Eurographics Conference on Visualization (EuroVis) 2012
S. Bruckner, S. Miksch, and H. Pfister
(Guest Editors)

Volume 31 (2012), Number 3

MarketAnalyzer: An Interactive Visual Analytics System for Analyzing Competitive Advantage Using Point of Sale Data

S. Ko†1, R. Maciejewski‡2, Y. Jang§3, and D. S. Ebert¶1

1Purdue University
2Arizona State University
3Sejong University

Abstract

Competitive intelligence is a systematic approach for gathering, analyzing, and managing information to make informed business decisions. Many companies use competitive intelligence to identify risks and opportunities within markets. Point of sale data that retailers share with vendors is of critical importance in developing competitive intelligence. However, existing tools do not easily enable the analysis of such large and complex data. therefore, new approaches are needed in order to facilitate better analysis and decision making. In this paper, we present MarketAnalyzer, an interactive visual analytics system designed to allow vendors to increase their competitive intelligence. MarketAnalyzer utilizes pixel-based matrices to present sale data, trends, and market share growths of products of the entire market within a single display. These matrices are augmented by advanced underlying analytical methods to enable the quick evaluation of growth and risk within market sectors. Furthermore, our system enables the aggregation of point of sale data in geographical views that provide analysts with the ability to explore the impact of regional demographics and trends. Additionally, overview and detailed information is provided through a series of coordinated multiple views. In order to demonstrate the effectiveness of our system, we provide two use-case scenarios as well as feedback from market analysts.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Pixel-based visualization, Linked Views, Geospatial, Temporal, Multi-variate, Business Intelligence, Competitive Intelligence, Market Analysis—

1. Introduction

The underlying goal of a business is to increase (or at least maintain) its current market share and to maximize its profits within the market. In order to pursue this goal, analysts must constantly explore and analyze market share data changes that are relevant to their current business sector. Their goal is to forecast changes in the market as a means of controlling and expanding the company's current market share. This exploration, analysis, and prediction of the market share is termed *competitive intelligence (CI)* [Kah98]. Companies use CI to compare themselves to other companies, to identify market risks and opportunities and to evaluate the potential impact of new sales strategies.

In order to generate intelligence reports, many companies extract information from a variety of sources using various methods of data collection and analysis (e.g., networking with company rivals, examining security filings, patent application analysis). One key data source is point of sale data that retailers share with vendors. This point of sale data is temporal, multivariate, and spatial in nature; therefore, it is well suited for analysis in a visual analytics environment. However, it is difficult to find systems that manage the characteristics of point of sale data effectively. In this paper, we present MarketAnalyzer, a visual analytics system for exploring, comparing, analyzing, and predicting trends of point of sale data. We have worked directly with analysts to provide proper and accurate analysis of their point of sale data (e.g., 288 stores with 36 different products) to increase their understanding and improve their market insight. We use an enhanced pixel-based visualization approach [KK94, Kei00] in MarketAnalyzer to efficiently utilize limited screen space for the large store and product

† ko@purdue.edu
‡ rmacieje@asu.edu
§ jangy@sejong.edu: corresponding author
¶ ebertd@purdue.edu

information. Our system allows analysts to explore current sales volume, trend, and temporal market share growth rates using a series of linked views including pixel-based visualization matrices, line graphs, stacked bar graphs, and choropleth maps.

MarketAnalyzer has several benefits compared to other tools for performing market analysis tasks. MarketAnalyzer enables analysts to investigate the status of the market by observing all the characteristics of point of sale data at the same time. In addition, the status of the competition in point of sales, trends and growth rates is projected onto a map for regional market analysis. MarketAnalyzer provides forecasts for both individual products and different stores utilizing statistical models, such as linear trend estimation [DS98] and ARIMA (Auto-Regressive Integrated Moving Average) [BJ76]. In order to reduce the perceptual difficulties inherent in pixel-based visualizations, a local magnification lens is also provided for focus + context analysis. Additionally, CUSUM [Pag54] and normalized trend filtering are provided for data filtering. For evaluation, we provide two case studies that describe how sales data can be analyzed with MarketAnalyzer. Although the case studies are business domain specific, it is easy to extend our system to other multivariate, spatial, temporal datasets such as property sales, crime and disease data to provide comparisons, insight, and new intuition.

2. Related Work

Traditionally various tasks, such as discovering market trends and predicting future prices of assets have been addressed with charts and line graphs in the financial data domain [Mur99, EM01]. While charts and line graphs provide useful visualizations of univariate data, they quickly become clutter as new dimensions are added. Analysts often use tree map visualizations to represent the market [Sma, VvWvdL06]. Unfortunately, these maps only provide a snapshot of the current market value whereas analysts often wish to explore short or long term trajectories within the market [TA03, STKF07].

In order to display the maximum amount of data relative to the screen space, Keim et al. introduce pixel-based [OJS*11] or pixel-oriented [KK94, Kei00] visualization techniques. In these techniques, each data element is assigned to a pixel. Then, a predefined color map is used to shade the pixel to represent the range of the data attribute. Thus, the amount of information in the visualization is theoretically limited only by the resolution of the screen. In the context of our work, pixel-based visualizations are visualizations that utilize small areas of the screen to encode one data item. Note that the areas may not necessarily be pixels, as the use of small rectangles also falls under the accepted classification of pixel-based visualizations [OJS*11].

Borgo et al. [BPC*10] present how the usability of the pixel-based visualization varies over different tasks and block resolutions. Oelke et al. [OJS*11] studies visual boosting techniques for pixel-based visualization such as halos and distortion. Ziegler et al. [ZNK08] presents how the pixel-based visualization helps analysts gain insight for long-term investments.

Many systems have been developed for visually exploring multivariate data (e.g., Xmdv [War94], Spotfire [Ahl96], XGobi [SCB98], GGobi [SLBC03], Comvis [MFGH08], Polaris [STH08], Tableau [Tab]). Common amongst these systems is the extensive use of interactive techniques (brushing, linking, zooming, filtering) to refine the user's queries. However, such systems often do not support market forecasting or geographical analysis. In discussing design strategies with our market analysts, it was noted that forecasting future trends and understanding outperforming geographical locations are important in market analysis. Of the systems previously listed, Tableau software [Tab] allows analysts to easily access and analyze their data by offering flexible operations. Although multivariate and time-series data analysis is possible in the tool, comparison between multivariate attributes with geographical information is not well supported by Tableau. In MarketAnalyzer, all attributes of the data is visualized in multiple linked views for simultaneous comparison, and analysts can investigate future sale trends based on statistical models and market share growth rates.

3. Visual Analytic Environment

Analysts tend to easily understand competition within a market and quickly draw conclusions when maximal information is presented. In order to present the most information for analysis, coordinated and multiple linked views have been used in various applications [SFOL04, WFR*07, CGK*07, Rob07, SGL08]. In this work, we also employ coordinated multiple linked views to visualize attributes from point of sale data. Figure 1 shows how MarketAnalyzer provides complete information in multiple linked views. Note that all color maps are chosen to fit perceptually with the data being analyzed. Both sequential and divergent color maps from ColorBrewer (which have been previously tested and evaluated [HB03]) are used. Sequential color scales are chosen to show ordered data, while divergent color maps are chosen to show differences in data values with respect to some point of interest.

Companies are displayed in a selectable list in (a), stores in (b), and products in (c). Note that in window (b), there are two selectable lists. The leftmost store list is used to select multiple stores for computing the sales average, while the rightmost list is used to select a single store, whose sales will be compared to the computed sales average. This single store selection is the anchor for view (h) and (i) that shows the sales of the primary company (h) or competitor company (i). In (h) and (i), the selected store's sales of the products chosen in the list view (c) are plotted in green, and the average sales of the products across a group of stores selected

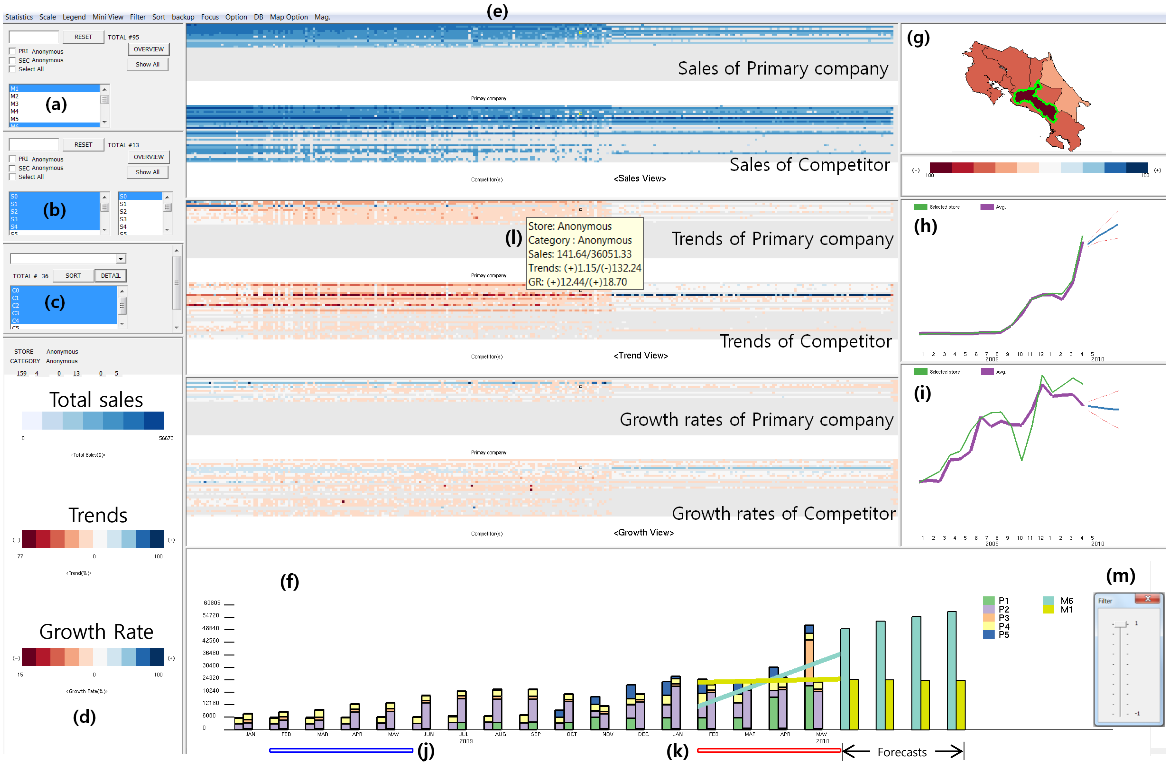


Figure 1: *The MarketAnalyzer interface. MarketAnalyzer consists of multiple coordinated views linked with interactive filters: (a) Company filter, (b) Store filter, (c) Products filter, (d) Legend view, (e) (Sorted) Matrix view for sales, trends, and growth rates. (f) Stacked bar view, (g) Geographical view, (h) and (i) Line graph small multiples views, (j) and (k) Time slider widgets and aggregation tools for temporal comparison. (l) Tooltip. (m) Filter. In the legends, the blue indicates positive and the red represents negative measurements in sales, trends, or growth rates.*

in the leftmost list of (b) are plotted in purple. At the end of the line graph, a four week forecast for sales based on an ARIMA model is plotted as a blue line bounded by two red dotted lines representing the upper and lower error bounds. Note that the ARIMA forecasts are calculated in *R* [R D06], which is integrated directly into our system.

3.1. Pixel-Oriented Display Matrix

A fundamental challenge in the visualization of large multivariate data is that screen space limits the amount of information that can be simultaneously presented to a user. This scalability problem causes various difficulties in analysis, such as inefficiency in comparison and tedious jumping back and forth to adjust different parameters. In order to alleviate this problem, we incorporate a pixel-based visualization [KK94, Kei00] that is effective when the screen space is limited. Sales, trends, and market share growth rates for stores and products (e.g., 288 stores, 36 products) are effectively presented in MarketAnalyzer as shown in Figure 1 (e). Note that we place different stores and products in different columns and rows, respectively. The matrix view in (e) consists of three views: sales, trends, and growth views. Each view has two matrices for a primary company and its competitor that are chosen in Figure 1 (a) by users. We place small squares side-by-side in each matrix with all matrices positioned vertically. The rationale behind this arrangement is that its conceptual simplicity makes the comparison and discovery of trends easier. Also comparing data side-by-side is more efficient than jumping back and forth and memorizing previously shown data, based on the principle of small multiples [Tuf90].

In the sales view, we use Equation 1 to define the sales ($S_{i,j}$) for each square as a sum of the sales during the user-selected time interval.

$$S_{i,j} = \sum_{t=m}^{n} Sales(t,i,j), \tag{1}$$

where i and j indicate the i_{th} row (product) and the j_{th} column (store) while m and n are the first and the last month in the time interval. The darker the blue, the more units are being sold. One frequent question that decision-makers might have is "Are sales increasing in this specific time period?". In order to answer this question, we present the user with a trend view in which the slope (variable b in Equation 2) of our linear trend estimation is visualized using a divergent color scale. Positive slopes (indicating that sales are trend-

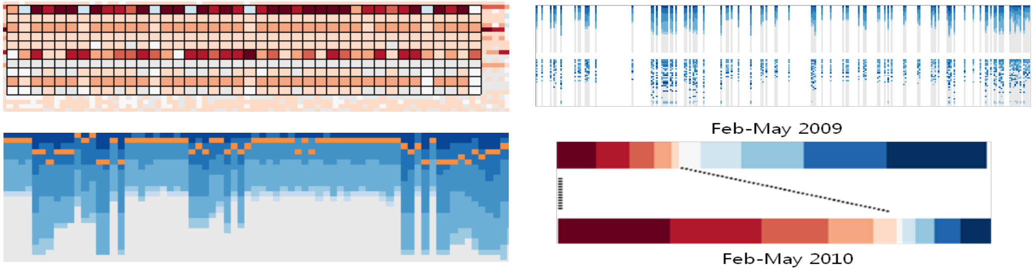


Figure 2: *(top-left) Magnification is applied for detailed comparison, (top-right) the CUSUM filtering method with the strict option is applied on Feb 2010, (bottom-left) SimulSort is applied to the sales view, (bottom-right) Proportional legend.*

ing upwards) are mapped towards the blue values, negative slopes (indicating sales are trending downward) towards the red values, with a slope of zero being white (indicating that sales are stable).

$$y = a + bx, \tag{2}$$

where x, y are variables, a is the intercept point of the regression line and the y axis, b is the slope from the linear trend estimation for a certain time interval. b is computed in Equation 3,

$$b = \frac{n(\sum_{m=1}^{n} X_m Y_m) - (\sum_{m=1}^{n} X_m)(\sum_{m=1}^{n} Y_m)}{n(\sum_{m=1}^{n} X_m^2) - (\sum_{m=1}^{n} Y_m)^2} \tag{3}$$

Here, n represents the number of months in the specified time interval, Y_m is the index of each month (e.g., 1, 2, ..., n), and X_m is the sales for the month.

Along with sales trends, we also define the growth rate of a business using Equation 4.

$$G_{i,j} = \frac{S_{i,j,cur} - S_{i,j,past}}{S_{i,j,past}} \times 100, \tag{4}$$

where $S_{i,j,cur}$ is defined as the sum of the sales between the first (M_f) and the last (M_l) month using the red time slider (Figure 1-(k)), as shown in Equation 5.

$$S_{i,j,cur} = \sum_{T_{red}=M_f}^{M_l} Sales(T_{red}, i, j). \tag{5}$$

$S_{i,j,past}$ is similarly defined for a past time interval (T_{blue}) from the blue time slider (Figure 1-(j)). A divergent color scale is applied to represent the growth rate. When $S_{i,j,past}$ is zero, the growth rate is set to the maximum rate in the range. This indicates a case where a product was not supplied at the selected time but was provided later at the time specified by the red time slider.

We calculate the competitive advantage for the primary company ($C_{P_{i,j}}$) using Equation 6 when the analyst activates the comparative mode, as shown in Figure 5.

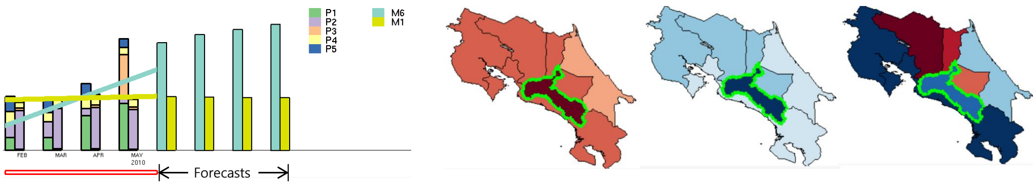
$$C_{P_{i,j}} = M_{P_{i,j}} - M_{C_{i,j}}, \tag{6}$$

where $M_{P_{i,j}}$ is a measurement of the i$_{th}$ product from the j$_{th}$ store for the primary company and $M_{C_{i,j}}$ is that of the competitor. The measurement can be any input data attribute, such as sales, trends, and growth rates used in these examples. This calculates a new matrix representing the competitive advantage of the primary company. Analysts should be alerted about a red row or a red column because it explicitly indicates the primary company is losing sales in comparison to its competitor as shown in Figure 5.

In order to provide more detailed information, we use a tooltip and magnification lens. A tooltip, as shown in Figure 1 (l), provides numerical information of the product in the store where a mouse is hovering. Squares are gray when no product is supplied to the stores. In order to help recognition and comparison, MarketAnalyzer provides a magnification lens as shown in Figure 2 (top-left).

We reorder rows in the matrix to emphasize the importance of data where squares near the top-left locations are more important than squares near the bottom-right. For the sales matrix, rows and columns are sorted based on sales sums. Products (rows) are sorted by the sales sum across the entire stores and then stores (columns) are re-sorted by the sales sum across the entire products. The topmost row indicates the top-selling product, and the leftmost column represents the top-selling store in the market as shown in Figure 1 (e). Note that matrices in the trend and growth views are also sorted corresponding to the sorting of the sales matrix. Negative measurements in trends and growth rates represent adverse situations and analysts want to find out the adverseness promptly in their analysis. In this case we use adverseness for the importance.

We also incorporate SimulSort [HY09] to sort columns independently as an auxiliary approach. In each store, products are sorted by sales. Thus, squares with higher sales tend to go upward in matrices. This enables analysts to quickly evaluate sales performance of products in a store. In order to see the performance of a product across all stores, MarketAnalyzer highlights the product in orange for all stores as presented in Figure 2 (bottom-left) where the sales performance of the product is good in most stores with some minor variations.



(a) Stacked bar view (b) Geographical view

Figure 3: *(a) The stacked bars represent trends in individual products. Analysts can see sudden increase in sales of product P3 in May. Note that M1 and M6 indicate company names. (b) Competitive advantages in February–May 2010 are linked on the maps to represent regional competitions for sales (left), trends (middle), and growth rates (right) compared to the past time interval, March–June 2009.*

3.2. Filtering Methods

MarketAnalyzer helps analysts filter out uninteresting stores by providing two filtering methods: the cumulative summation method (CUSUM) [Pag54] applied to sales and the trend filtering method based on the computed trends. The CUSUM filtering method has two modes: default and strict. In the default mode, the CUSUM method in MarketAnalyzer filters out stores during the visualization process that have a negative $S_{st,n}$ from Equation 7.

$$S_{st,n} = \sum_{m=1}^{n} (X_{st,m} - \mu_{st}) \tag{7}$$

Here, $S_{st,n}$ is the CUSUM of the *st* store in the n_{th} month (where the n_{th} month is the first month of the red time slider). $X_{st,m}$ is the sum of all products of store *st* at the m_{th} month (where the m_{th} month is started from the first month of our data set–January 2009). μ_{st} is the mean sales of st over the time period *n-m*. The CUSUM method can also be changed from the default parameter to a strict mode. In the strict mode, stores are filtered out only when the CUSUM of the n_{th} month and the $n-1_{th}$ month are negative. If the CUSUM of the n_{th} month is negative but that of the $n-1_{th}$ month is positive, then the store will still be visualized.

When the trend filtering method is used instead of CUSUM, MarketAnlayzer maps trends from -1 to 1. Then, MarketAnalyzer visualizes stores whose trends are below the threshold specified in the filter (Figure 1 (m)). The normalized trend N_{st} for store *st* is computed in Equation 8.

$$N_{st} = \frac{T_{st} - T_{mean}}{\sqrt{T_{var}}} \tag{8}$$

Here, T_{st} is the sum of the trends of all products in the store *st*, T_{mean} and T_{var} are the mean and variance of trends across all stores. Then, N_{st} is mapped to M_{st} in Equation 9.

$$M_{st} = \left(\frac{N_{st} - N_{min}}{N_{max} - N_{min}} - 0.5 \right) * 2, \tag{9}$$

where N_{max} and N_{min} are the maximum and minimum in normalized trends.

3.3. Proportional Legends

The legend view in Figure 1 (d) provides numeric and color map information for each view. The scale in MarketAnalyzer can be adjusted to have the denominator represent either measurements (sales, trends, and growth rates) at the local level (comparing measurements within a company) or at the global level (comparing measurements across the primary company and its competitor). In the local scale mode, six legends (two for each view) are provided, while three legends (one for each view) are provided in the global scale mode.

Generally a legend consists of evenly divided intervals. However, analysts often need to walk through a huge volume of data within a relatively short period to check for abnormalities. For the market analysis, the legend can be as important as the data itself [Cle94, Wil05, TLH10]. In order to enhance understanding abnormalities of the data, the intervals for the legends in MarketAnalyzer can be nonlinearly mapped according to importance as shown in Figure 2 (bottom-right). Analysts need to be alerted when red intervals are much wider than blue intervals. For instance, the change of the width for negative measurements during investigation explicitly presents the worst growth rates.

3.4. Geographical View

The geographical view as shown in Figure1 (g) supports the visual analysis of spatiotemporal patterns. The regional status of competition is a measure of the difference between two selected companies within the selected geographical area and is represented by colors. For instance, analysts are able to see that the primary company has lower sales (red) compared to its competitor in the left most image in Figure 3 (b). However, the market looks optimistic (blue) for the primary company because the sales trend is increasing (middle image in Figure 3 (b)). Note that when a mouse hovers on a store in the matrix view, the corresponding region on the map is highlighted.

Dynamic querying for specifying time intervals also plays an important role in geographic analysis because sales patterns evolve over time. This querying helps analysts identify complicated spatiotemporal patterns in the competition and facilitate appropriate strategies. For instance, analysts can easily find the time when the leading competitor started losing its competitive advantage. Then, the analyst can start investigating reasons for the loss by investigating informa-

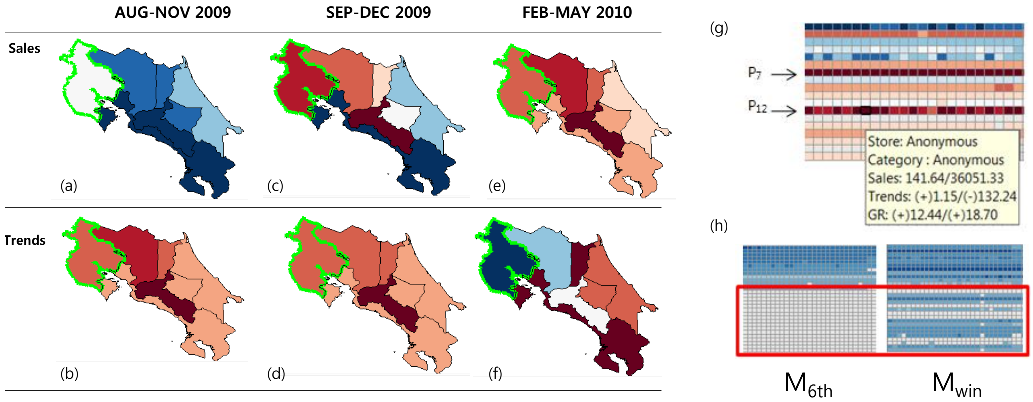


Figure 4: *(a)–(f) Analysis from the geographical view. The sales row shows the process of losing competitive advantage while the trends row presents forecasts for each column interval (August 2009–May 2010). The blue color represents good performance while the red color represents bad performance for the primary company compared to its competitor. (g) The trend view helps an analyst design short term tactics such as promotions. This example shows a decreasing overall sales trend of the competitor in some stores. The analyst notes that the competitor has the worst downward trend in products P_7 and P_{12}. (h) The red box represents possible new markets for the new company M_{6th} but its competitor M_{win} has already started its business in various products.*

tion on the matrices and other views. These processes during analysis are described in Section 4.

3.5. Stacked Bar View and Time Sliders

Well-designed stacked graphs are popular because of their aesthetics and ease of perception [HHWN02, BW08]. Since it is important to verify combinatorial trends of multiple products, we employ stacked bar graphs to investigate these trends as shown in Figure 1 (f) whose products and stores are chosen from (b) and (c). When a person buys two products together, there could be various assumptions to explain why these two are chosen. It may be because of complementary relationship between two products (e.g., ketchup and mustard), similar purchase cycles (e.g., beer and diapers), impulse purchases of a product (e.g., candy bar), or undiscovered reasons, such as happenstance [MAG99]. Figure 3 (a) shows an example where one can see that as the sales of P_1 increases and reaches a certain level, the sales of P_3 rapidly increase as well.

Draggable and length-changeable time sliders are used to select time intervals for the stacked bar view, as shown in Figure 1 (j) and (k), respectively. The red slider (k) is for specifying current time interval. For the growth rate computation, a past time interval is required that is selected using the blue slider (j). When the slider is dragged or the range of the slider is changed, the time interval in the analysis also changes, updating all other views with the new time interval.

4. Case Studies

We describe two scenarios with anonymous manufacturers, stores, and products for privacy. The first scenario presents the process of designing a strategy for a young company that wants to increase its market share. The second scenario illustrates the analysis of a defending champion that needs to verify and understand its competitive advantage compared to 94 other competitors in the market.

4.1. Analysis to Step into a New Market

An analyst in a two year old company (M_{6th}) is asked to look for potential areas in which the leading competitor (M_{win}) may currently be showing signs of weakness. By finding these weaknesses, the new company can begin making inroads into the market as it expands its base. The exploration begins in the pixel-based matrix views, as shown in Figure 1 (e). In the top matrix view, the y-axis represents all 36 products that are currently sold in the market. The x-axis represents all 288 stores that sell its products. The analyst chooses to sort the matrix by sales. Here, the darker the blue, the more units are being sold. In the local mode, the analyst can see which stores and products are the top sellers for M_{6th} and M_{win}. On the other hand, in the global mode, as shown in Figure 4 (h), the analyst sees that the competitor has more blue squares, meaning it is outperforming M_{6th}. In the red box in Figure 4 (h), the analyst easily verifies the products and stores in which the new company is not supplying any product while the competitor has been earning additional profit.

Before deciding on which market opportunities might be the most profitable, the analyst needs to understand the weakness and trends within the market. In the trend view, the analyst sees the rate of growth or decline over the last sales period. First, the sales period is selected using one time slider (e.g., February–May 2010) in the lower portion of the

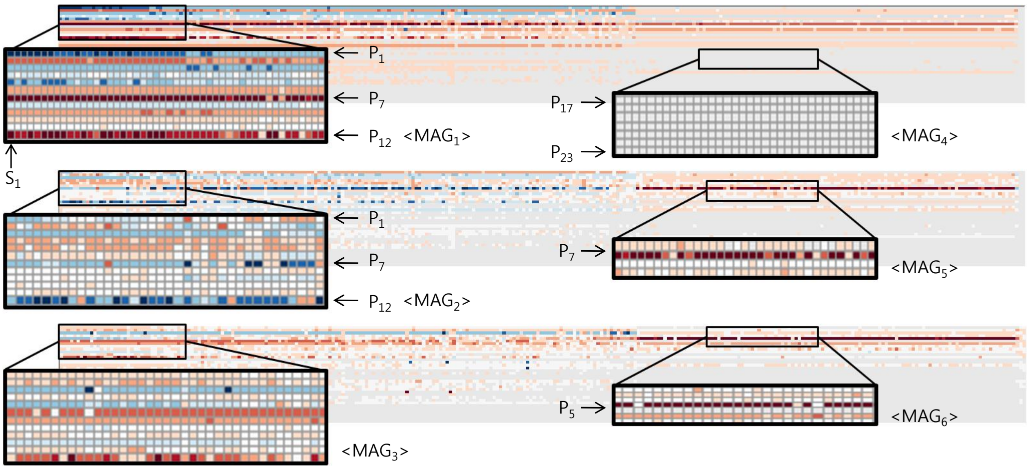


Figure 5: *Our pixel-based matrix views using the comparative mode. Each row of the matrix represents a product, and each column represents the store selling the product. The three views present three types of information (from top to bottom): sales, trend, and growth rate comparisons between two selected companies. Note that gray color indicates zero. The blue color presents positive and the red color represents negative in difference of the two measurements.*

stacked bar view as shown in Figure 1 (k) (red slider). Here, the analyst sees that many of the stores that the new company supplies are showing a downward growth trend. While alarming, it is important to note that the leading competitor is also displaying negative trends for similar products during this time period. Looking at the trend analysis, the analyst notes that the P_7 and P_{12} products of the competitor have the worst trends, which might indicate they are becoming less popular. In that case, they could be the targets for the new company to take the market share as shown in Figure 4 (g).

Of primary interest to the analyst is the matrix view under the comparative mode, where the analyst can explore how the current product sales is performing with respect to M_{win}. If a square is blue, then M_{6th} is outperforming M_{win}, and if red, then M_{win} is still outperforming the new company. Variations in the red and blue hue show the degree of sales performance. The analyst notices the following from Figure 5.

- P_1 is the best-selling product, outperforming its competitor (MAG_1),
- P_7 records the worst sales performance in almost all stores (MAG_1). The sales of the P_7 in two thirds of stores is expected to increase (MAG_2) while those in other stores will still keep decreasing (MAG_5),
- P_5 has not been sold much in one third of the stores (MAG_6),

In Figure 3 (a), the analyst sees that the sales of P_3 suddenly increased in May 2010 while those of P_1 has grown gradually. Thanks to increasing sales in these two products, M_{6th} has outperformed its competitor since May 2010 in those selected products and it is predicted to outperform the competitor with those products in June 2010 according to the ARIMA forecast.

The analyst further wants to explore the distribution of its performance across stores and chooses to use the geographical view for regional competition analysis. Competition in the last quarter (February–May 2010), as presented in Figure 3, can be summarized as follows:

- Sales (Figure 3 (b) Left): M_{6th} has difficulty in overcoming its competitor in all regions. It is notable that the company is losing in the most important region including the capital city that is colored in the darkest red.
- Trends (Figure 3 (b) Middle): The trends are optimistic in all regions and the most important region has the highest upward trend.
- Growth rates (Figure 3 (b) Right): Sales in some regions have grown compared to sales in March–June 2009 although there are regions where growth rates are negative.

Therefore, the analyst will propose attacking the weaker points of the leading competitor (M_{win}). For instance, P_7 and P_{12} of the competitor in the trend view in Figure 4 (g) are expected to keep losing competitive advantage. Thus, they might be proper targets for launching additional new products, to take over the market. Conversely, to keep or increase the sales, a local promotion focusing on the capital city or a global advertisement should be designed for the P_7 product, as shown in MAG_1 in Figure 5. Giving up P_5 in MAG_6 might not be a good strategy since it still remains in 5th place in sales for the company. Watching other companies' strategies when they launch a new product, could be an effective solution at the moment. In addition, the analyst notices that there are many gray products where M_{win} does not supply its products. These would be the secondary locations where M_{6th} could launch new products for earning more profit. Lastly, finding reasons for the abnormal sales pattern between P_1 and P_3 in Figure 3 (a) is important as well since they are in the 1st and 3rd places in sales.

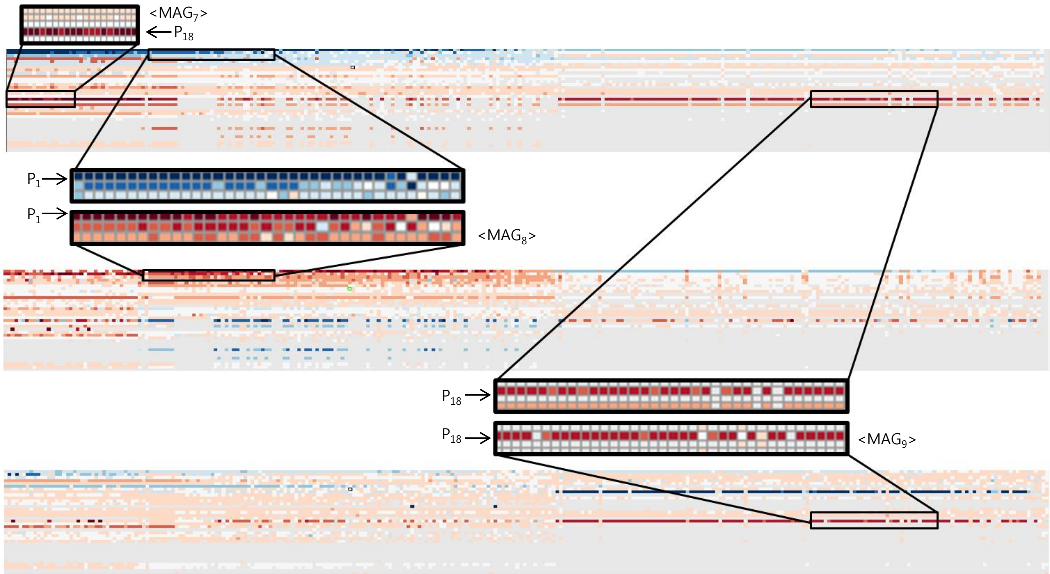


Figure 6: *The matrix view under comparative mode with sorting provides direct information for the competition. These matrices imply a possible strategic failure that could cause the loss of a big market.*

4.2. Analysis of a Defending Champion

Often, the leading competitor's analyst (company M_{win}) is asked to analyze market share competition that requires a performance comparison of all companies in all regions. The analyst might begin exploration in the geographical view that quickly shows the regional competition status for a given time interval. In Figure 4 (a), the analyst sees that M_{win} had market share higher than 50% before November 2009 but the trend during the period decreased in all regions (Figure 4 (b)). The analyst notes that the most important region, colored in the darkest red, was expected to have the most severe decrease in sales. From December 2009, as shown in Figure 4 (b), M_{win} started losing competitive advantage in three regions but still had three winning regions (Figure 4 (c)). It is notable that the regions that had moderate downward trends (Figure 4 (b)) still kept a competitive advantage, while others with darker red colors were losing the competitive advantage. At the same time, forecasts kept warning of downward trends in all regions (Figure 4 (d)). In the end, M_{win} did not have any winning region during February–May 2010 (Figure 4 (e)). However, Figure 4 (f) shows two regions having upward trends in sales (darker blue colors). During the analysis, the analyst sees that the proportional legends keep changing. For instance, the red area representing an adverse situation, only occupied a small portion of the growth legends in the February–May 2009 time period but it ends up taking almost 80% of the area in the February–May 2010 period, as shown in Figure 2 (bottom-right). The *strict* CUSUM filtering also suggests deeper investigation in February 2010 because the number of visualized stores is largest (93 stores) while 28 stores are visualized in January 2010. This implies the largest number of stores turned profit into loss in cumulated sales in February 2010 as shown in Figure 2 (top-right). The default CUSUM filtering also effectively reduces the number of stores, which, for example, visualizes 150 stores on average during October 2009–January 2010 (48% reduction).

In order to find the causes for the loss of competitive advantage, a view of primary interest to the analyst is the matrix view under the comparative mode, as shown in Figure 6. In this view, the analyst sees one of the problematic products. From October 2009, the company started losing competitive advantage in P_{18}. The effect of the loss was not significant at that time. However, during its sharpest decrease between December 2009 and March 2010, the sales of product P_{18} decreased in January–April 2009. Even worse, P_{18} had the worst growth rate in about 40% of the stores severely (MAG_9). This implies a strategic failure that could cause the loss of a big market. On the other hand, the analyst sees additional important information in MAG_8 meaning that the three top-selling products (P_1–P_3) will severely lose competitive advantage in the next month. Through the analysis, the analyst verified how the leading company has been losing its market share due to the decreasing sales performance. The analyst also verified that as a product became less popular in the market, the trend, growth, and proportional legend view reflected the adverse situation while the geographical view highlighted the region where the sales performance is lowest. Although discoveries should be interpreted from various perspectives in management strategies, we believe that the two use-case examples show the effectiveness of the system for competitive advantage analysis using point of sale data.

For our analysis, we collaborated with two groups of pro-

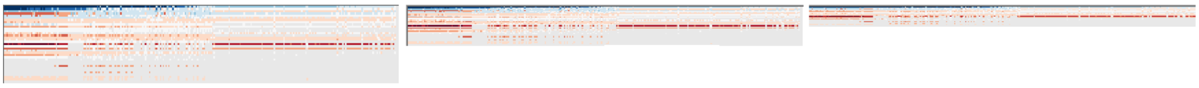


Figure 7: *288x36 resized pixels with assumptions of 63 (left), 126 (middle), and 252 (right) products in 1000 stores.*

fessionally employed analysts. The first group consisted of four analysts that had not previously used automated tools for analyzing CI. Working in a supervised learning environment in our lab, they described that they required tools for report generation and trend discovery. They noted that they usually compare singular data aspects one at a time using conventional tools such as Excel. The second group consisted of three analysts who were working with custom-made tools for CI analysis. Their tools utilized line charts, bar graphs and data aggregation schemes. The analysts in group 2 noted that the support of geographical visualization and the coupling of advanced analytical methods were a marked improvement over their current tools. Both groups utilized MarketAnalyzer and reported on their increased ability to assess sales performance against their competitors. The greatest benefit in the analysis is that analysts can recognize the overall competition status at a glance without tedious selections. This enables time-efficient investigation and easy comparison in various dimensions that can lead to discovery of unexpected trends. The system is also very helpful for people who are responsible for analyzing a competitive advantage in wide areas such as countries. In addition, the system is easy to understand and use. A manager from the company's finance department immediately gained new insight without difficulty by using the system for few minutes.

5. Discussion

In this work, we discussed a relatively small data set using 288 stores and 36 products. In this section we discuss some of the potential limitations of our chosen techniques when scaling to larger datasets. One key scaling issue is the limitation of the pixel-based visualization. Figure 7 illustrates the effects of resizing the matrix. We show results using 63, 126 and 252 products (from left to right) in 1000 stores. Here, we see that we are still able to distinguish between blue and red stores (P_1– P_3 and P_{18}) as well as discrete patterns of light red pixels (left, middle). As the number of products increases, the ability to distinguish between such groupings becomes more difficult. As such, intelligent filter controls are necessary to help show the most important aspects of the data (thus the motivation for the CUSUM and trend filtering). In our current data set, of the 288 stores, approximately 157 stores (std: 4.74) would be removed utilizing the $default$ mode, and 229 stores (std: 23.86) would be removed utilizing the *strict* mode. Future work will focus on other importance metrics for guiding the data analysis. Another method for overcoming the scaling issue is to use focus+context interaction methods. Our current system provides a single-level zoom magnification lens. Future iterations of this tool will employ further interaction techniques to better facilitate problems of scale. Finally, the stacked bar chart method employed also becomes less effective as the data size increases. However, the purpose of this particular view is to look at a small combination of products. Furthermore, future iterations will utilize more statistical methods and machine learning algorithms for directing users to particular products of interest. In ARIMA, blue and red are chosen due to their maximum distance in a diverging color scheme [HB03] to contrast with each other.

6. Conclusions and Future Work

We have introduced a new visual analytics tool for market and competitive business analysis incorporating multiple tightly integrated interactive visualizations with integrated trend analytics. Our zoomable and sortable matrix displays presents sales, trends, and growth rates with enhanced pixel-based visualization, while linked line graph views and stacked bar views aid analysis and awareness acquisition of global and specific product/store information. The linked choropleth maps enable geospatial, temporal, and regional competition analysis. We presented two use-case examples using real point of sale data to illustrate the use and potential of the system. Our system can be easily applied to analysis with any other multivariate spatiotemporal data. In the future, we plan to improve the geographical view with advanced selecting and filtering to investigate correlations between point of sale data, demography, and geolocations for more advanced business analysis. To find strength and weakness, a user study comparing to other alternative techniques such as small multiple horizon graphs is needed. We also plan to deploy our system with our corporate partner and start a longitudinal study.

Acknowledgements

This work was supported in part by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001. The authors would like to thank the analysts and corporate partners for feedback, advice and evaluation during all stages of the design process.

References

[Ahl96] AHLBERG C.: Spotfire: An information exploration environment. *ACM Special Interest Group on Management of Data Record 25*, 4 (1996), 25–29. 2

[BJ76] BOX G., JENKINS G.: *Time Series Analysis: Forecasting and Control*. Holden–Day Press, San Francisco, U.S.A., 1976. 2

[BPC*10] Borgo R., Proctor K., Chen M., Janicke H., Murray T., Thornton I.: Evaluating the impact of task demands and block resolution on the effectiveness of pixel-based visualization. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 963–972. 2

[BW08] Byron L., Wattenberg M.: Stacked graphs - geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1245–1252. 6

[CGK*07] Chang R., Ghoniem M., Kosara R., Ribarsky W., Yang J., Suma E., Ziemkiewicz C., Kern D., Sudjianto A.: WireVis: Visualization of categorical, time-varying data from financial transactions. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology* (2007), pp. 155–162. 2

[Cle94] Cleveland W. S.: *Visualizing Data*. Hobart Press, New Jersey, U.S.A., 1994. 5

[DS98] Draper N. R., Smith H.: *Applied Regression Analysis*. Wiley-Interscience, New Jersey, U.S.A., 1998. 2

[EM01] Edwards R., Magee J.: *Technical Analysis of Stock Trends*. Amacom Press, New York, U.S.A., 2001. 2

[HB03] Harrower M. A., Brewer C. A.: Colorbrewer.org: An online tool for selecting color schemes for maps. *Cartographic Journal 40*, 1 (2003), 27–37. 2, 9

[HHWN02] Havre S., Hetzler E., Whitney P., Nowell L.: ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics 8*, 1 (Jan. 2002), 9–20. 6

[HY09] Hur I., Yi J. S.: Simulsort: Multivariate data exploration through an enhanced sorting technique. In *Human-Computer Interaction. Novel Interaction Methods and Techniques, 13th International Conference, HCI International 2009, San Diego, CA, USA, July 19-24, 2009, Proceedings, Part II* (2009), vol. 5611 of *Lecture Notes in Computer Science*, Springer, pp. 684–693. 4

[Kah98] Kahaner L.: *Competitive Intelligence: How to Gather Analyze and Use Information to Move Your Business to the Top*. Touchstone Press, New York, U.S.A., 1998. 1

[Kei00] Keim D. A.: Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics 6*, 1 (2000), 59–78. 1, 2, 3

[KK94] Keim D. A., Kriegel H.: VisDB: Database exploration using multidimensional visualization. *IEEE Computer Graphics and Applications 14*, 5 (1994), 40–49. 1, 2, 3

[MAG99] Manchanda P., Ansari A., Gupta S.: The shopping basket: A model for multicategory purchase incidence decisions. *Journal of Marketing Science 18*, 2 (1999), 95–114. 6

[MFGH08] Matkovic K., Freiler W., Gracanin D., Hauser H.: ComVis: A coordinated multiple views system for prototyping new visualization technology. In *Proceedings of 13th International Conference on Information Visualisation* (2008), pp. 215–220. 2

[Mur99] Murphy J. J.: *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. Prentice Hall Press, New Jersey, U.S.A., 1999. 2

[OJS*11] Oelke D., Janetzko H., Simon S., Neuhaus K., Keim D. A.: Visual boosting in pixel-based visualizations. *Computer Graphics Forum 30*, 3 (2011), 871–880. 2

[Pag54] Page E. S.: Continuous inspection scheme. *Biometrika 41* (1954), 100–115. 2, 5

[R D06] R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0. URL: http://www.R-project.org. 3

[Rob07] Roberts J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In *Proceedings of Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization* (2007), pp. 61–71. 2

[SCB98] Swayne D. F., Cook D., Buja A.: XGobi: Interactive dynamic data visualization in the X window system. *Journal of Computational and Graphical Statistics 7*, 1 (1998), 113–130. 2

[SFOL04] Shimabukuro M., Flores E., Oliveira M., Levkowitz H.: Coordinated views to assist exploration of spatio-temporal data: A case study. In *Proceedings of Second International Conference on Coordinated and Multiple Views in Exploratory Visualization* (2004), pp. 107–117. 2

[SGL08] Stasko J. T., Görg C., Liu Z.: Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization 7*, 2 (2008), 118–132. 2

[SLBC03] Swayne D. F., Lang D. T., Buja A., Cook D.: GGobi: Evolving from XGobi into an extensible framework for interactive data visualization. *Journal of Computational Statistics & Data Analysis 43*, 4 (2003), 423–444. 2

[Sma] SmartMoney. http://www.smartmoney.com/marketmap/. Accessed by 12 Aug 2011. 2

[STH08] Stolte C., Tang D., Hanrahan P.: Polaris: A system for query, analysis, and visualization of multidimensional databases. *ACM Communications 51*, 11 (2008), 75–84. 2

[STKF07] Schreck T., Tekusova T., Kohlhammer J., Fellner D. W.: Trajectory-based visual analysis of large financial time series data. *ACM Special Interest Group on Knowledge Discovery and Data Mining Explorer Newsletter 9*, 2 (2007), 30–37. 2

[TA03] Taskaya T., Ahmad K.: Bimodal visualisation: A financial trading case study. In *Proceedings of Seventh International Conference on Information Visualization* (2003), pp. 320–326. 2

[Tab] Tableau. http://www.tableausoftware.com, Accessed on Aug 2011. 2

[TLH10] Talbot J., Lin S., Hanrahan P.: An extension of Wilkinson's algorithm for positioning tick labels on axes. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1036–1043. 5

[Tuf90] Tufte E. R.: *Envisioning Information*. Graphics Press, CT, U.S.A., 1990. 3

[VvWvdL06] Vliegen R., van Wijk J. J., van der Linden E.-J.: Visualizing business data with generalized treemaps. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 789–796. 2

[War94] Ward M. O.: XmdvTool: Integrating multiple methods for visualizing multivariate data. In *Proceedings of IEEE conference on Visualization* (1994), pp. 326–333. 2

[WFR*07] Weaver C., Fyfe D., Robinson A., Holdsworth D., Peuquet D., MacEachren A. M.: Visual exploration and analysis of historic hotel visits. *Information Visualization 6*, 1 (2007), 89–103. 2

[Wil05] Wilkinson L.: *The Grammar of Graphics (Statistics and Computing)*. Springer-Verlag, New Jersey, U.S.A., 2005. 5

[ZNK08] Ziegler H., Nietzschmann T., Keim D. A.: Visual analytics on the financial market: Pixel-based analysis and comparison of long-term investments. In *Proceedings of International Conference on Information Visualisation* (2008), IEEE Computer Society, pp. 287–295. 2