) will change sizes in the same way as it did in the overview. One additional phenomenon that can be seen in the details level is the formation of new IPOs (represented by a blue shade due to lack of sufficient data to calculate a performance) in the center of the sector cluster. This is easy to spot because they originate in the center and will cause an increase in the size of the sector as they push themselves out and try to reach a stable location without overlap. This kind of effect is generally seen in the case of a sector which is harboring a bubble because investor excitement causes many IPOs to be established which may or may not be able to sustain their growth in the market after the crash occurs.

Interactions which are supported in this view include manipulation of the location of entities by a click and drag mechanism, bidirectional navigation of the time line, and selection of entities for highlighting and showing details. Users can select market entities in the Bubble View and see in a small information tab, the name of the selected sector or firm and a few details about it such as number of firms that belong to it (in case of sector), its current market capitalization value, its percentage performance change over the past year and the stage of bubble it is in. They can be dragged around to relocate them and when they are let go, they will again be pulled toward the cluster and head for the most stable position like in a force directed layout scenario. This way, if the user is interested in looking at a certain number of sectors, they can all be brought near each other while the animation plays, so that a visual comparison can be made easier due to their close proximity in space. The user can also search for a specific entity using its name, or SIC code in case of sectors and Permanent Number in case of stocks and by doing so can automatically

highlight that entity in the Bubble View. The selection will be preserved during navigation of the time line in that the selected entities will remain highlighted until they are deselected by clicking again or using an option to clear all selections. The selection made in the Bubble View will also be reflected in the other views as will be explained in coming sub-sections. Thus this visualization will work better than an animated treemap by solving the issue wherein abnormal behavior needs to be discernible by using the jitter that entities will exhibit. Other concerns of the treemap that will be addressed by our design include that of showing overall market condition where the scale of effect of an entity's abnormal behavior can be observed. Relative sizes of the entities is easy to compare because the change in dimension of a circle clearly affects its size unlike in the case of rectangles wherein change in dimensions gives the impression of size change even if the area has not altered. Space on the screen is also given meaning unlike in the case of treemaps where locations of entities spread throughout the screen space have the same meaning. Our design on the other hand is built around a focal point on the screen around which all the entities try to cluster. Their distances from the center are meaningful in that larger ones tend to navigate toward the periphery and smaller move toward the center. Also since the entities are clustered around the center, they do not take up the entire space of the screen because of which it is still relatively simple to keep track of those on the periphery. Using Gestalt's principles [13], the contour around the entities which represents the whole market gives the indication of grouping which keeps the user's focus within the market bubble [36].

3.1.3 Formative Study

The proof-of-concept formative study conducted to test the Bubble View of our visualization used the feedback from 19 participants to verify that the design is indeed fruitful in enabling the user to identify sudden changes in the behavior of stock market entities. All participants had a background of either finance, business or accounting studies, 17 from the graduate level and 2 from the undergraduate level. Participants with such backgrounds were chosen because they were required to possess a basic knowledge of what a stock market bubble means and have a clear understanding of the attributes of stock market data. Participants used data of 100 firms in the NASDAQ stock exchange that contributed the most to the dot com crash of 2000. This is a stock market crash which had one of the most damaging effects in recent times, which is why we expected the visualization to clearly show the bubble's formation, peak and burst. However the specifics of the data was hidden from the participants because we did not want them to use their own personal knowledge of the crash to figure out where and how the bubble occurs in the visualization they were seeing. They were only told the aim of the visualization, the attribute-channel mapping and that they were to use it to find around what time they can observe the bubble in the market. Their response times were recorded. After noting the point of the bubble burst, they were asked to answer a few questions based on their experience.

Participants were asked to rate their agreement with statements on a 7-point Likert scale with one being strongly disagree and seven being strongly agree. Two statements were related to the sizes of the circles used in the visualization while two other statements were related to the color

Table 3.1: Participant responses to visualization statements

Statement	Average 100Percent Mode Response	Average in Means Response	Difference (Std Error) GMR Mode	T-stat
The size of the circles was useful to me for identifying when the bubble burst.	6.42	6.42	0.00 (0.1)	0.00
The color of the circles was useful to me for identifying when the bubble burst.	5.84	4.68	-1.15 (0.43)	2.67**
The frequency of color changes was useful to me for identifying when the bubble burst.	4.94	4.10	-0.84 (0.39)	-2.15**
The frequency of size changes was useful to me for identifying when the bubble burst.	5.57	5.57	0.00 (0.18)	0.00
The visualization (design) was easy to understand.	6.15	5.73	-0.42 (0.28)	-1.46
There were too many stocks (circles) represented in the visualization.	3.15	2.47	-0.68 (0.27)	-2.48**
Overall I was able to identify the bubble burst in the visualization.	5.68	5.73	0.05 (0.14)	0.37

Notes: *, **, * denote significance at the 10%, 5%, and 1%, respectively.**

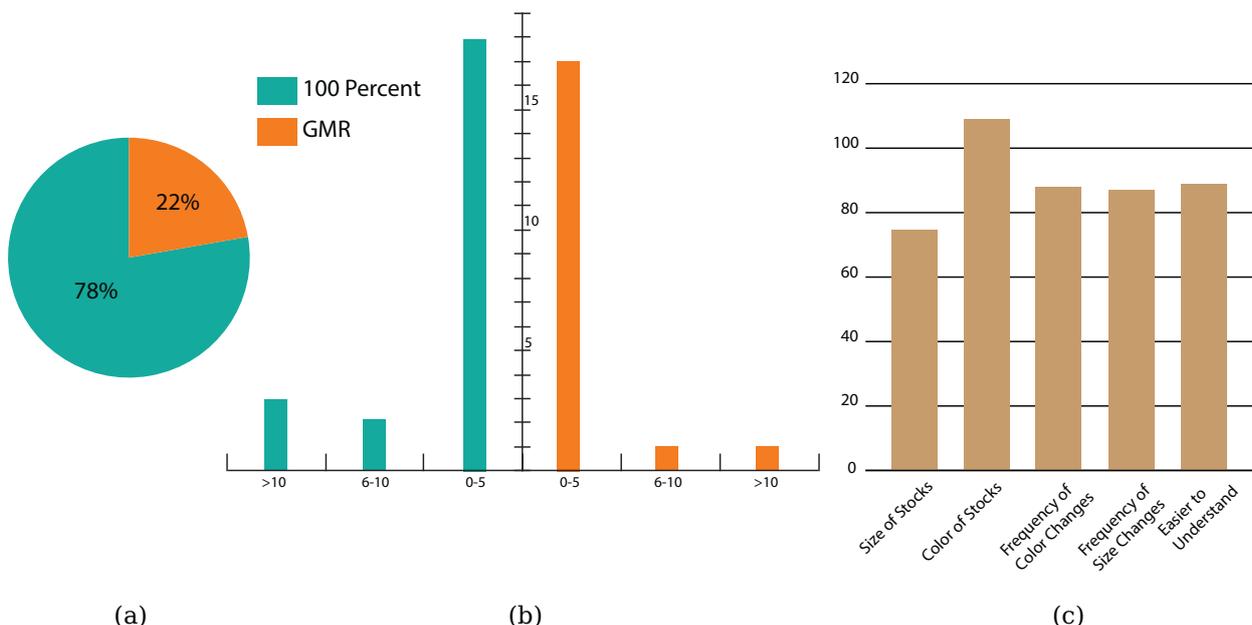


Figure 3.2: User Study Statistics (a) Preference of users (b) Distance from correct answer (c) Answers regarding most useful feature

schemes used. The results, as tabulated in Table 3.1, indicate that participants rated the color schemes as more significant to their overall decision making than the size of the circles and the frequency of size changes. These results may be due to the possibility that the color changes were more easily understood by the participants than the changes in the size of the circles. Circle sizes were based on the market capitalization of the individual firm. Although many participants may have had an understanding of shares outstanding and stock price, which are the two elements that comprise market capitalization, the participants may not have fully understood the overall importance of market capitalization. However, color changes were dependent on whether the stock has increasing or decreasing performance over the past year. This is almost certainly easier to understand for participants. On the other hand, it is possible all participants completely understood the concept of market capitalization and their responses were based upon the fact that they preferred the usefulness of the color scheme over the size of the circles. Both effects may be driving the differences in response rates in Table 1. Figure 3.2(a) shows that 78% of the users preferred the 100Percent mode to the GMR mode however from Figure 3.2(b) it can be seen that users were able to answer more accurately with the GMR mode. Figure 3.2(c) shows various factors that the users found useful because of which they picked the mode they preferred. From it, it can be seen that though most of the features were useful, the most important one is the color of the stocks which is why most users preferred the 100Percent mode to the GMR mode. Thus users mostly pick the mode that most fits their existing cognitive model.

Table 3.2: Regression Estimations

Variable	Parameter Estimate	Standard Error	T-stat
Intercept	-0.499	1.897	-0.26
FirstVizGMRDummy	-4.003	1.783	-2.24**
GMRDummy	-1.900	3.644	-0.52
Seconds	0.054	0.016	3.31***
SecondsInteract	0.035	0.042	0.83
PlayPause	-0.468	0.248	-1.89*
PlayPauseInteract	0.319	0.315	1.01
Keystrokes	0.006	0.019	0.33
KeystrokesInteract	-0.026	0.026	-0.99
Adjusted R2	50.56%		

Notes: *, **, * denote significance at the 10%, 5% and 1% respectively. The dependent variable is distance which is calculated as the absolute value of the distance between the participant's response as the highest stock market capitalization value and the actual value.**

Table 3.2 shows the parameter estimates from a regression of the following equation:

$$\begin{aligned}
 \text{Distance} = & \beta_0 + \beta_1 \text{FirstVizGMRDummy} + \beta_2 \text{GMRDummy} \\
 & + \beta_3 \text{Seconds} + \beta_4 \text{SecondsInteract} + \beta_5 \text{PlayPause} \\
 & + \beta_6 \text{PlayPauseInteract} + \beta_7 \text{Keystrokes} \\
 & + \beta_8 \text{KeystrokesInteract}
 \end{aligned}$$

where, Distance which is calculated as the absolute value of the distance between the participant's response as the highest stock market capitalization value and the actual value; *FirstVizGMRDummy* is equal to one if the first visualization the participant saw was the GMR visualization; *GMRDummy* is equal to one if the distance dependent variable is for the GMR visualization; *Seconds* is the number of seconds the participant took to analyze the visualization; *SecondsInteract* is equal to an interaction between the variables *Seconds* and *GMRDummy*; *PlayPause* is equal to the number of times the participant hit the space bar during the visualization which pauses, then plays the visualization; *PlayPauseInteract* is equal to an interaction between the variables *PlayPause* and *GMRDummy*; *Keystrokes* is equal to the number of times the participant hit either the up or down keys to move the visualization forward or backward on period of time; *KeystrokesInteract* is equal to an interaction between the variables *Keystrokes* and *GMRDummy*.

The variable *FirstVizGMRDummy* is included to account for participant learning effects between the two visualizations. The design of the study was among subjects meaning each participant looked at both visualizations. Because there are learning effects from the first visualization to the second, we alternated the order of the visualizations so the effects will not be systematic and

influence the results of the study. The coefficient on the `FirstVizGMRDummy` variable indicates the learning effects were most pronounced for the participants who saw the GMR visualization followed by the 100Percent visualization. These participants, on average, were over 4 months more accurate with their estimation of the peak of the stock market bubble for both visualizations. This may be due to the fact that the GMR visualization used colors which were different from the participants expectations. This may have led the participants to pay closer attention to the visualization when faced with the GMR visualization first, and when they subsequently saw the 100Percent visualization, which had a color scheme that matches expectations, the participants may have had a clearer idea of when the stock market bubble occurred.

The results shown in Table 3.2 also indicate that, if a participant took more time evaluating the visualization, the distance between the highest peak of the stock market capitalization and their estimation of that point was greater. That is, the more time the participants took, the worse their estimations were. This effect is captured in the positive and significant parameter estimate on the `Seconds` variable. We attribute this effect to the different backgrounds of the participants. Some had backgrounds in finance and may have been more familiar with stock market visualizations. They may have been quicker in deciding which month had the highest market capitalization. Other participants had backgrounds in non-financial fields such as tax and information technology. These participants may not have been very familiar with stock market visualizations and may have taken longer to perform the task. These conclusions are conjectures because we did not collect extensive participant background data on previous stock market investing and visualization experience.

Table 3.2 also shows the accuracy of the estimated peak of the stock market bubble is greatly improved by increased effort on the part of the participant in the form of pausing and playing the visualization which is captured in the `PlayPause` variable. As a participant is trying to figure out which month is the peak of the stock market bubble, their effort in trying to figure it out is captured in the number of times they played and paused the visualization. The significance of the `PlayPause` variable shows that, the more a participant paused and played the visualization, the more accurate their estimation of the peak of the stock market bubble became.

We include interaction variables between `GMRDummy` and all other variables of interest. If significant, these will show a difference in the variables of interest between the GMR and 100Percent visualizations. None of the interaction variables is significantly different from zero. The lack of significance of the interaction variables indicates there is no statistical difference in the effects of `Seconds`, `PlayPause`, and `Keystrokes` on the accuracy of the participants estimations of the peak of the stock market bubble.

3.2 Statistics View

In order to support and further investigate hypothesis formed using the Bubble View, we use other views which are synced to show statistics of the visualization in the Bubble View. Four types of statistics are used in these views.

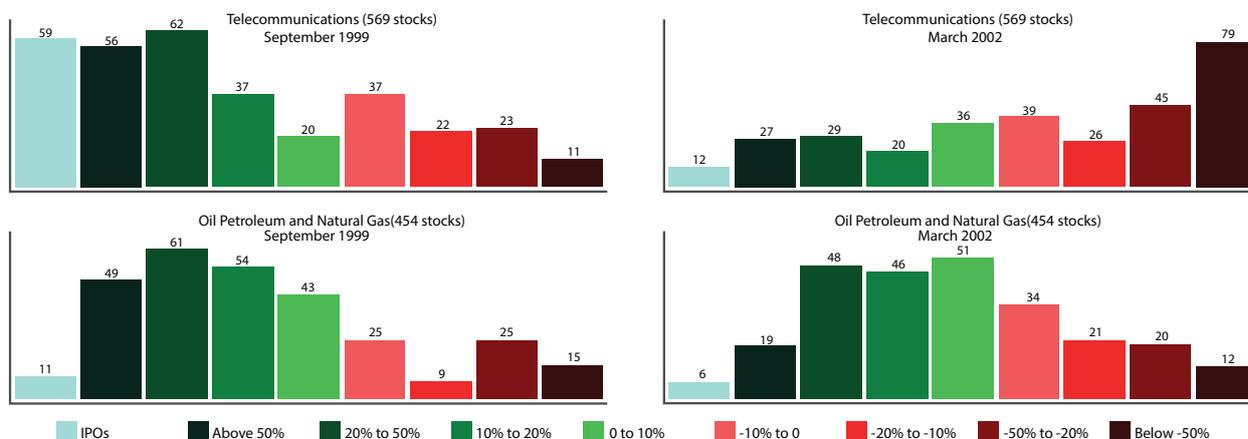


Figure 3.3: Details view for Telecommunications and Petroleum sectors before and after the Dot Com crash. Notice the number of IPOs has fallen drastically for the bubble sector Telecommunications and has remained almost the same for the non-bubbling sector Petroleum

Histogram This shows with help of a bar chart, how many entities belong to each stage of the bubble. The colors of the bars match the color assigned to the particular stage of the bubble that it represents. Thus at any given point on the timeline, this view will show the histogram of what is seen in the Bubble View. Clicking on a bar in the histogram will highlight those entities in the view for the duration of the click as a quick way to visually link both the views. The same applies for the details view when stocks are seen in the Bubble View. In this case there are as many histograms as the number of sectors selected and an additional bar in each histogram also shows how many firms in the Bubble View are still IPOs (as shown in Figure 1(f)). Firms are counted as IPOs for one year after their establishment until they accumulate enough data to be able to calculate their performance. If a sector containing a bubble and one without a bubble have selected for the details view, the histograms for these sectors should ideally show reverse trends as the time line is traversed, i.e. during the bubble formation and burst the bubbling sector will show very high green bars and few red while after the crash it will show the opposite. The non-bubbling sector on the other hand should may be affected by the bubble in the market in the opposite way if funds are connected or not be affected at all (See Figure 3.3).

Line Chart Line charts, as mentioned before have been traditionally used for a long time by market professionals and students as the standard method of visualizing market data in the form of graphs. Whether they are interactive visualization or just a graph to summarize a subset of data, individuals in the Finance community are extremely accustomed to using them on a regular basis and are quite adept at reading them to get near-accurate estimates of the data that they are seeing. Since line charts are, by the way they are designed, useful for showing historic data of the entire time, we incorporate two types of line charts in our attempt to design the best tool for stock market visualization. The first line chart shows the performance of the sectors of the market and maps them against the time line. The second line chart shows the actual market capitalization values

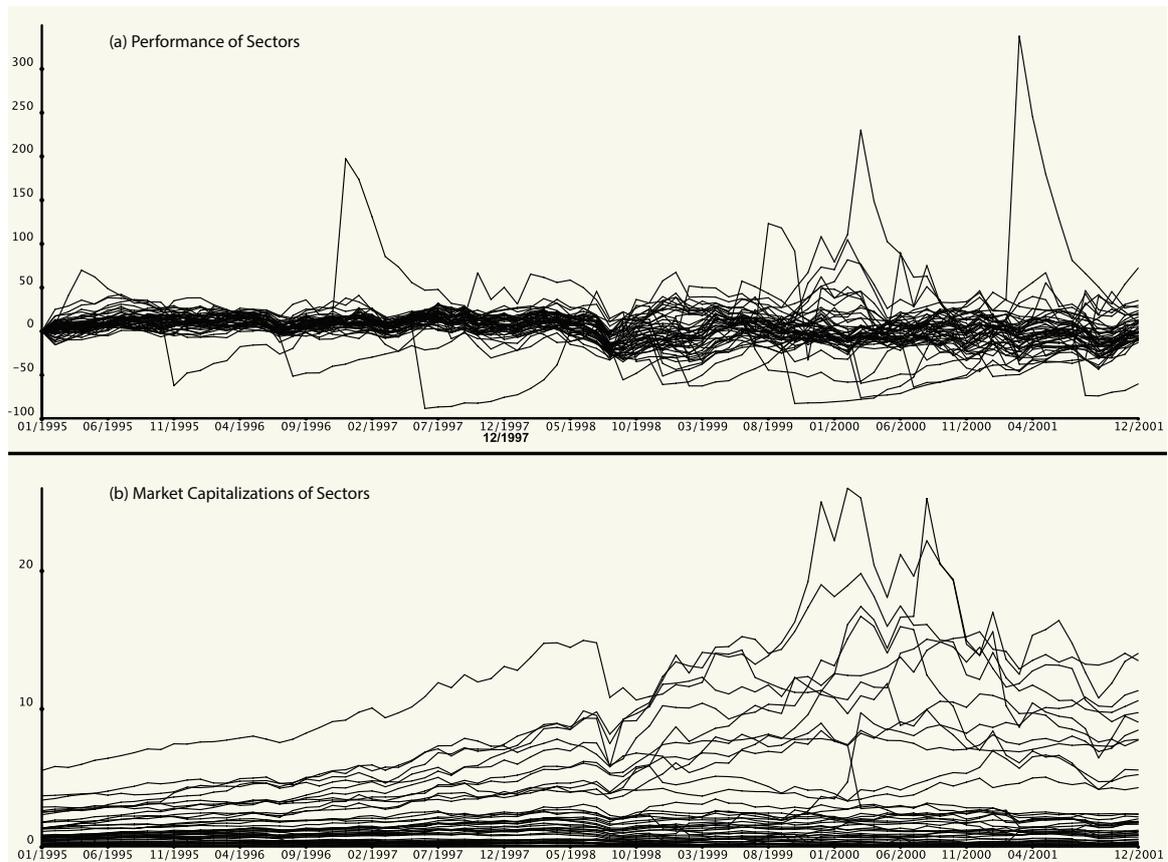


Figure 3.4: Line Charts showing 48 sectors of the Market (a) Performance Chart (b) Market Capitalization Chart

of the sectors against the time line. The line chart plots the market capitalization of all sectors against time and highlights all the dates that have already been traversed by the time line in the Bubble View. There are as many lines as there are sectors in the Bubble View. This is helpful when studying the movement of one particular sector and gives a sense of its movement in the past. It can also be used to compare how much it has grown in the past. The lines which belong to sectors that are selected in the Bubble View will be highlighted so that tracking it is easier through the animation. This also eliminates confusion because the line chart can get very cluttered due to the large number of lines. This chart only for the overview level as observing hundreds of lines for the details view of stocks will cause overload and is not worth the cognitive strain it causes (See Figure 3.4).

Median and Strike Chart This is a bar chart which shows how unique each sector is when compared to the rest of the market. It can be used to judge the abnormality of the performance of a sector after it has been identified in the Bubble View. The median for Sector S1 is taken among the values we get by comparing the performance of S1 against each of the other sectors in the market (S_i) which is

$$P_c(S_1, S_i) = \frac{P_c(S_1) - P_c(S_i)}{|P_c(S_i)|} \quad (3.5)$$

The median value may be positive or negative. Positive values are shown with green colored bars and negative medians are shown as red colored bars. There are as many bars as there are sectors in the overview level of the Bubble View. Such a calculation basically gives an indication of how a sector has performed when compared individually with all the other sectors. Median is a better way to calculate it than average of these values because it cannot be skewed to one direction due to the presence an outlier value. An outlier value may occur because that particular sector which the selected sector is being compared to is abnormal and not because the sector in question is abnormal. Thus positive values of medians indicate that sector is doing well when compared to the other sectors in the market at the current instant of time. However what is interesting is how long the sector has been doing well compared to the other sectors in the market. This will give a better indication about how serious the bubble contained in that sector is and how close it is to causing a burst and crash. This is what the strike is calculated for. The strike for each sector is shown as an inverted bar chart and is aligned with the median of each sector. The strike count of one sector at an instant of time shows for how many time points before the current time point the sector has shown a positive or negative median of performance comparison. Thus a green strike bar of value 5 means this sector has shown positive medians for the last 5 time stamps when compared to the rest of the market. This chart also is only for the overview level as observing hundreds of bars for the details view of stocks will cause overload and is not worth the cognitive strain it causes (See Figure 4.6).

Chapter 4

Case Studies

4.1 Using Different Performance Measures

One data set that the StockBubbler uses is that of values of 103 firms of the NASDAQ listing between the years 1994 and 2004 which is the period where the dot com bubble burst occurred. Technology companies were booming during the first half of this period because of which many investors were excited to heavily invest in them. Their stock values rose exponentially until there was a massive sell-off that occurred in March of the year 2000 which is now identified at the point of the bubble burst. We test StockBubbler to verify whether an animation of stock values during this period using both modes reveals the bubble bursting or not.

GMR Mode

In this mode, as mentioned in the previous section, the colors of the stocks represents the geometric mean return of the stocks. During this period, most firms were not giving dividends so equation (1) can be used to calculate this return for each stock. Since deeper shades of red indicate higher returns, as one moves closer to the point of the burst, the red colors of most of stocks keep deepening until the peak after which the reverse begins to happen. The start date can be set by the user in this mode. The default start date is initially set to the start date of the data set which is January of 1994. When StockBubbler plays from this start date, the GMR mode shows an increase in most GMR values of the stocks but point of the peak is not very evident unless the more booming firms are selected and actual values can be seen to be changing, that is, the trend in colors do not drastically change after the bubble peaks. However once a region of interest (a part of the time slider when it seems like something significant has happened) has been noticed, moving the start date closer to this date, will cause the change in color trend to start becoming more and more pronounced. Thus we begin to enter start dates closer to this value, the peak can be precisely identified by seeing

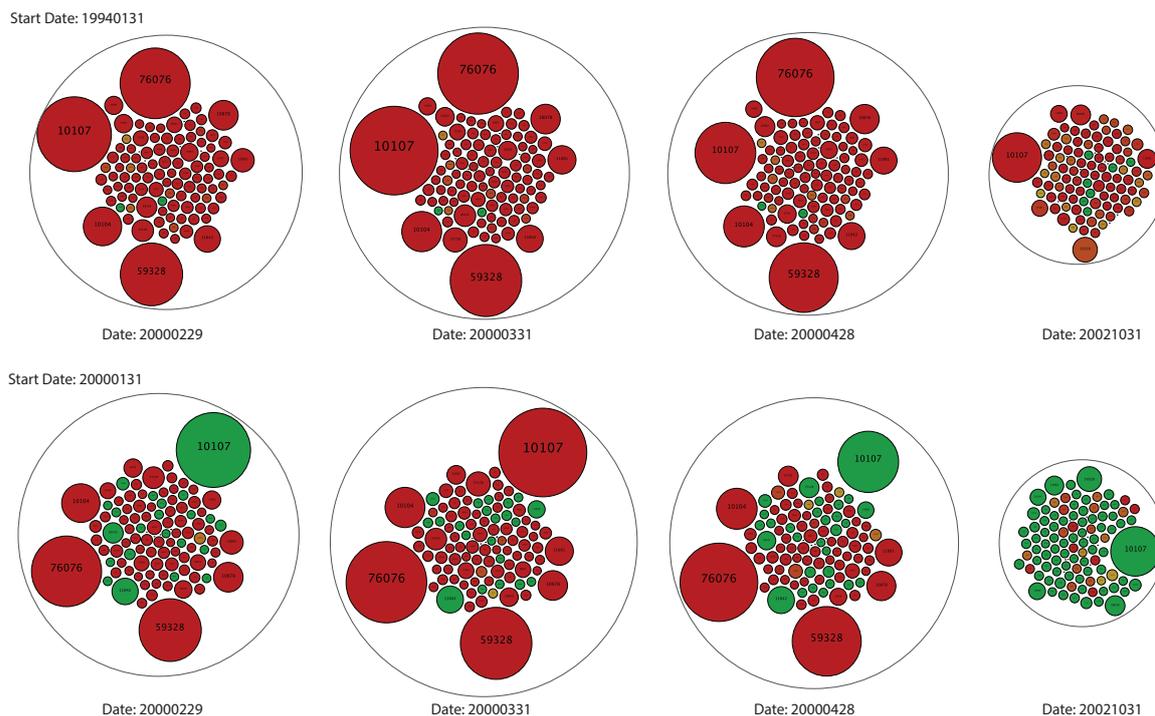


Figure 4.1: GMR Mode: Effect on colors by Selecting different start dates

most of the stocks turn into lighter shades of red and even green (if the start date is only one or two months before the peak). As it can be seen in Figure 4.1, when the start date is long before the point of the bubble peak, like January of 1994, the colors are mostly all red and the changes are not that drastic even for October 2002. However when a date close to the peak is selected like January of 2000, the drastic drop in performance is clearly visible.

100Percent Mode

The most easily observable trend in the 100Percent mode is, because of the presence of the threshold function, the presence of many gray circles is a clear indication of a stable market with insignificant changes in stock performances. Alternately, many fluctuations in colors as the animation plays is a strong indication of an unstable market that may result in a bubble burst. In the 100Percent Mode, the effect of many IPOs coming into existence before the peak occurs is very easily observable, since in this mode, a 12 month moving window is used and a firm that comes into existence on a certain date will not contain any color until it completes one year of existence in the market since that date. The first 12 months of the mode will show only blue bubbles since there isn't enough data available to calculate the 12 month percentage change in performance. Once the 12 month period has been crossed, the stocks take on the appropriate colors based on their performance changes. As can be seen in Figure 4.2, once the bubble burst happens, an instant effect other than the reduction

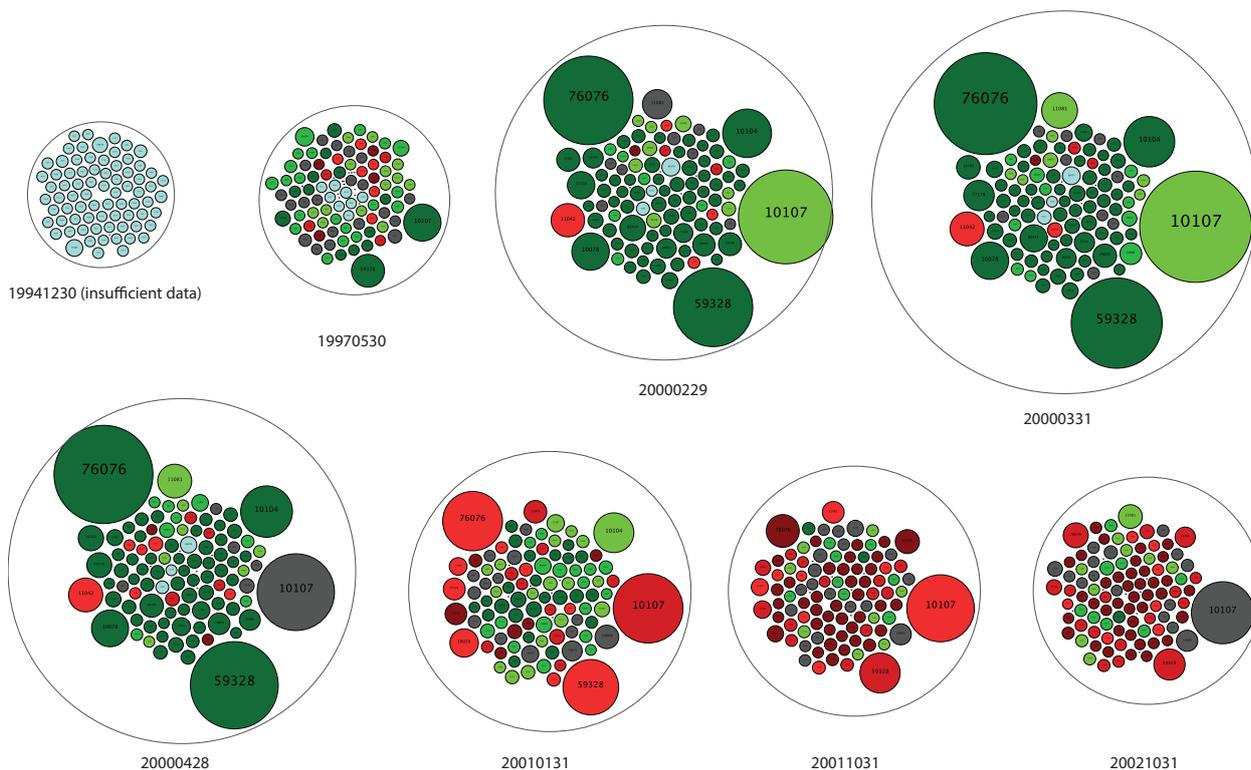


Figure 4.2: 100Percent Mode: Flipping of colors trend after burst occurs

in size is the lightening of the greens in the stocks. This is because the performance starts to fall as investors begin to sell their stocks. A year from that point stocks begin under-performing and this is visible in the red color of many of the stocks in the market.

4.2 Details on Demand Level

A market regulator is usually tasked with duties such as [10] maintaining market confidence and reducing crime and frauds in the market. Thus, the visualization should help the regulator detect areas for investigation which may seem suspicious of fraudulent activity before a bubble has peaked and caused a crash in the market. In order to test whether the visualization is able to facilitate the user to detect such frauds, a sector that was bubbling in the overview, Telecommunications, was selected for details in the Bubble View. In this view, the animation was played until right before the crash and 5 stocks were asked to be identified as potentially fraudulent. Details of the stocks however were hidden so that knowledge of the market would not interfere. Out of 5 stocks that were selected, two were stocks belonging to firms that were found to have committed financial fraud (WorldCom Inc., and Lucent Technologies Inc.). The darkest bar in the histogram was clicked to first take notice of stocks that seemed to be doing extremely well and those that showed

an increase in sizes and remained dark green were finally picked. We were told that it was easy to do so because these large sized stocks kept coming to the periphery of the cluster which made it easy to track them as the animation played. It is notable that the animation was played just once before the observation was made. Thus keeping track of entities in the cluster's periphery seems to be a task that does not tax the user and cause cognitive overload of any kind even though there were roughly 500 stocks in that sector.

4.3 Using Multiple Views

The visualization was used by two finance professionals from the Pamplin School of Business at Virginia Tech and they were able to explore and analyze monthly stock market data of 6 years from 1997 till 2002. This period was chosen by them because there were two distinct market phenomena that occurred and they wanted to test if the visualization was able to catch it and if yes how obvious it was going to be. Sectors of the market were created using the Fama-French [11] sector classification which causes stocks of the market to be classified into one of 48 sectors. The following case studies outline some interesting insights that were gained about the market using the implementation of the design that is proposed in this paper.

For observing the difference between a sector that contributes to the market bubble and one that does not, we pick Hotels, Rooming, Houses and Camps for the non-bubbling sector and Business Services for the one that contributes significantly to the bubble and analyze them at different times before, during and after the burst in the enterprise view. Figure 4.3 (a) shows that though size can be indicative of the growth in the market capitalization values, the performance is indicated by the colors of the circles that represent the firms. Comparing the ratio of green firms to red in the three stages of the bubble, it can be clearly seen how Business Services contributed more to the bubble burst than did Hotels, Rooming, Houses and Camps. Thus once the user has selected the interesting sectors, the actual values of the firms contributing to the bubble and how much so can be observed in the enterprise view.

The Dot Com Crash of 2000

One obvious event that occurred during this period was the crash of the dot com market bubble in 2000. Many telecommunication, business services and electronics manufacturing firms experienced enormous growth in the run up to the bubble peak. However when investors realized that the expected high returns would not be achieved, a massive sell-off occurred in March of 2000 when the market crashed. In the Bubble View it can clearly be seen that around 1999 and 2000 many sectors are doing exceedingly well and it can be verified from the histogram, which shows many sectors are failing and few that are falling into categories that are observing above 20% returns growing because people are excited to invest in just them as seen in Figure 4.4. On selecting

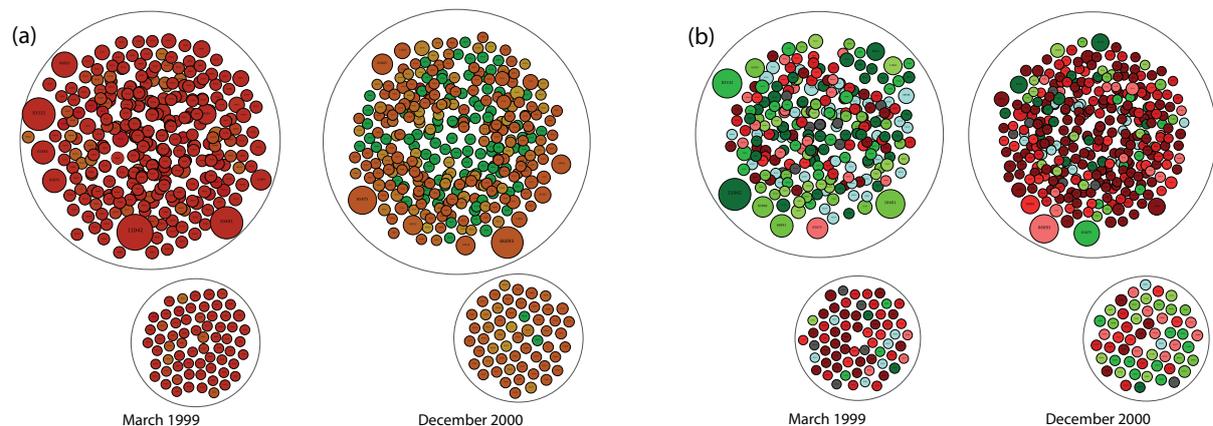


Figure 4.3: Enterprise view where larger sector is “Communications” and smaller sector is “Hotels, Rooming, Houses and Camps”. (a) Geometric Mean Return (b) 100 Percent.

these sectors for details in the Bubble View it is seen that they are sectors like Telecommunication, Business Services, Electronics Equipment and Computers. When these sectors are highlighted, those corresponding to them on the line chart also become highlighted and it can be seen how they stand above the rest of the market at this period of time. However in order to see what they did to the market as a whole, the visualization was played as an animation from that time onward. On March of 2000, it can be seen suddenly most of these sectors have shrunk in size drastically making the market circle small thus showing that the poor performance of these sectors affected the performance of the entire market. On playing the animation further, it can be seen how these sectors tried to recover but were finding it difficult to do so. However the market can be seen to recover as a whole, as people began moving their investments into other sectors after having lost investments in the technology sectors. Because of this it was difficult for the technology sectors to bounce back up.

This led to the observation of sectors like Utilities and Petroleum. These sectors were spotted after the crash because they seemed to be falling into the bars that were showing sector performances above 20%. When they were highlighted, it was seen from their line graphs that they were fairly flat throughout the time period of 6 years that were under observation, i.e., before, during as well as after the crash. The reason for this is because, sectors like these are always invested in because they are necessities that are not affected by the state of the market and people who did invest in them did not lose money during the dot com crash. Many others however began investing in these safe sectors again after the crash which is why they fair better than other sectors even though their performance is not great even after the crash. That is, they stand out above the others (See Figure 4.5) because they didn't go down as others did and not because their rate of return was increasing substantially.

When observing the histograms for the details view for a sector that is bubbling and one that isn't, the differences become quite evident due to the establishment of IPOs in the bubbling sector. As can be seen in Figure 3.3, a little before the crash happened (September 1999), the Telecommuni-

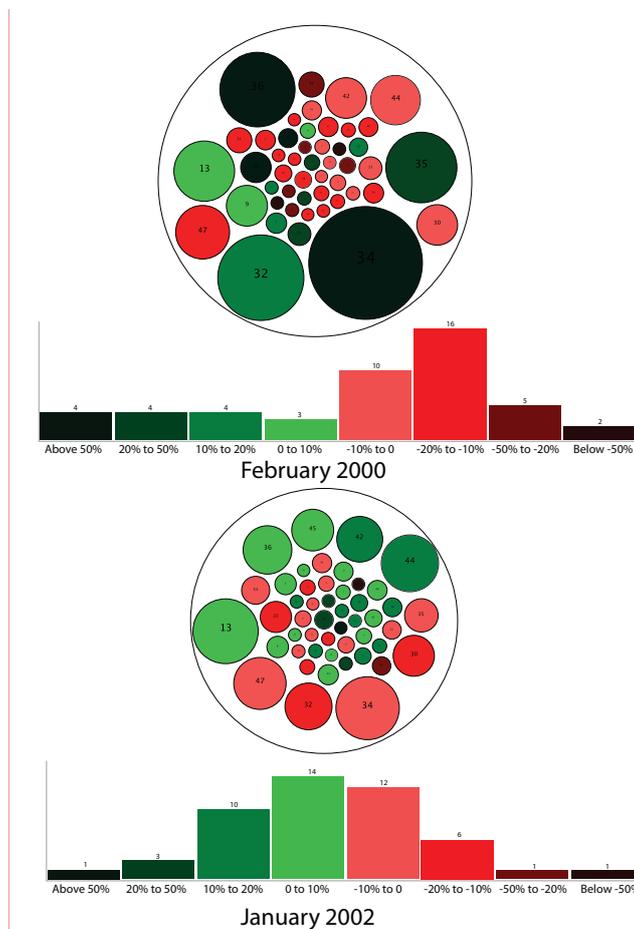


Figure 4.4: Histogram and Bubble Chart for the market, before and after the Dot Com crash of 2000

cation sector has around 60 new IPOs whereas only a handful of IPOs were seen in the Petroleum sector. However almost two years after the crash (March 2002), Telecommunications seems to be struggling to recover with almost 33% of its stocks performing poorly while Petroleum still seems to be pretty much steady with only 19% of its stocks in a similar condition. The dark red bar is definitely higher for Telecommunications which is a clear indication of how badly the sector is performing. The number of IPOs has become substantially lower for Telecommunications while Petroleum does not see a considerable change in that number.

Another interesting observation that was made was the early detection of bubbling sectors much before they peaked. Much before the dot com crash, when the strike chart is sorted, it can be seen which sectors have had a positive median value and for how long all arranged toward the right side of the chart. If the animation is now played, the sector order in this chart keeps changing while the strikes are still sorted. When sectors like Business Services (ID: 32) and Telecommunications (ID: 34) are selected in the Bubble View due to their large sizes, it can be seen that their bars in

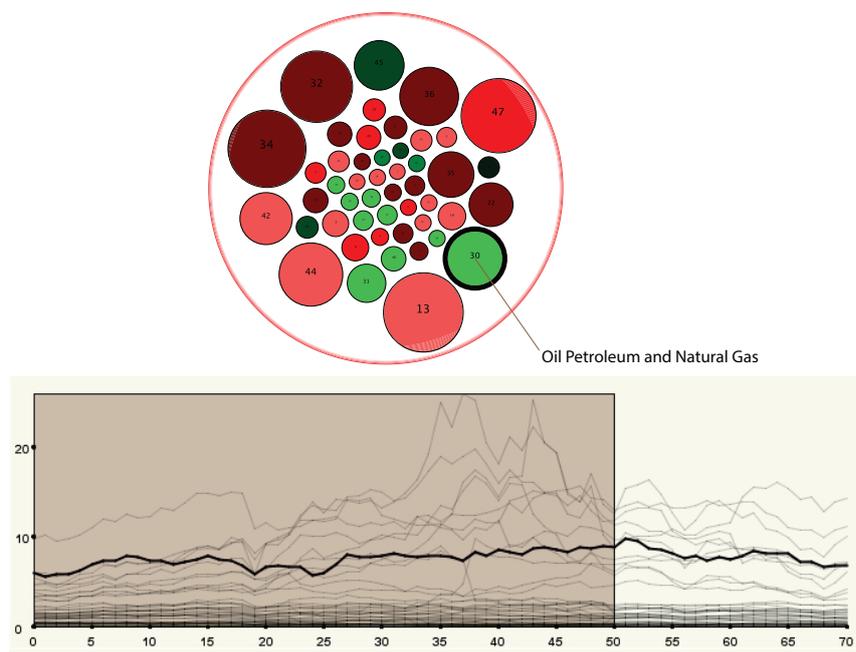


Figure 4.5: Line chart and Bubble chart highlighting the Petroleum sector for March 2001, which is one year after the dot com crash

the Median and Strike Chart, slowly keeps moving toward the right as the animation plays when the chart is sorted according to strikes. This means that they continually remain in the bubble condition and this is getting worse at time progresses. As can be seen in Figure 4.6, just before the burst, these two sectors are not only to the far right of the chart actually staying above the media for almost 28 months, followed closely by other technology based sectors like Computers (ID: 36) and Electronics (ID: 35) but the next strike just before them is much smaller at only 15 consecutive months above the median, meaning these technology sectors have been in a bubble condition far longer than the other sectors; clearly indicating their burst will cause the most devastating effects to market. This is also a point where investors should start pulling out their money lest they lose it in a crash that seems near. On the other hand, as the animation continues to pay after the crash, sectors like Petroleum are seen to move the right of the chart as technology sectors begin moving to the left and also go red as their median switches to negative values due to bad performances. Thus, even though the line for the Petroleum sector in the line graph seems to be relatively flat, from the Median and Strike chart it can be observed how it becomes a more profitable sector to invest in because the condition of the market has changed drastically after the crash.

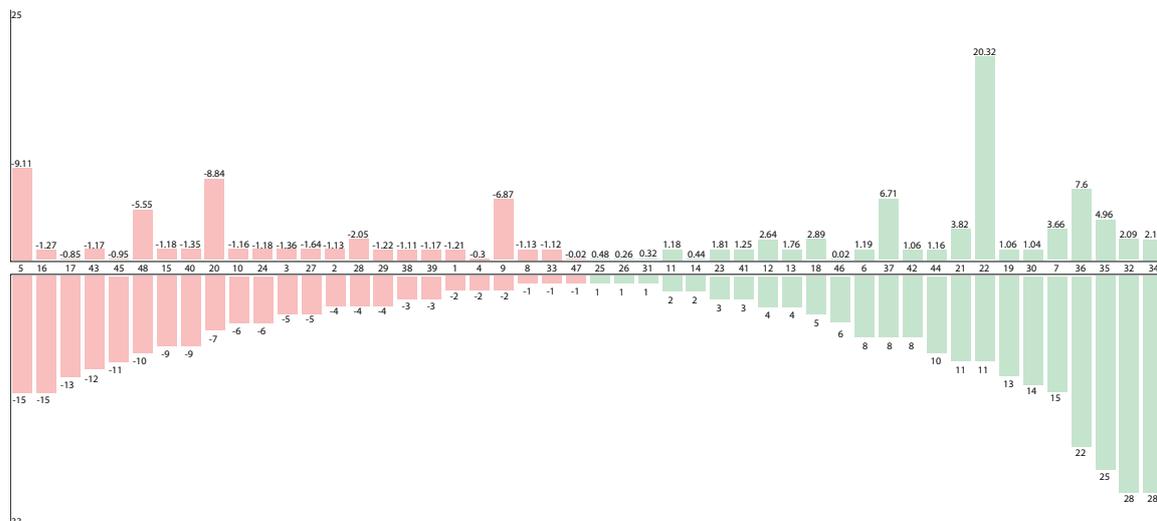


Figure 4.6: This shows the Median and Strike Chart for a date one month before the crash. The strikes have been sorted and sectors to the far right show those that have contained bubbles for long periods.

The 1998 Market Correction

The other significant market event which was expected to be shown by the visualization for this time period was the 1998 Stock Market Correction. It occurred in mid-1998 when financial crisis in Asia, spread to Russia and finally reached the United States when some hedge funds began to fail and threatened the stability of the entire international financial system as a whole. However banks in the United States were able to prevent very disastrous consequences. Even though this is not a full blown stock market bubble and crash, it was still an event that caused the loss of significant amount of investment all over the world.

The 1998 stock market crash is clearly visible in sectors that were closely related and contributing to it. As can be seen in Figure 4.7, in January of 1998 which is fairly close to the crash of August 1998, there are many sectors which seem to be doing extremely well compared to their performance in the previous year. In January of 1998 there seem to be very few sectors which exhibit such a negligible change in performance. By May of 1998, many of the noncontributing sectors seem to be performing badly and in June 1998 this number seems to have increased. However in both these periods, the sectors contributing to the crash continue to be performing extremely well compared to the others. By the end of August in 1998, the crash is very clearly visible wherein the sizes of the contributing sectors has reduced along with a significant drop in their performance. It is also interesting to note that since the contributing sectors are mostly finance based some of the smaller sectors also seem to be taking the hit even though they did not contribute to the crash. This is because many firms in those sectors were dependent on finance sectors for their capital and they too had to face the adverse effects of the correction.

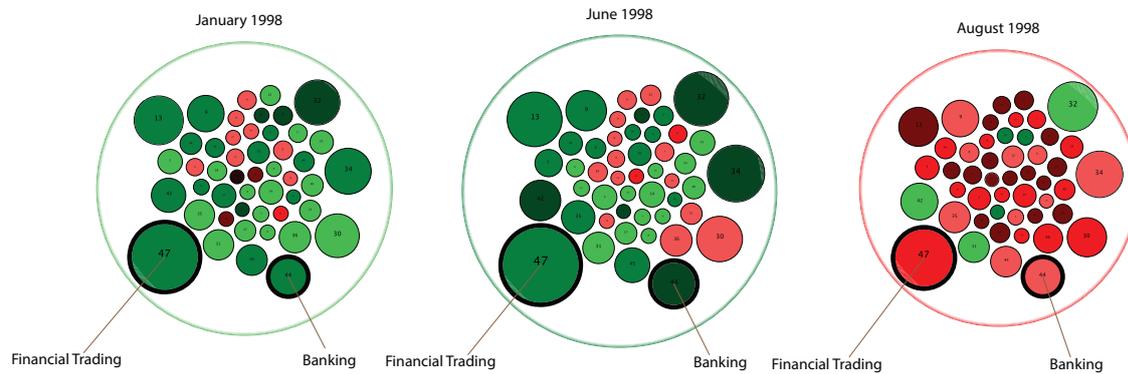


Figure 4.7: The 1998 Market correction occurred in August 1998 because of finance based sectors like Trading, Banking and Insurance

Even though the time frame of the 1998 correction is much shorter than the dot com boom, the strikes identify that these sectors are above the median for substantially longer than others, except those in the dot com bubble, during this period. It is this extended time above the median that clearly shows the sectors that are substantively above overall market performance, for both the long term and short term bubbles. This is indicated in the bar chart of strikes by the larger relative differences in strike counts compared to the other sectors in the market.

Chapter 5

Pilot Test

A small scale pilot test was conducted using graduate level students from Virginia Tech belonging to departments in the Virginia Tech College of Engineering and the Pamplin College of Business. The aim of this test was to verify the use of different views of our visualization system in performing different types of tasks in the study of market bubble formations and bursts. All visualizations showed the same data for 6 years from 1995 to 2001. They also only show the overview level containing the sectors of the market. Since the aim was to test the effectiveness of the visualizations for different visualization tasks and not to actually visualize a market bubble (which would require much knowledge and expertise in financial data analytics that generally only market regulators or auditors possess), individual stock visualizations were considered to cause cognitive overload and confuse the novice user. Data for 48 sectors of the market were shown to the participants. The 48 sectors were formed based on the Fama French [11] sector classification.

5.1 Design of the Test

The pilot test requested the participation of 12 graduate students who had never seen any of the visualizations prior to the test. Participants were divided into three groups of 4 participants each as Group 1, Group 2 and Group 3. There were 3 visualizations that needed to be tested for effectiveness in facilitating different types of tasks. Participants of Group 1 did tasks for Visualization 1 and then those for Visualization 3. Those in Group 2 did tasks for Visualization 2 and then those for Visualization 3. Group 3 participants did tasks only for Visualization 3. While viewing one visualization, participants are presented with tasks one at a time. They are presented with the next task once they answered the question to the previous task. The tasks are presented in a separate window from the visualization where they were to give their answers and move on to the next task. Questions and the tasks that they represent have been outlined in Table 5.1. At the end of completing all tasks, they were asked for their opinions regarding the visualizations using a survey with a 7-point Likert Scale. The three visualizations are detailed further.

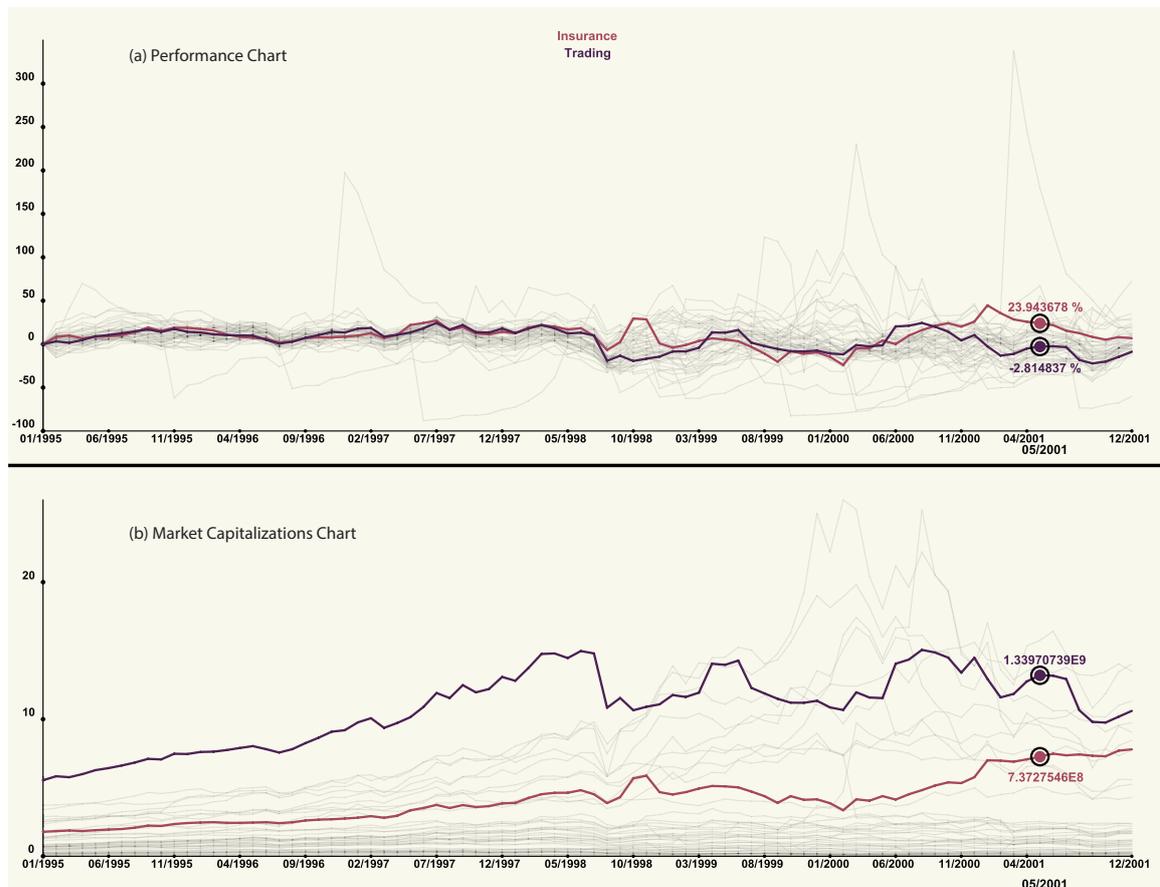


Figure 5.1: Line Charts showing 48 sectors of the Market with Insurance and Trading sectors highlighted (a) Performance Chart (b) Market Capitalization Chart

Visualization 1: Line Chart View This view shows two line charts for 48 sectors of the market placed one on top of the other. The bottom chart shows the market capitalization value for 48 sectors of the market for a period of 6 years. The top chart shows the change of market capitalization value using the 100 Percent method for the same 48 sectors of the market and for the same period of 6 years. Functionalities of this view included selection based on just clicking and highlighting a line. The X-position of the mouse pointer would highlight the point on the line that is selected to show the actual value of the Y-axis as shown in Figure 5.1. A search box was also provided so that searching for a specific sector line will not require manually searching through all the lines or remembering which lines represent which sector. A maximum of 2 sectors can be selected at a time in which case the lines are color coded to match the names of the sectors displayed at the top of the view. Users were also provided with a *clear selections* option to remove the highlighting of the selected sectors so they don't have to deselect each sector individually.

Visualization 2: Bubble and Strike Chart View This view shows the bubble chart at the top and the strike chart at the bottom of the view. The market capitalization values are represented by the

sizes of the circles in the bubble chart and the colors of the bubbles indicate the performance of the sectors using the 100 Percent calculation. In other words, the same data attributes which are shown in the Line Charts of Visualization 1 are shown in the bubble chart but as a snapshot of time. The strikes in the strike chart of this visualization however only show the strikes of the colors that the bubble chart is showing. The inclusion of the median and median-strike chart was omitted because this is processed data which the participant can obviously not see using the Line Charts. Instead the strikes in this chart simply show for how long before the current time in the visualization, the sector has shown the current performance trend. Thus a green bar of height 5 in the strike chart shows that the sector is showing positive performance currently and has been showing a positive performance for the last 4 time stamps. There are 48 glyphs in both charts, one for each sector in the market. Since this view is based on animation and navigating the time line, the date for which data is currently being seen is displayed at the very top of the view along with all the controls that the users have. Here as well, at most only 2 sectors can be selected at a time and they are highlighted on both charts at the same time. There is a *clear selections* option in this view as well, in addition to the search bar for easy selection of a specific sector. One additional function that users were given for this visualization was the sorting and un-sorting of the strike chart.

Visualization 3: Combined Multiple Coordinated View This view contains two windows where one contains the line charts and the second contains the bubble and strike chart as described above. The two views were synced to each other such that selection in any one view would highlight the same sector in the other view. *Clear selections* would be effective in both views. The views were provided to the participant side by side in the same screen so they don't have to distribute their attention among multiple screens. They were allowed to change the locations of the views according to their convenience but not allowed to change the dimensions of the views.

5.2 Hypothesis of the Study

Each participant was asked to perform the same tasks for every visualization that he or she viewed. We expected there to be a learning effect for participants in Groups 1 and 2, while they were viewing the second visualization (i.e. Visualization 3). This is because what they saw in their first visualization is repeated in what they view second with the addition of one more view synced to what they saw first. The tasks were designed and chosen such that some are difficult to perform using the line charts and some are difficult to perform using the bubble and strike chart. Thus the combined visualization with the multiple coordinated views should facilitate the performance of all the tasks. Though we expected that the users would take longer using the combined visualization, we expected them to take lesser time if they have already done a task using one of the visualizations because they are tuned to what they need to look for.

Table 5.1: Participant Tasks for Each Visualization

No.	Question	Task
1	What would have been the best time to sell an investment in the Telecommunication sector in the period between 01/1995 and 10/1998?	Locate
2	Are the Business Services and Telecommunication sectors correlated in the period between 01/1995 and 10/1998? (Yes: When one increases, the other decreases / No: Both increase or decrease together / Can't Say)	Correlate
3	When was the market at its lowest performance between 01/1995 and 10/1998?	Locate
4	At this time of lowest performance, which sector shows the sharpest performance drop?	Correlate
5	List 3 sectors that are increasing in performance the most during 07/1999 and 07/2000?	Cluster
6	List 3 sectors that are decreasing in performance the most during 07/1999 and 07/2000?	Cluster
7	How long will the Business Services sector continue to increase after the last time you see in the visualization before it peaks (1 month / 3 months / 6 months / 1 year / longer than 1 year)?	Predict
8	Which sector do you think will perform the worst in 1 year from the last time you see in the visualization?	Predict

5.3 Participant Task Results

All participant responses are given in Appendix B. Table 5.3 shows the average answers for tasks by all participants and how they should be analyzed. We analyze the answers in terms of distance from the right answer, whether the answer is correct or incorrect and how many answers were correct. Table 5.2 shows how long it took participants to perform the different tasks in seconds.

Locate Task 1: Locate Task 1 was to find the time in the data being visualized which would have been the best for selling an investment in a particular sector. The correct answer was a date when the investment showed the maximum market capitalization value. For the Telecommunications sector between 01/1995 and 10/1998 the correct answer was 10/1998. Participants of Group 2 who saw Visualization 2 first got the most wrong answers but when they did Visualization 3 right after, they got the most right answers. This may be because the addition of the line chart allowed them to easily pick the time when the market capitalization value was the highest. There was also a slight decline in more right answers when participants of Group 1 did Visualization 3 after viewing Visualization 1. Thus the line chart was more useful overall for this task.

However in many cases of Visualization 3 when the participant has answered with 07/1998, there was a peak in the value of the market capitalization which then fell down and started peaking again. Though this is not exactly the right answer since the peak of 10/1998 was higher than the peak of

Table 5.2: Time to Perform Tasks (in seconds)

Group 1		Group 2		Group 3
Visualization 1	Visualization 3	Visualization 2	Visualization 3	Visualization 3
155.282	137.32975	114.25	79.43875	103
87.9925	74.77975	62.151	94.57925	43.70625
130.41425	115.3245	45.1545	127.74425	49.71075
126.90875	74.8965	73.695	197.40825	40.4865
291.83375	176.25675	176.43225	245.28375	250.526
185.325	213.72475	149.0105	261.96425	194.49225
111.46075	120.78025	147.432	234.09675	92.38375
90.467	92.70825	75.80475	251.0605	73.9555

7/1998, the logic that the participants have applied to get to this answer seems to be sound. When considering the time taken to complete this task, Line Charts seem to have taken longer and the addition of a visualization has reduced the time as expected. Also, participants of Group 3 seem to have done best in terms of time taken to perform this task.

Correlate Task 1: This task was to identify the correlation between two sectors of the market. Through most of the time period which was to be observed, the two sectors seem to be uncorrelated, that is, both are increasing and decreasing at the same time. There are however points of intersection when one increases and one decreases. This is why the correct answer is *Can't Say*. However *No* would be the next best logical answer because through most of the time period under observation the sectors do move in the same direction. If only taking the correct answer into consideration, i.e. *Can't Say* the best results were seen in Group 1 when participants saw Visualization 1 and then saw Visualization 3. This is also the case when *No* and *Can't Say* are both considered correct; wherein participants of Group 1 have done best and did better when they saw Visualization 3 after Visualization 1. This result is attributed to the fact that when looking for correlations among changing variables through a span of time, it is better to have the values for the entire time period as in Visualization 1 than having to remember what one has seen before as is the case in Visualization 2.

Group 3 participants performed this task much faster than the other two groups as well. The time taken to perform this task has increased for participants of Groups 1 and 2 due to the addition of another visualization, perhaps because they now have to look at one more visualization for a task that involves observation over a period of time.

Locate Task 2: This task asked the participants to locate a time in the period between 01/1995 and 10/1998, when the entire market was at its lowest performance. The answer is 8/1998 when there was a correction in the market, and most sectors performed very poorly all of a sudden. Since the effect was drastic and seen across the market, we analyze results for this question as correct or incorrect. As can be seen in the Table 5.3, participants of Group 1 did the best in Visualization 1 and did similarly well in Visualization 3 as well. Participants of Group 2 did poorly for Visualization 2

but seemed to do better with the addition of Line charts in Visualization 3. Thus line charts seem to have been useful to find the time point when the entire market was performing poorly.

A general upward movement in performance by most sectors is seen as an anomaly at 8/1998 when they take a sudden dip and then continue to rise. This however can be confusing the Visualization 2 because there are 2 points when the entire market is seen to be shrunken and comparatively small in size. The addition of the line charts in Visualization 3 for participants of Group 2 seems to have increased their time to perform the task which is expected because they need to observe a new visualization to get to the correct answer now.

Table 5.3: Average of Responses

No	Evaluation Method	Target	Group 1		Group 2		Group 3
			V1	V3	V2	V3	v3
1	Distance from right answer (0)	Minimize	3.25	3.75	4.25	1	4.25
2	Correct(0) or Not Correct (1)	Minimize	1	0.5	0.5	1	1
3	Distance from right answer (0)	Minimize	0.5	0.5	1	0.5	0.75
4	Correct(0) or Not Correct(1)	Minimize	0.5	3.75	3.75	0.5	3.75
5	Count of right answers out of 3	Maximize	1.75	1.75	1	1	1.5
6	Count of right answers out of 3	Maximize	1	1.25	1.5	2	1
7	Distance from right answer (0)	Minimize	3	1.5	0.75	1	2.5
8	Correct(0) or Not Correct(1)	Minimize	1	0.5	0.75	0.25	0.75

Correlate Task 2: In this task, which follows the previous task, was to identify which sector was performing the worst when the entire market was suffering and at it's worst. The responses from participants show that those who were in Group 1 did the best followed by those in Group 3 who did Visualization 3. This means that the line chart was the most useful for identifying the correct sector. As can be seen in Figure 5.2, the Fabricated Products sector at this period, is easy to spot with the Line charts. Group 3 participants however were much faster at identifying the worst performing sector because they could easily select the interesting sector and observe it at the same time.

Cluster Task 1: In this task, the participants were asked to find 3 sectors which were showing performance improvements in the period between 7/1999 and 7/2000. As can be seen in the Table 5.3, performance increases seem to be more easily spotted by Visualization 1. Participants of

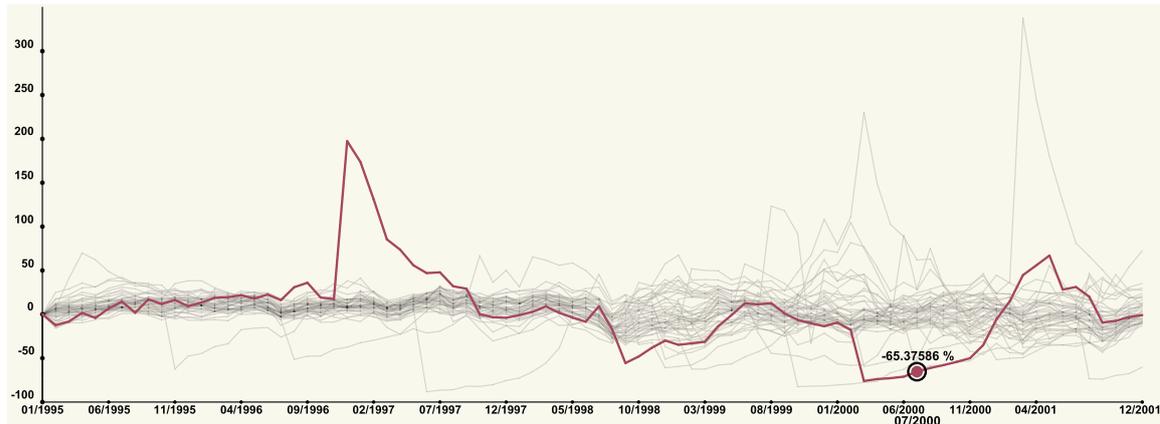


Figure 5.2: Line Charts showing the performance of Fabricated Goods at 07/2000

Group 1 who did Visualization 1 and then Visualization 3 seem have found most right answers followed by those of Group 3 who only did Visualization 3. This is because, this period includes the Dot Com Crash of 2000 when the few technology sectors in the market did exceedingly well and they were easy to spot in the Line Charts since they stand out from the rest of the sectors.

Cluster Task 2: In this task, the participants were asked to find sectors which performed poorly during the period between 7/1999 and 7/2000. As can be seen in the Table 5.3, participants of Group 2 had the best results when they used Visualization 2 and did even better when they used Visualization 3. Since this period includes the dot com crash of 2000, when most sectors in the market suffered the adverse effects of the failing economy in the market, Visualization 2 seemed to be easier to perform the cluster task. So most sectors were showing a decrease in performance during this period, but clustering was easier in Visualization 2 because it does not have the clutter that occurs in the Visualization 1 due to similar movements by most sector lines.

Predict Task 1: In this task, specific sector's peak was asked to be predicted and as can be seen in the Table 5.3, participants of Group 2 had the best results. They did best when viewing Visualization 2 and did slightly worse with Visualization 3 though still better than other groups. Those in Group 1 did poorly when viewing Visualization 1 but showed much improvement after viewing Visualization 3. Thus the addition of the Bubble and Strike Chart helped in the performance of this task. When asked for feedback from the users regarding this task, most in Group 2 answered that the sorting of strikes helped them make the prediction. At the last date they saw in the visualization Business Services has a small strike bar, which helped them say that it will probably take longer than 1 year for it to make another peak.

Predict Task 2: In this task, the participants were asked to find the sector which they thought would be performing the worst, 1 year in the future. Here again, participants of Group 2 seemed to be doing best when they saw Visualization 2 but did better when they were also given the line charts in Visualization 3. There is also an improvement in performance from Visualization 1 to Visualization 3 for those in Group 1. Thus, the addition of the bubble chart helped them in this

task. When asked about how they made the prediction for this task, those who were in Group 3 answered that they used the Bubble chart to select the sector of interest and used the line chart to see if it was showing an upward or downward trend. Those in Group 3 also showed good results which goes to show that both views are important for performing of this kind of task. This task took a significantly lesser time than the previous task overall because all the participants were more focused on what they were currently seeing rather than trying to visualize in their minds what would happen in the future as was the case in the previous task.

5.4 Participant Opinions on Visualizations

As mentioned before, all participants were given a survey to take with optional questions and were asked for their feedback. Out of 12, 7 participants took the survey about the Visualization 1, 10 took the survey for Visualization 2 and all of them took the one for the Visualization 3. They have indicated how much they agreed or disagreed to statements about the visualizations using a 7-point Likert Scale. 76% of the users agreed that both views were helpful in performing the tasks and 59% users found that the line chart was more useful for the tasks.

We hypothesized that participants would find *selection of sector* based tasks easier in the bubble and strike chart than the line chart. Responses from participants that support this hypothesis indicate that 72% found the selection of sectors to be difficult in the line chart visualization while 100% agreed that it was easy on the bubble and strike chart. 84% of the participants, after completing tasks for Visualization 3 have indicated that the bubble and strike chart is more *easy to use* than the line chart view.

Another hypothesis was that the depiction of 48 sectors of the market may be too cluttered for the line chart and not so in the bubble and strike chart. While only 60% of the participants agreed that the bubble and strike chart had too many bubbles, 100% of the users agreed that there were too many lines. Many of the participant feedback regarding what they did not like about the visualization mentioned that the line chart was too cluttered and close to each other. This is not the case in bubble charts because their representative glyphs are visually separate so even if they are cluttered, selecting individual bubbles is not difficult.

Chapter 6

Conclusion

6.1 Intellectual Merits

The ease with which complex market data can be used to study stock market bubbles over the years is the principal intellectual merit of this research. As mentioned before, the target (heavy) users for this tool are the market regulators who study trends of investment all over the market on a global scale. Furthermore, they may have access to confidential data for their unique jurisdiction which can be plugged into the tool to visualize potential bubbles that may already exist, see the beginning of formation or be close to bursting. Other than regulators, this tool is of interest to any number of financial scholars to visualize the intricate movement of capital through the different sectors of the market over time. The tracking feature especially is one that can be useful to even light users who wish to see the trends of a particular stock that they have a vested interest in. Another important merit of this research is in being able to study how markets achieved a state of stability after a bubble burst occurred and how long it took to arrive at this state by analyzing a bubble burst that happened in the past. This can be extremely useful in understanding consequences of a future bubble if present conditions of the market are similar to what happened before the previous bubble burst.

6.2 Broader Impact

The proposed research has a very broad impact in the world of finance and global market studies. Being able to analyze in real time, market trends that are indicators of a potential stock market bubbles can be used by market regulators in preventing such an occurrence. Market Regulators, whether in a government or non-government institution, have the responsibility and interest in keeping the market as stable as possible for protecting customers. Since many different visualizations are used currently, none of which have been designed specifically for the particular intent

of detecting formation of market bubbles, this tool can provide them with a novel application that contains the best of all the features they currently use while providing an intuitive metaphor and a small learning curve before which they can start making practical observations in the market with it. Since changes in the current market can be seen and regulators can use the tool to see trends over a range of period, they have the ability to study bubbles that have been latent for a long time but may be imminent in the near future. Projected data can be plugged into this tool to see if a bubble will occur and how it may affect the rest of the market. This can be especially useful when trends can go one of many ways and forecasts need to be made for each of these conditions. Early identification of bubbles can save the global economy billions of dollars to say the least. Investors, governments and corporations can make smarter investment decisions based on conclusions of these studies thereby being mutually beneficial.

6.3 Future Work

This thesis outlines the iterative process of the design of a visualization system that is based on animated, interactive views, that can help finance professionals with the analysis of stock market. The initial research direction was to create an animation of stock movements in the market with the aim of spotting abnormal entity behavior and then using this identification to further examine it closely. This led to the addressing of the shortcomings of using only animation and making it interactive so that the user has the capability to traverse the time line as desired. The strengths of other visualizations like line charts and bar graphs were further included to give the best possible tool for the analyst to verify hypotheses that may be formed after viewing the animation. Finally a second level of detail was provided with the individual stocks of selected sectors, to enable market regulators to be able to spot fraudulent activities in the market. Trading strategies are currently being developed with the aim of building an exploratory tool for aiding the user in making decisions regarding the buying and selling of stocks in specific sectors while providing comprehensive knowledge about the state of the market to make prudent investment decisions. These strategies will be extensively tested before being applied to the current market. Another formal user study is planned to be conducted with market professionals and regulators in order to test the effectiveness of the visualization in making trading decisions and detecting market bubbles. This study will focus on using participants who already have extensive knowledge in financial data analysis and use their expertise in validating and verifying the utility of this research.

Bibliography

- [1] Yahoo's finance charts: <http://finance.yahoo.com/>.
- [2] D. Abreu and M. K. Brunnermeier. Bubbles and crashes. *Econometrica*, 71(1):173–204, 2003.
- [3] W. alexander Jungmeister and D. Turo. Adapting treemaps to stock portfolio visualization. Technical report, 1992.
- [4] L. Bartram. Perceptual and interpretative properties of motion for information visualization. In *Proceedings of the 1997 Workshop on New Paradigms in Information Visualization and Manipulation*, NPV '97, pages 3–7, New York, NY, USA, 1997. ACM.
- [5] D. Brodbeck, M. Chalmers, A. Lunzer, and P. Cotture. Domesticating bead: adapting an information visualization system to a financial institution. In *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pages 73–80, Oct 1997.
- [6] I. Cooper. Arithmetic versus geometric mean estimators: Setting discount rates for capital budgeting. *European Financial Management*, 2(2):157–167, 1996.
- [7] C. Csallner, M. Handte, O. Lehmann, and J. Stasko. Fundexplorer: supporting the diversification of mutual fund portfolios using context treemaps. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pages 203–208, Oct 2003.
- [8] H. T. Dao, A. L. Bazinet, R. Berthier, and B. Shneiderman. Nasdaq velocity and forces: An interactive visualization of activity and change. *J. UCS*, 14(9):1391–1410, 2008.
- [9] T. Dwyer and P. Eades. Visualising a fund manager flow graph with columns and worms. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 147–152, 2002.
- [10] B. Ely. Financial regulation. <http://www.econlib.org/library/Enc/FinancialRegulation.html>, 2008.
- [11] E. F. Fama and K. R. French. Industry costs of equity. *Journal of financial economics*, 43(2):153–193, 1997.

- [12] J. Guerra-Gomez, M. Pack, C. Plaisant, and B. Shneiderman. Visualizing change over time using dynamic hierarchies: Treeversity2 and the stemview. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2566–2575, Dec 2013.
- [13] G. W. Hartmann. Gestalt psychology: A survey of facts and principles. 1935.
- [14] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09*, pages 1303–1312, New York, NY, USA, 2009. ACM.
- [15] M. L. Huang, J. Liang, and Q. V. Nguyen. A visualization approach for frauds detection in financial market. In *Information Visualisation, 2009 13th International Conference*, pages 197–202, 2009.
- [16] S. Huron, R. Vuillemot, and J.-D. Fekete. Visual sedimentation. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2446–2455, Dec 2013.
- [17] K. imuni. Visualization of stock market charts. In *In Proceedings from the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2003 (2003), Plzen-Bory (CZ), 2003*, 2003.
- [18] B. Johnson and B. Shneiderman. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In *Visualization, 1991. Visualization '91, Proceedings., IEEE Conference on*, pages 284–291, Oct 1991.
- [19] D. A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, Jan. 2002.
- [20] A. Komlodi, P. Rheingans, U. Ayachit, J. R. Goodall, and A. Joshi. A user-centered look at glyph-based security visualization. In *In Proceedings of the 2005 Workshop on Visualization for Computer Security*, pages 21–28, 2005.
- [21] C. S. Merino, M. Sips, D. A. Keim, C. Panse, and R. Spence. Task-at-hand interface for change detection in stock market data. In *Proceedings of the working conference on Advanced visual interfaces, AVI '06*, pages 420–427, New York, NY, USA, 2006. ACM.
- [22] K. Nesbitt and S. Barrass. Finding trading patterns in stock market data. *Computer Graphics and Applications, IEEE*, 24(5):45–55, 2004.
- [23] C. North, U. Farooq, and D. Akhter. Datawear: Revealing trends of dynamic data in visualizations, 2001.
- [24] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1325–1332, Nov 2008.

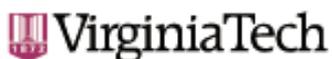
- [25] H. Rosling. Gapminder. *GapMinder Foundation* <http://www.gapminder.org>, 2009.
- [26] B. L. Ruth, L. R. Bartram, and S. F. University. Enhancing information visualization with motion, 2001.
- [27] T. Schreck, T. Tekušová, J. Kohlhammer, and D. Fellner. Trajectory-based visual analysis of large financial time series data. *SIGKDD Explor. Newsl.*, 9(2):30–37, Dec. 2007.
- [28] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343, 1996.
- [29] B. Shneiderman and M. Wattenberg. Ordered treemap layouts. In *Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on*, pages 73–78, 2001.
- [30] A. S. Spritzer and C. M. D. S. Freitas. A visual tool to support technical analysis of stock market data. In *Proceedings of the working conference on Advanced visual interfaces, AVI '06*, pages 512–515, New York, NY, USA, 2006. ACM.
- [31] J. Stasko and E. Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pages 57–65, 2000.
- [32] J. Steele and N. Iliinsky. *Beautiful visualization*. O'Reilly Media, Inc., 2010.
- [33] B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: Can it facilitate? *Int. J. Hum.-Comput. Stud.*, 57(4):247–262, Oct. 2002.
- [34] W. Wang, H. Wang, G. Dai, and H. Wang. Visualization of large hierarchical data by circle packing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06*, pages 517–520, New York, NY, USA, 2006. ACM.
- [35] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '00*, pages 110–119, New York, NY, USA, 2000. ACM.
- [36] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.
- [37] M. Wattenberg. Visualizing the stock market. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems, CHI EA '99*, pages 188–189, New York, NY, USA, 1999. ACM.
- [38] W. Wright. Research report: information animation applications in the capital markets. In *Information Visualization, 1995. Proceedings.*, pages 19–25, 1995.

- [39] W. Wright. Research report: information animation applications in the capital markets. In N. D. Gershon and S. G. Eick, editors, *INFOVIS*, pages 19–25. IEEE Computer Society, 1995.
- [40] W. E. Young and R. H. Trent. Geometric mean approximations of individual security and portfolio performance. *Journal of Financial and Quantitative Analysis*, 4:179–199, 6 1969.

Appendix A

Formative Study

This study was done with graduate level students from the Pamplin College of Business at Virginia Tech. This study was approved by the IRB at Virginia Tech. This appendix gives the IRB approval letter for this study, the recruitment document, the experimenter instructions, participant questionnaire and data that was collected from the participants.

**Office of Research Compliance**

Institutional Review Board
North End Center, Suite 4120, Virginia Tech
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-4606 Fax 540/231-0959
email irb@vt.edu
website <http://www.irb.vt.edu>

MEMORANDUM

DATE: November 27, 2013
TO: Steve Sheetz, Sruthi Ganesan Iyer
FROM: Virginia Tech Institutional Review Board (FWA00000572, expires April 25, 2018)
PROTOCOL TITLE: Stock Market Bubble
IRB NUMBER: 13-1089

Effective November 26, 2013, the Virginia Tech Institutional Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

<http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

PROTOCOL INFORMATION:

Approved As: **Expedited, under 45 CFR 46.110 category(ies) 7**
Protocol Approval Date: **November 26, 2013**
Protocol Expiration Date: **November 25, 2014**
Continuing Review Due Date*: **November 11, 2014**

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Invent the Future

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

Visualization 1 questions.

Please answer each of the questions in relation to visualization 1.

What month did the bubble burst? (enter a month index number) _____

For each of the following questions, please respond using a scale of strongly disagree, disagree, slightly disagree, neither agree or disagree, slightly agree, agree, strongly agree by placing an X in the box indicating your response.

	strongly disagree	disagree	slightly disagree	neither	slightly agree	agree	strongly agree
The size of the circles was useful to me for identifying when the bubble burst.							
The color of the circles was useful to me for identifying when the bubble burst.							
The frequency of color changes was useful to me for identifying when the bubble burst.							
The frequency of size changes was useful to me for identifying when the bubble burst.							
The visualization was easy to understand.							
There were too many stocks represented (circles) in the visualization.							
Overall I was able to identify the bubble burst with this visualization.							

Can you think of any additional information that would be useful for identifying bubbles for this visualization?

Visualization 2 questions.

Please answer each of the questions in relation to visualization 2.

What month did the bubble burst? (enter a month index number) _____

For each of the following questions, please respond using a scale of strongly disagree, disagree, slightly disagree, neither agree or disagree, slightly agree, agree, strongly agree by placing an X in the box indicating your response.

	strongly disagree	disagree	slightly disagree	neither	slightly agree	agree	strongly agree
The size of the circles was useful to me for identifying when the bubble burst.							
The color of the circles was useful to me for identifying when the bubble burst.							
The frequency of color changes was useful to me for identifying when the bubble burst.							
The frequency of size changes was useful to me for identifying when the bubble burst.							
The intensity of the colors was useful to me for identifying when the bubble burst.							
The frequency of changes in the intensity of the colors was useful to me for identifying when the bubble burst.							
The visualization was easy to understand.							
There were too many stocks represented (circles) in the visualization.							
Overall I was able to identify the bubble burst with this visualization.							

Can you think of any additional information that would be useful for identifying bubbles for this visualization?

Final Selection and Demographics

Do you prefer visualization 1 or visualization 2? 1 2 Don't know.

	strongly disagree	disagree	slightly disagree	neither	slightly agree	agree	strongly agree
I prefer my selection because the size of the circles was more useful in it versus the other visualization.							
I prefer my selection because the color of the circles was more useful in it versus the other visualization.							
I prefer my selection because the frequency of color changes of the circles was more useful in it versus the other visualization.							
I prefer my selection because the frequency of size changes of the circles was more useful in it versus the other visualization.							
I prefer my selection because it was easier to understand.							
Overall I was able to identify the bubble burst better with this visualization.							

Why do prefer your selection?

What is your age? _____

Please circle your gender: female male

What is your major? _____

What is your class standing (circle one): Freshman Sophomore Junior Senior Graduate

Experimenter Instructions for Stock Market Bubble Visualization

This document describes the activities of the experimenter during the experiment. Including selection of the order of the visualizations, instructions to participants, demonstrating the operation of the program, providing the questions to participants and thanking them for participating.

Step 1: Before the participant arrives determine which visualization the participant will see first and second. If the last participant saw the GMR visualization first, the next participant should see the 100% visualization first.

Select the next numbered consent form for the participant. Write the participant number on the two visualization question sheets and the final question sheet to be used for this participant.

Step 2: Ask the participant to complete the informed consent form.

“Thank you for participating in our study of stock market bubble visualizations. Before we start please review the informed consent form, which has been approved by the VT Internal Review Board for human subjects protection. If you have any questions please feel free to ask me. “

Give the participant the consent form with participant number already completed. Then get the signed consent form from the participant and place it in the folder of consent forms.

“Thank you. You will be seeing two visualizations during the experiment. In both visualizations you will be asked to identify the month index in which the bubble bursts.”

Step 3: Read the instructions for the first visualization to the participant. Use the instruction pages for the appropriate visualization.

Step 4: Demonstrate the first visualization for the participant.

Step 5: Ask the participant to use the visualization to identify the stock market bubble.

Step 6: When the participant indicates they have identified the month index for the bubble, record it on the questions page and press the ESC button on the key board to close the program.

Step 7: Give the participant the questions page and ask them to answer the question for the first visualization

Step 8: Read the instructions for the second visualization (if they did GMR first they will then do 100%) to the participant.

Step 9: Explain that the second visualization is operated in the same manner as the first. Ask if the participant would like another demonstration of how the visualization works. If requested demonstrate the second visualization for the participant.

Step 10: Ask the participant to use the second visualization to identify the stock market bubble.

Step 11: When the participant indicates they have identified the month index for the bubble, record it on the questions page and press the ESC button to close the program.

Step 12: Give the participant the questions page and ask them to answer the question for the second visualization.

Step 13: Give the participant the final questions page and ask them to answer the questions comparing the visualizations.

Step 14: Thank the participant for their participation and inform them that the drawing will be held next week and that they will be contacted via email if they are successful in the drawing.

“Thank you for participating in our study. You are now entered in the drawing for a \$20 Wal Mart gift card. The drawing will be held next week. Dr. Sheetz will contact you via email if you are successful in the drawing and you will need to meet with him to receive it and sign a receipt.”

Step 15: Rename the file “100Percent_Key_Press_Data” or “GMR_Key_Press_Data” to start with the participant number, e.g., “P01_100Percent_Key_Press_Data.”

Instructions for GMR visualization.

“Each circle represents stock of one company. Size of the circle represents the market capitalization for the stock in the month of the index. Remember that market capitalization is the price of the stock times the number of outstanding shares held by investors. This value changes every month, so some circles get larger and others smaller depending on the value of the market capitalization of the stock in that month.

Colors of the circles indicate the rate of change of the market capitalization from the first index month shown in the visualization. Red tones indicate the stock is rising and is in a bubble condition. Deeper shades of red indicate a larger rise in the market capitalization and more intense bubble in the stock. Green tones indicate the stock the stock is falling in market value.

Do you have any questions?” Answer any questions about size and colors from the participant.

IF THIS THE FIRST VISUALIZATION THE PARTICIPANT IS SEEING DO THE FOLLOWING

“To operate the visualization press the “play” button (point to it on the screen) on the lower right corner of the screen, you can stop the visualization by pressing the “pause” button shown when the visualization is operating. Or you can stop and start the visualization using the space bar.”

Demonstrate this for the participant.

“You can also use the slider bar at the bottom of the screen to select where to start the visualization and change the periods displayed.”

Demonstrate this for the participant.

“Lastly, when the visualization is paused, you can use the up arrow to step to the next index month and the down arrow to see the previous index month.”

Demonstrate this for the participant.

“Do you have any questions about operating the visualization?” Answer any questions of the participant and demonstrate features again as needed.

After demonstrating you must close the program by pressing the ESC key and restart it for use by the participant. This ensures that the timing will be correct for the participant.

IF THIS THE SECOND VISUALIZATION THE PARTICIPANT IS SEEING DO THE FOLLOWING

“This visualization operates in the same way as the first visualization you completed. Would you like me to demonstrate any of the operations again?” Answer any questions of the participant.

“It is now time for you to use the visualization to identify the stock market bubble burst in these data. Please begin now. Indicate when the bubble bursts by identifying the month index with the highest market capitalization”

Instructions for 100% visualization.

“Each circle represents stock of one company. Size of the circle represents the market capitalization, which is the price of the stock times the number of outstanding shares held by investors, for the stock in the month of the index. This value changes every month, so some circles get larger and others smaller depending on the value of the price and outstanding shares.

Colors of the circles indicate the rate of change of the market capitalization from the first index month shown in the visualization using a 36 month moving average. Red tones indicate the market capitalization of the stock is falling. Deeper shades of red indicate a larger fall in the market capitalization. Green tones indicate the market capitalization of the stock is rising. Deeper shades of green indicate a larger rise in the market capitalization. Blue circles indicate that the market capitalization has not changed significantly for the stock in for the last period. Gray circles indicate stocks that were initially offered in this month and therefore do not have enough data to calculate the 36 month moving average.

Do you have any questions? “

IF THIS THE FIRST VISUALIZATION THE PARTICIPANT IS SEEING DO THE FOLLOWING

“To operate the visualization press the “play” button (point to it on the screen) on the lower right corner of the screen, you can stop the visualization by pressing the “pause” button shown when the visualization is operating. Or you can stop and start the visualization using the space bar.”

Demonstrate this for the participant.

“You can also use the slider bar at the bottom of the screen to select where to start the visualization and change the periods displayed.”

Demonstrate this for the participant.

“Lastly, when the visualization is paused, you can use the up arrow to step to the next index month and the down arrow to see the previous index month.”

Demonstrate this for the participant.

“Do you have any questions about operating the visualization?” Answer any questions of the participant and demonstrate features again as needed.

IF THIS THE SECOND VISUALIZATION THE PARTICIPANT IS SEEING DO THE FOLLOWING

“This visualization operates in the same way as the first visualization you completed. Would you like me to demonstrate any of the operations again?” Answer any questions of the participant.

“It is now time for you to use the visualization to identify the stock market bubble burst in these data. Please begin now. Indicate when the bubble bursts by identifying the month index with the highest market capitalization”

Participant	100pct_Q1	100pct_Q2	100pct_Q3	100pct_Q4	100pct_Q7	100pct_Q8	100pct_Q9
P1	7	7	5	4	6	2	6
P2	7	6	5	5	5	2	6
P3	6	4	5	6	6	2	4
P4	6	7	6	6	6	5	5
P5	6	6	7	6	6	5	5
P6	6	6	5	6	4	3	5
P7	7	7	6	6	7	2	6
P8	7	6	6	5	7	6	6
P9	6	6	6	7	6	4	6
P10	6	5	4	6	6	2	6
P11	7	5	5	7	7	1	7
P12	6	7	4	6	7	1	7
P13	7	6	7	4	6	4	6
P14	6	3	3	6	6	6	5
P15	7	7	5	5	7	6	5
P16	6	6	6	6	6	2	5
P17	6	7	4	6	7	3	6
P18	7	5	2	3	6	2	6
P19	6	5	3	6	6	2	6

Participant	First_Viz	GMR_Guess	GMR_Correct	GMR_Distance	100pct_Guess	100pct_Correct	100pct_Distance
P1	100pct	565	560	5	789	786	3
P2	GMR	562	560	2	792	786	6
P3	100pct	565	560	5	801	786	15
P4	GMR	576	560	16	816	786	30
P5	100pct	566	560	6	806	786	20
P6	GMR	561	560	1	792	786	6
P7	100pct	561	560	1	786	786	0
P8	GMR	561	560	1	787	786	1
P9	100pct	560	560	0	786	786	0
P10	GMR	560	560	0	786	786	0
P11	100pct	565	560	5	786	786	0
P12	GMR	565	560	5	786	786	0
P13	100pct	565	560	5	786	786	0
P14	GMR	560	560	0	786	786	0
P15	100pct	560	560	0	786	786	0
P16	GMR	560	560	0	786	786	0
P17	100pct	560	560	0	788	786	2
P18	GMR	560	560	0	791	786	5
P19	100pct	561	560	1	786	786	0

Participant	GMR_Q1	GMR_Q2	GMR_Q3	GMR_Q4	GMR_Q5	GMR_Q6	GMR_Q7	GMR_Q8	GMR_Q9
P1	7	1	1	6	4	4	5	2	5
P2	7	5	5	6	4	2	6	1	6
P3	6	6	6	6	6	6	6	2	4
P4	7	6	6	6	5	6	5	4	5
P5	7	5	3	6	3	3	6	2	5
P6	6	5	4	6	5	4	6	2	5
P7	7	3	3	6	3	2	6	1	6
P8	6	5	6	6	5	6	5	5	6
P9	6	7	4	5	3	3	3	4	5
P10	6	4	4	6	5	4	6	2	6
P11	7	2	2	7	2	2	6	1	7
P12	6	7	4	6	4	6	7	1	7
P13	7	4	6	4	7	2	6	2	7
P14	5	3	3	6	3	3	6	6	5
P15	7	7	5	5	3	3	7	2	7
P16	6	7	4	6	4	5	7	2	5
P17	6	5	5	6	6	5	4	4	6
P18	7	5	2	2	5	2	6	2	6
P19	6	2	5	5	2	2	6	2	6

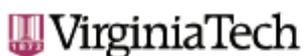
Participant	Overall Preference Order	Overall Preference	Overall_Q 1	Overall_Q 2	Overall_Q 3	Overall_Q 4	Overall_Q 7	Overall_Q 8
P1	1	100pct	2	7	6	4	6	6
P2	1	GMR	7	6	3	6	6	6
P3	2	GMR	2	6	6	6	6	4
P4	2	100pct	4	7	4	4	6	6
P5	2	GMR	7	5	6	5	6	6
P6	2	100pct	4	6	3	4	4	3
P7	1	100pct	4	7	7	5	5	6
P8	2	100pct	6	6	7	6	6	6
P9	1	100pct	4	6	5	4	4	5
P10	2	100pct	2	6	6	5	4	6
P11	1	100pct	1	7	1	1	1	1
P12	2	100pct	4	7	4	7	7	4
P13	1	100pct	5	2	5	6	6	6
P14								
P15	2	GMR	7	5	3	5	5	7
P16	2	100pct	4	7	7	4	4	5
P17	1	100pct	4	6	6	6	6	6
P18	2	100pct	4	6	4	4	4	5
P19	1	100pct	3	6	4	4	2	6

Participant	Age	Gender	Major	Class_Stan- ding	GMR_Play_Pause	GMR_Up- _keystrokes	GMR_Down- _keystrokes	GMR_Millise- conds	Hundpct- Play_Pause	Hundpct- Up_keystr- okes	Overall_P- reference- _Order	Overall_P- reference
P1	28	M		5	1	9	29	54699	1	21	1	100pct
P2					0	85	84	164973	1	81	1	GMR
P3	40	F	Acct	5	0	0	0	50924	0	0	2	GMR
P4	34	M	Acct	5	0	79	90	202432	0	145	2	100pct
P5	30	M	Acct	5	1	23	35	67505	1	102	2	GMR
P6	33	M	Acct	5	1	75	83	118439	0	185	2	100pct
P7	37	F	Acct	5	0	0	0	98630	0	0	1	100pct
P8	23	M	ISE	4	0	0	0	110321	0	0	2	100pct
P9	19	M	BIT	2	0	188	156	144005	8	17	1	100pct
P10	20	F	GEN	2	0	47	28	115145	0	44	2	100pct
P11	34	M	Acct	5	1	12	14	59936	1	18	1	100pct
P12	30	M	Acct	5	1	42	44	102994	1	19	2	100pct
P13	27	M	Finance	5	23	29	28	139565	19	57	1	100pct
P14	28	M	Finance	5	2	31	24	90496	1	30		
P15	40	F	Finance	5	0	0	0	65653	1	0	2	GMR
P16	34	F	Acct	5	3	17	11	77173	1	25	2	100pct
P17	32	M	Acct	5	2	15	18	88490	4	5	1	100pct
P18	37	F	Acct	5	0	26	48	92729	0	36	2	100pct
P19	35	M	Finance	5	8	2	32	53246	2	0	1	100pct

Appendix B

Pilot Test

This study was done with graduate level students from Virginia Tech. This study was approved by the IRB at Virginia Tech. This appendix gives the IRB approval letter for this study, the recruitment document, the experimenter instructions, participant questionnaire and data that was collected from the participants.



Office of Research Compliance
 Institutional Review Board
 North End Center, Suite 4120, Virginia Tech
 300 Turner Street NW
 Blacksburg, Virginia 24061
 540/231-4606 Fax 540/231-0959
 email irb@vt.edu
 website <http://www.irb.vt.edu>

MEMORANDUM

DATE: April 15, 2014
TO: Yong Cao, Steve Sheetz, Sruthi Ganesan Iyer
FROM: Virginia Tech Institutional Review Board (FWA00000572, expires April 25, 2018)
PROTOCOL TITLE: Design Study for Stock Market Bubble Visualization
IRB NUMBER: 14-438

Effective April 15, 2014, the Virginia Tech Institutional Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

<http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

PROTOCOL INFORMATION:

Approved As: **Expedited, under 45 CFR 46.110 category(ies) 5,7**
 Protocol Approval Date: **April 15, 2014**
 Protocol Expiration Date: **April 14, 2015**
 Continuing Review Due Date*: **March 31, 2015**

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
An equal opportunity, affirmative action institution

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

Questions about the Visualization

- (1/8) What would have been the best time to sell an investment in the Telecommunication sector in the period between 01/1995 and 10/1998?
- (2/8) Are the Business Services and Telecommunication sectors correlated in the period between 01/1995 and 10/1998? (Yes: When one increases, the other decreases / No: Both increase or decrease together / Can't Say)
- (3/8) When was the market at its lowest performance between 01/1995 and 10/1998?
- (4/8) At this time of lowest performance, which sector shows the sharpest performance drop?
- (5/8) List 3 sectors that are increasing in performance the most during 07/1999 and 07/2000?
- (6/8) List 3 sectors that are decreasing in performance the most during 07/1999 and 07/2000?
- (7/8) How long will the Business Services sector continue to increase after the last time you see in the visualization before it peaks (1 month / 3 months / 6 months / 1 year / longer than 1 year)?
- (8/8) Which sector do you think will perform the worst in 1 year from the last time you see in the visualization?

For Line Chart: Select one option for each of the following questions

It was easy for me to select the sectors I wanted to observe

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

It was easy for me to identify the fastest growing sectors

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The viz was useful for identifying the fastest growing sectors

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The viz was useful for predicting decreases in sector market capital

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

There were too many lines in the visualization

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

For Bubble Chart: Select one option for each of the following questions

It was easy for me to select the sectors I wanted to observe

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

It was easy for me to identify the fastest growing sectors

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The viz was useful for identifying the fastest growing sectors

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The viz was useful for predicting decreases in sector market capital

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

There were too many bubbles in the visualization

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

For Combined Vis: Select one option for each of the following questions

It was easy for me to select the sectors I wanted to observe

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

It was easy for me to identify the fastest growing sectors

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The viz was useful for identifying the fastest growing sectors

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The viz was useful for predicting decreases in sector market capital

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The addition of one more view helped in performing tasks

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

Overall, the line chart was more useful than the bubble viz

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

The bubble viz was easier to use than the line chart

Strongly disagree Disagree Somewhat disagree Neutral Somewhat agree Agree Strongly agree

Optional Questions

What I liked most about the visualization:

What I did not like most about the visualization:

Additional comments/suggestions:

Description Given to All Participants

Market Bubble Visualization Study

This is a study to test the effectiveness of different kinds of visualizations that help in analyzing and studying stock market data for bubble formations and bursts.

A stock market bubble is a phenomenon that occurs when the value of a stock or sector goes much higher than its true value because of investor excitement in a certain kind of stock. In such a situation investors continue to invest in a specific sector expecting very high returns. When they realize that the expected returns cannot be met (may be due to bad performance or the company/sector failing), a massive sell off occurs and this is when a bubble bursts.

Some investors know that this is about to happen but still invest in the bubbling sector thereby contributing to the bubble growth because they know that selling the stock just before the bubble bursts, right at its peak will allow them to make a huge profit.

We thus analyze 2 values while studying stock market bubbles, 1. Is the market capitalization value of the stock which is price times shares outstanding and the performance of the stock which is the comparison of the stocks market cap value against a one year moving average of its own market cap value. Thus performance shows how the market cap has changed for a particular stock in the last one year.

Both the visualizations that you will see depict data for 48 distinct sectors of the market.

Line Charts

In the line chart here, you see the top chart is one that maps performance of the sectors against time. The data is monthly because daily data contains a lot of noise. The bottom chart shows the actual market cap value of the sectors against time.

Interactions with this visualization include:

- Selection and de-selection of up to 2 sector lines
- Observe how the name of sector appears on the top of the chart
- Clear selections button
- Search button to search for a sector by its name

Bubble and Strike Chart:

This is an animated interactive visualization where you see snapshots of the market data in time and are allowed to move across time using a time slider. There is a bubble chart on the top and a strike chart in the bottom.

The bubble chart has 48 sectors where their sizes indicate their market cap values and their colors show their performance in one year with a moving average. Greens show positive performance with darker shades showing higher performance than lighter shades and reds show negative performance where darker shades show lower performance than lighter shades.

The strike chart will show long before the current time on the time slider, the sector has remained in the current performance trend. So a green bar of height 5 will show that the sector is green now and has been green for the last 4 time stamps.

You can interact with this visualization by using the up and down arrow keys or by just dragging the time slider to any point in the visualization. You can select any bar or any bubble to highlight that sector and look at details about that sector. There is also a search bar and clear selections options.

Combined Vis:

In this visualization there is one more view added from the previous visualization that you just saw.

Participant IdD	1	2	3	4
P1_G3V3	Jul-98 23.538	no 2.985	Sep-97 64.828	29 7.183
P2_G2V2	Oct-98 43.121	Can't say 10.433	Oct-98 30.534	14 74.318
P2_G2V3	Sep-98 21.895	yes 282.939	Oct-98 343.922	20 454.197
P3_G1V1	Mar-98 107.017	Yes 83.486	Aug-98 128.801	34 180.634
P3_G1V3	Jul-98 142.686	Yes 52.986	Aug-98 65.182	19 116.814
P4_G3V3	Jun-98 157.367	Yes 45.126	Sep-98 34.132	27 50.422
P5_G3V3	Mar-98 135.344	cantsay 127.335	Oct-98 163.391	29 77.35
P6_G2V2	Oct-95 6.98	Can't Say 5.067	Jan-95 6.099	28 20.459
P6_G2V3	Oct-98 75.553	Yes 24.714	Jan-95 13.883	28 33.812
P7_G3V3	Jul-98 99	Yes 102	Aug-98 86	20 70.529
P8_G1V1	Jul-98 107.426	No 69.652	May-98 92.27	29 23.166
P8_G1V3	Jun-98 61	No 30	Aug-98 172	13 123
P9_G2V2	Jul-98 196.637	No 172.553	Sep-98 72.985	20 41.334
P9_G2V3	Oct-98 103.307	No 35.664	Aug-98 82.172	27 125.624
P10_G1V1	Jul-98 34.495	No 54.495	Aug-98 115.95	20 260
P10_G1V3	Jun-98 128.633	cant say 77.133	Jul-98 41.116	1 29.886
P11_G2V2	Jul-98	yes	Jan-95	27

	212.956	60.551	71	158.669
P11_G2V3	Jul-98	yes	Aug-98	20
	117	35	71	176
P12_G1V1	Oct-98	Yes	Sep-98	20
	372.19	144.337	184.636	43.835
P12_G1V3	Jun-98	cant say	Jul-98	1
	217	139	183	29.886

5	6	7	8
22 36 11	20 5 9	longer than 1 year	9
22.368	218.593	20.361	61.937
35 36 37	25 40 32	longer than 1 year	18
55.867	211.837	36.1	72.283
34 32 36	5 48 20	longer than 1 year	39
654.972	750.485	800.575	840.033
36 22 34	9 44 18	3 month	9
432.687	369.844	85.734	174.834
34 36 32	9 5 48	3months	36
120.197	256.855	104.261	100.947
36 32 34	31 23 40	6months	11
318.162	311.097	199.654	81.153
21 7 35	5 18 22	1month	1
317.302	247.973	63.526	88.055
32 34 36	20 18 5	longer than 1 year	39
144.969	80.033	162.369	18.049
36 7 35	1 5 16	LongerThan1Year	39
96.247	75.745	25.253	21.638
34 36 32	9 22 37	1month	5
565.327	172.534	124.267	131.094
11 45 29	20 18 48	2 months	18
181.52	50.318	96.819	42.5
32 35 36	8 1 5	3months	6
174	102	52	28
22 36 34	5 48 9	Longer than 1 Year	32
298.755	216.238	122.186	165.12
34 22 37	5 48 20	Longer than 1 year	29
167.901	83.627	48.779	67.628
22 36 37	5 9 20	6 month	29
120.037	124.035	53.953	41.049
36 22 21	1 18 48	longerthan1year	5
167.415	146.022	144.78	120.943
21 22 7	1 5 16	3 months	9

	206.138	87.934	269.073	47.767
22 36 6		5 20 18	1month	5
	62.015	138	61.78	74.943
22 36 37		5 48 20	1 month	48
	433.091	197.103	209.337	103.485
36 22 21		1 18 48	longerthan1year	5
	243.415	350.022	182.08	120.943

Survey for Line Chart Visualization							
	It was easy for me to select the sectors I wanted to observe	It was easy for me to identify the fastest growing sectors	The viz was useful for predicting decreases in sector market capital	There were too many lines in the visualization	What I liked most about this visualization	What I did not like about this visualization	Additional Comments/Suggestions
1	Strongly disagree	Neutral	Somewhat agree	Strongly agree	I can understand the time scale very well	The lines are overlapping and it is hard to see individual lines and select one	Make each line have a unique color may improve the visualization. When mouse over a region, enlarge the region may make it easier to select
2	Strongly agree	Agree	Agree	Somewhat agree			
3	Strongly agree	Somewhat agree	Somewhat agree	Strongly agree	Interactive	Information overload	
4	Disagree		Agree	Agree	I liked the fact that the stuff greyed out when you selected particular sectors	I wish that I could have selected more sectors	Compared to the other visualization, there weren't nearly as many brushing and linking techniques in the visualization as there were in the bubble version.
5	Somewhat disagree	Somewhat agree	Neutral	Agree	mouse hovering for detail informaiton	lots of lines	
6	Strongly disagree	Agree	Agree	Strongly agree	able to easily identify increase and decrease	too many lines together	
7	Somewhat disagree	Disagree	Strongly disagree	Agree	Interactively displaying time points simultaneously on the two charts based on	Too many lines in the bottom part of the line vis; only two sectors can be selected	

Survey for Bubble and Strike Chart							
	It was easy for me to select the sectors I wanted to observe	It was easy for me to identify the fastest growing sectors	The viz was useful for predicting decreases in sector market capital	There were too many bubbles	What I liked most about the viz:	What I did not like most about the viz:	Additional comments/suggestions :
1	Strongly agree	Somewhat agree	Agree	Agree	Versatility	Sector numbers weren't clear on the bubbles.	Nicely done.
2	Strongly agree	Neutral	Somewhat disagree	Strongly agree	The color of the bubbles	There are too many bubbles	It would be nice to quantify the size of the bubbles because it is hard to tell the difference among the small bubbles.
3	Strongly agree	Agree	Neutral	Disagree			
4	Strongly agree	Somewhat agree	Somewhat agree	Agree			
5	Strongly agree	Somewhat agree	Somewhat agree	Agree	Interactive	Information overload	
6	Agree	Somewhat disagree	Somewhat disagree	Disagree	It was easy to select particular sectors in the visualization based on their growth	It was weird how I couldn't use the arrow keys to move passed 1999 (forget the month), which disoriented me and added to the length of time it took me to complete some of the tasks.	
7	Strongly agree	Agree	Somewhat agree	Agree	Easy to see market cap represented by size	A lot of information to take in and process at once	
8	Agree	Somewhat agree	Neutral	Somewhat disagree	The animation	it is hard to compare history values.	
9	Strongly agree	Somewhat agree	Somewhat agree	Agree	timeline	colors	
10	Strongly agree	Agree	Disagree	Strongly disagree	Bubble with animation is intuitive, helps to tell the trend of the market	Only two bubbles can be selected simultaneously; Display size of the whole bubble window is too small	

Survey for Combined Visualization									
	It was easy for me to select the sectors I wanted to observe	It was easy for me to identify the fastest growing sectors	The viz was useful for predicting decreases in sector market capital	The addition of one more view helped in performing tasks	Overall, the line chart was more useful than the bubble viz	The bubble viz was easier to use than the line chart	What I liked most about the viz	What I did not like most about the viz	Additional comments/suggestions:
1	Agree	Agree	Somewhat agree	Somewhat agree	Agree	Neutral	Ability to search for sectors, See trends and patterns, Relative performance,	Selecting one of the 48 sectors in the line graph was difficult, Prediction of future values was not easy	The timeline scale can have markers to mark ends, Zooming in would ease selecting a particular line
2	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Somewhat			
3	Strongly agree	Strongly agree	Somewhat agree	Strongly agree	Disagree	Strongly agree	1. It is easier to select the individual sectors in bubble vis. 2. Bubble viz gives	It is hard to estimate the length of time just based on bubble view. However when combine bubble	
4	Strongly agree	Agree	Somewhat disagree	Strongly agree	Somewhat agree	Somewhat agree	Interface was super easy compared to the lines	Trends were a little difficult to interpret	If both are used together, it will be very useful if the selection of dates is coordinated (i.e., if a select a
5	Agree	Somewhat agree	Somewhat agree	Neutral	Strongly agree	Strongly disagree	It was interactive	It was almost too much information in trying to decide	
6	Strongly agree	Agree	Neutral	Strongly agree	Somewhat disagree	Somewhat agree	The full time period of the line chart and the color of the bubble viz	The lines (bubbles) are too close to each other so it is hard to identify the trend (size) of one	N/A
7	Strongly agree	Agree	Agree	Neutral	Agree	Somewhat agree	easy select and compare	N/A	N/A

8	Agree	Agree	Agree	Agree	Somewhat disagree	Somewhat agree	I could switch back and forth between the visualizations	The views weren't as coordinated as I wanted them to be (see comment below)	The timeline should have been linked with the bubble view so that when you moved the time slider, the dot indicating particular values on the timeline would move with it. Although the timeline
9	Agree	Strongly agree	Strongly agree	Strongly agree	Agree	Agree	Ability to cross reference information	Too easy to focus only on one viz	
10	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree	Neutral	Agree		Too many views may distract my	
11	Strongly agree	Agree	Agree	Agree	Somewhat agree	Somewhat	combination		
12	Agree	Agree	Disagree	Somewhat disagree	Strongly disagree	Strongly agree	The combination of bubble and bar vis with animation help to see the change of the market intuitively; Bars help to increase	Too many lines at the bottom part of the line chart; Only two sectors can be selected simultaneously	

Appendix C

Design Iterations

This appendix details other designs that were tried before the final version for visualization. Some were rejected due to implementation issues while others were improved upon for better interactions or simplicity. This research started with the idea of using bubbles as a visual metaphor for showing sectors of the stock market and finding where sizes of the bubbles would show price times volume of the stocks in the sector and the way they would increase and decrease in sizes would show how capital flows through the market. As can be seen in Figure C.1 the four larger bubbles

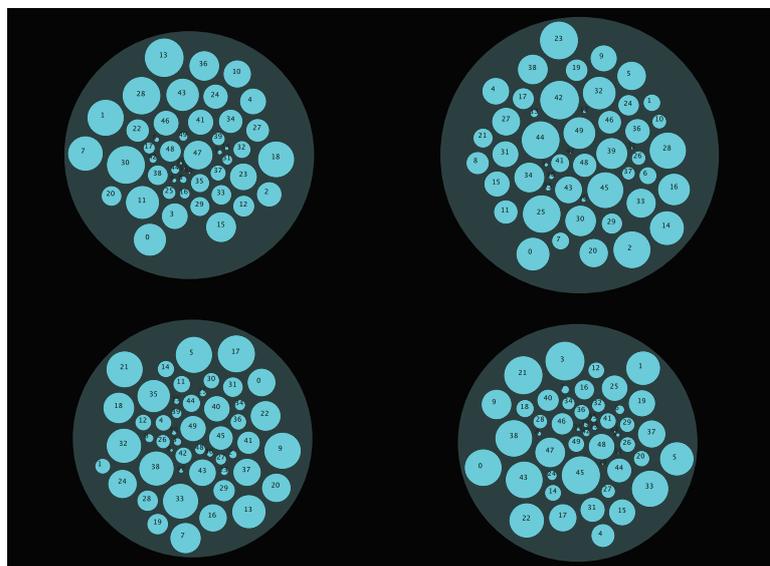


Figure C.1: Using circle packed bubbles to show stocks within sectors of the market

represent sectors of the market and the inner circles represent stocks that belong to those sectors. There were two main drawbacks to this design. This design though simple in concept was not feasible for visualizing real data of the stock market. There are too many industries, sectors and stocks to visualize this way. This is why the *overview-first, details-on-demand* approach works for

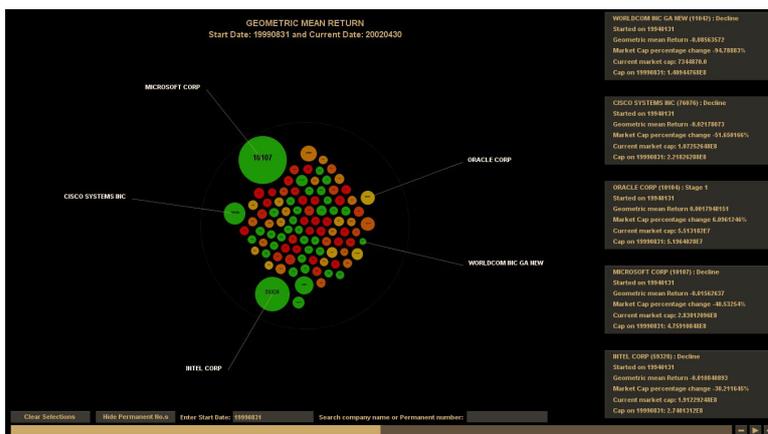


Figure C.2: Overview with few sectors and details selected

this particular data set. Another flaw was that there is no interaction between the sectors of the market as seen in such a visualization. The visualization needs to be able to show how the capital flows through various sectors. The overview was then designed with majority of the interactions in the bubble view. This can be seen in C.2 where the NASDAQ 100 companies are shown in the visualization. This dataset shows the 100 major stocks that contributed the most to the dot com crash of 2000. As can be seen here the user is allowed to select sectors to look at the details regarding that sector in a small information tab. This design decision was used in the final visualization as well.

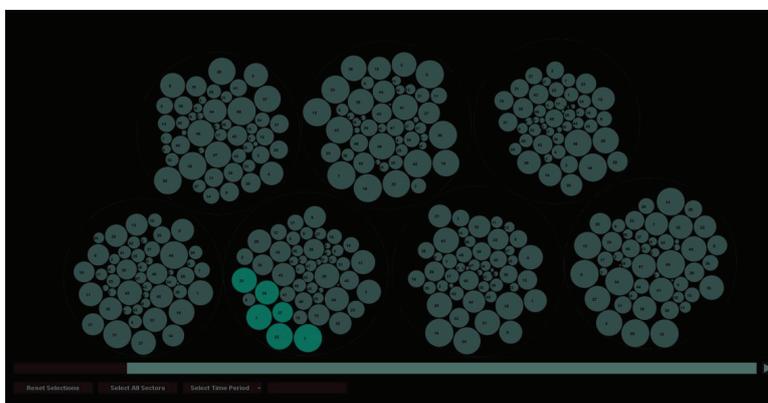


Figure C.3: Details view with circle packed sectors and circle packed stocks within sectors

While designing the details-on-demand approach for the visualization, the choice was made to allow the user the ability to select all sectors of interest and then view them with their containing stocks in the a circle packed form as shown in C.3. This however also becomes a very cluttered view because, if too many sectors have been selected from the main view with just the sectors, the space is almost completely filled with bubbles representing stocks. This is why the final visualization only allows the user to pick two sectors for the details view. Most comparisons that need to be

done in the stocks level is between bubbling and non-bubbling sectors. Hence this design decision does not harm the utility of the details view.

The final visualization was designed with a white background instead of the black so that the darker shades of green and red as well as the grays are clearly visible and distinguishable. The darker shades are those representing more extreme values in performance so it is imperative that they stand out visually and be distinctly visible.