Visualizing Time-Dependent Data in Multivariate Hierarchic Plots - Design and Evaluation of an Economic Application

Tatiana Tekušová
Fraunhofer Computer Graphics Institute
Darmstadt, Germany
tatiana.tekusova@igd.fraunhofer.de

Tobias Schreck
Technische Universität Darmstadt
Darmstadt, Germany
tschreck@gris.informatik.tu-darmstadt.de

Abstract

For successfully competing in a modern economy, large amounts of hierarchic time-dependent data need to be analyzed. As an example, one could consider the geographic composition of inflation in the European Union, or the revenue by product (sub) categories of a firm in the last month. Analysts wish to interpret the structure of the data not only at a single point in time, but examine the changes in the data categories through time. The analysts may need to consider additional dimensions to composition and time, such as the growth rate or profit rate. To reflect such analytic requirements, we have developed an interactive visualization of multi-dimensional, structured data taking the time dimension into account. The data are displayed in a three dimensional hierarchic circular or column plot. The time-dimension of the data is represented by animation. Our system provides interactive tools for the visual data analysis and variable set-up of the data display. For better orientation in the data space, we have enhanced the visualization with smooth transitions between different data selections in case of 3D hierarchic plots. The techniques presented can be applied to various data domains. A user study using European inflation data has shown the usefulness for effective economic analysis.

1. Introduction

In a modern economy, large amounts of hierarchic time-dependent data need to be analyzed. As an example, one could consider the (geographic) composition of consumer prices in the EU or the revenue by product (sub) categories of a company in the last month. Analysts may wish to investigate the structure of the data, e.g.: 'Which countries have predominantly contributed to higher consumer prices in the EU?' Moreover, they may need to examine the changes in the contribution of different components to an aggregate through time - 'Which consumer basket components have caused the higher price level in the last month, compared to the previous month?' The contribution of the items in the consumer basket to the overall rate of price level changes evolves over time as the price of each item fluctuates substantially from month to month. For example, a higher price of oil on the world market may significantly raise the price of fuel goods, or bad weather conditions may cause higher wheat prices. However, composition and time are not the only dimensions that need to be included in the analysis. Often, additional dimensions such as the pattern of inflation, or product profitability will need to be considered.

To reflect the data and analytic requirements, we have developed an interactive visualization of multi-dimensional structural data, which is able to display time variation in the data as well. Our system displays the data in a three dimensional hierarchic circular or column plot. We use animation to convey the data changes through time. The detailed time-dependent development of selected data items can be shown in an additional view. Our system provides interactive tools supporting the economic data analysis in a visual way. In addition, the users may set up the data display so that it best suits their analytical needs. We apply our system to an extensive real-world database of monthly European consumer price data from January 2000 to July 2007, taken from the Eurostat database [7], thereby illustrating its usefulness for effective visual analysis in a complex application area. The system can be used in other application areas, e.g., in the analysis of other economic or socio-demographic areas.

The remainder of this paper is structured as follows. In Section 2, we review relevant analytical tasks and data specifics of price analysis, and we describe the data used in our application. In Section 3, we recall important related work on visualization of time-dependent hierarchic data. In Section 4, we describe our visualization system and its functionality regarding analysis of European consumer price data. Section 5 provides results of a qualitative user study showing the usefulness of the system for economic analysis. Finally, Section 6 concludes.

2 Economic Analysis

In this section, we introduce the studied consumer price data and associated analytic tasks.

2.1 Consumer Price Data

Decision making in the economic and financial domain often relies on the analysis and evaluation of large amounts of time-dependent data in a timely way. The data often have a hierarchic structure. In this work, we address analysis of European consumer price data and its meta-data from the Eurostat database [7]. Eurostat reports the so called harmonised consumer price indices (HICPs), which measure the changes over time in the prices of consumer goods and services acquired, used or paid for by households. The index differentiates twelve main item categories ranging from food to clothing and education. Broken down to the lowest level of aggregation (5), the index for a country consists of 165 individual item series (see Figure 1, a).

The Eurostat database contains, among others, monthly data on the price index for each individual item category, data on the composition of the overall price index by item category (weights in the consumer basket), and the annual rate of change of the price series, the so-called rate of inflation for each component and the overall index. The inflation data is used for a wide variety of purposes, for example, as a guide for monetary policy of the European Central Bank (ECB), during wage negotiations for the indexation of earnings, or in public finance for calculating changes in national consumption or living standards [6]. We have downloaded the data for the European Union and each of its member states.

A Consumer Price Index is generally calculated as Laspeyres index [6]. A Laspeyres index for period t with reference to period 0 is the growth (from time 0 to time t) of a weighted average of prices p for every component c of consumer basket C, where the weight w for each item reflects the consumption of the item c in the consumer basket C (Equation 1).

$$I_{t_t,to} = \frac{\sum\limits_{c \in C} p_{c,t_t} \times w_{c,t_0}}{\sum\limits_{c \in C} p_{c,t_0} \times w_{c,t_0}} \tag{1}$$

The indices are often compiled also as a weighted arithmetic mean of component indices (Equation 2).

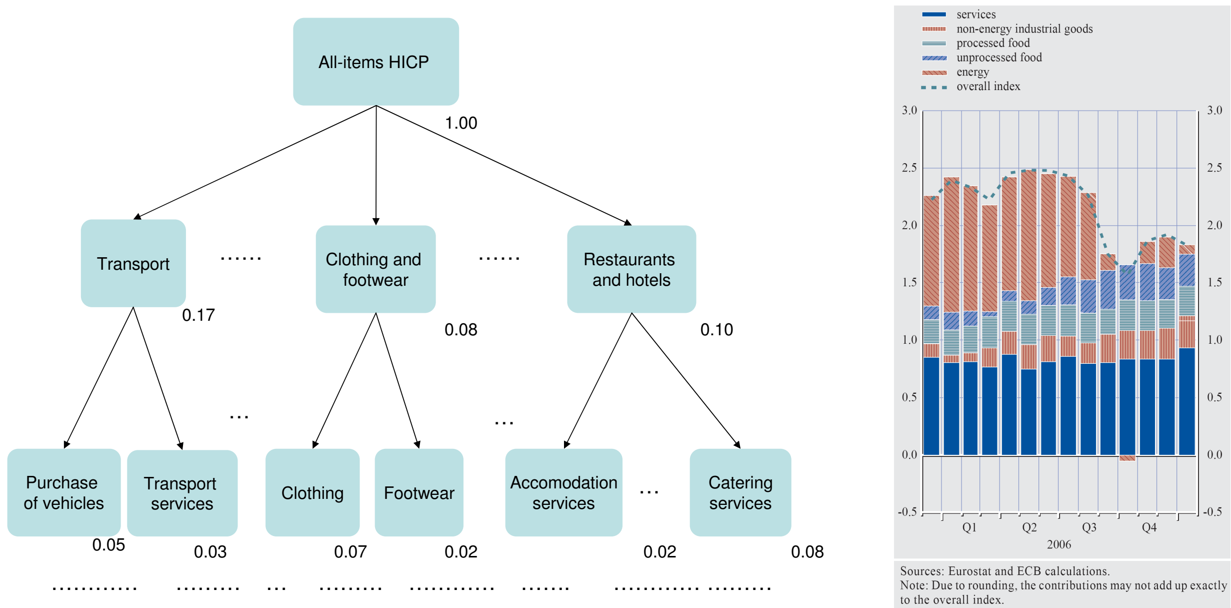
$$I_{t_t,t_0} = \sum_{c \in C} w_{c,t_0} \times I_{c,t_t,t_0} \tag{2}$$

For HICP, the weight of each item or category in the consumer basket may change every year and differ from country to country. Therefore, we need to be able to display these changes over time in the plot as well as to allow the user to compare the compositions of the price indices across countries.

2.2 Analytic Tasks

In economic terms, inflation is defined as the on-going increase in the price level. The analysis of inflationary developments therefore entails the need to *identify the co-movement in consumer prices across a broad set of item categories*. More specifically, it is important to distinguish between movements in the overall HICP index resulting from changes in the prices of any individual consumer goods and services, and movements in the HICP as a whole, reflecting a broadly based increase in the cost of living. For example, in recent years oil and other energy prices increased substantially, while on the whole, inflation in the EU remained low and stable. Generally, this could be due to a small weight of high inflation goods and services in the consumer basket, or due to a compensation by falling prices of other goods. Therefore, one of the main tasks in the analysis of inflation development is *to examine the composition of the inflation by its components*. In the current analysis, the development of the contributions to inflation by main categories is mainly visualized in a stacked column chart (see Figure 1, b). The chart shows the monthly development of the Euro area inflation over the period of the year 2006. The dotted line represents the total annual consumer price increase in this period, and the column segments represent the contribution (composite of weight and price increase) of a component to the total price increase. For example, the blue segment explains almost half of the total price increase in December 2006. However, this view does not allow for discrimination of whether this was caused by high inflation goods and services with small weight in the index, or by low inflation category with large weight. Moreover, it does not allow for displaying multiple levels of hierarchy of item categories (i.e. the composites of each category). This is supported by results of our user study, indicating that the main shortcoming of currently used visualizations "is the lack of presentational tools for multidimensional hierarchical problems [...] in a convenient, automatic way". Therefore, we have applied 3D hierarchic circular and column plots for visualization of inflation data (see Section 4). As the static hierarchic plot offers only display of inflation components regarding a single point in time, we have used animation to convey the data changes over time.

In recent years, economists have been increasingly interested in the calculation of so-called core inflation measures, which try to capture the broad-based nature of the inflation process. Such measures are calculated by deriving the median of price increases across the consumer basket or by calculating the mean of price changes trimmed at 75% or 90%



(a) Hierarchy of consumer basket items

(b) Contributions to HICP inflation

Figure 1. Illustration of a hierarchy of consumer basket items. Each node represents an item category. Item weights are given next to each category (a). Contributions to HICP inflation from main components (b; taken from ECB [5]).

quantile. Appropriate visualization methods can be used *to identify the items which fall out of the trimmed basket*. A display of a selected threshold would be of advantage for this task (see Section 4.2). Such threshold levels allow also for a quick identification of consumption items/categories whose inflation rate exceeds a set threshold (for example 2%) or is higher than a certain category. Similar tasks are relevant also to other economic and financial analysis areas, e.g., analysis of other macro-economic data, portfolio analysis, or analysis of product revenues.

3 Related Work

As the main contribution of the paper is the display of multivariate hierarchic time-dependent data, we concentrate on a review of literature for this type of data visualization. We mainly focus on relevant works applied to analysis of economic data. Although hierarchy visualization techniques, in general, are more widely applied in other areas of economic and financial analysis, from our survey it is apparent that only few systems apply these techniques to economic or financial data. A survey of techniques for visualization of tree graphs can be also found in [11].

Both Hao et. al. [10] and Schreck et. al. [18] concentrate on visualization of hierarchical financial time series. They use a modified treemap layout to show the hierarchical structure and time-varying values of prices of companies in a stock market. Treemaps are also used for the visualization of stock prices on the stock market (see Map of the Market [20]) and portfolio holdings (see Csallner et. al. [4]). Moreover, Ghoniem and Fekete [8] presented animated transition between two weight distributions in a treemap. Visualization of changes using treemaps has been recently presented by Tu and Shen [21]. Mansmann et. al. [15] propose an explorative framework for OLAP data to analyze data cubes of various data sources including financial data using graphs. Koenig et. al. [14] present a combination of DagMaps and Sugiyama Layout for navigation of hierarchical data on an example company subsidiary network. In [13], Keim et. al. visualize hierarchical data by so-called pixel bar charts.

In financial and economic applications, pie charts and stack column or area charts (see Section 2.2) are the techniques used for display of structural information of economic time series. As our system uses a circular presentation of data, we further concentrate on methods for hierarchic data presentation in a circle or column plot. Keim et. al. [12] present a combination of hierarchic visualization and circular layout techniques based on circle segments [2] for time dependent data. The hierarchic circle plots were also used in [12] and [3]. Although the cascading semi-circular disks were already presented in [1], only recently an animated transition between different focus views was

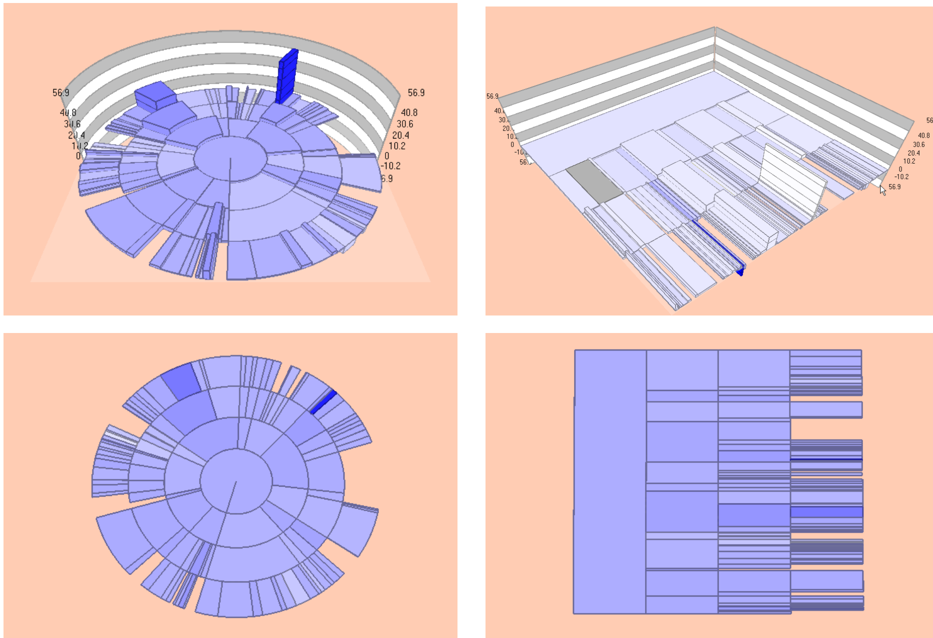


Figure 2. Circular data visualization (left) and column data visualization (right). 3D views are shown in the top row, 2D views are shown in the bottom row.

presented in [16]. Additionally, an application of circular graphs to business process analysis is shown by Hao et. al. [9].

Hierarchic circular plots have until now rarely been used for providing multiple information in 3D. In 2D, Keim [12] presents hierarchic plots with color as an additional dimension. Recently, 3D representation has been used in [17]. At the same time, multiple 3D treemap designs were presented in [19]. However, they often do not allow for display of time-dependent data in addition to the hierarchic data structure. To the best of our knowledge, there are no applications of 3D circular or column plots for multivariate time-dependent data in the economic or financial field.

4 Description of the System

To reflect the analytic requirements, we developed an interactive visualization of multi-dimensional structural time-dependent data. Our system displays the data in a three dimensional hierarchic circular or column plot. We use animation and time-series plots to show time-dependent changes in the data. Furthermore, our system provides interactive tools for analysis support. We apply our system on an extensive real-world database of monthly European inflation data from January 2000 to July 2007 (see Section 2.1), illustrating its usefulness for effective visual analysis in a complex application area.

4.1 3D Hierarchic Data Visualization

Our first prototype of the interactive visual analysis system displays the multi-dimensional hierarchic data at a certain point in time in a three dimensional hierarchic circular or column plot. We decided to use these graphical representations as they are often used by and familiar to financial analysts, so the users are able to interpret the displayed data very easily. The Treemap technique, another well-known hierarchic data visualization, has beeen deemed not suitable as it does not scale for multiple dimensions and, owing to the nesting-based layout of the method, overplots the parent nodes with their children nodes. Therefore, the values of the parents are not directly visible to the user. Moreover, the height dimension is difficult to convey without blending.

In the circle view (see Figure 2 left), the center of the plot represents the root of the hierarchy ('all items HICP'). Each layer of the circle represents one layer of the item hierarchy. For each node, the children are displayed as a segment in the next layer, bordering with the parent. The sweep of the child segment is calculated as a proportion of the weight of the child to the weight of the parent multiplied by the sweep of the parent. In our case, the item weight of the subcategory is proportional to the weight of the parent category in the consumer basket. The height of the segments is determined by another dimension of the data, the annual inflation rate of the item. As the rate of consumer price change can be negative, also negative heights are supported. We

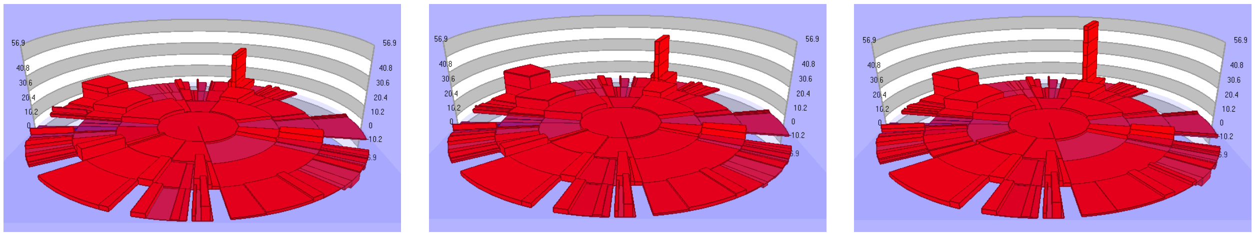


Figure 3. Smooth data animation over time showing transition between two time points.

have included colored gridlines, which are reflected also in the circle parts so that the analyst can better compare the data values encoded as height. The color of the graph segments is dependent on a chosen additional variable (e.g., the monthly price index).

The user can also choose to view the data in a column plot (see Figure 2 right) where the top column represents the root of the hierarchy, and the children are laid out similar to the circular view. In general, the data visualization is adjustable to user preferences. The analysts may choose between a 2D and a 3D view, they can turn on or off coloring of the segments, and switch between column and circular layout.

As users need to view the data changes over time, we have implemented animation over a selected time period (see Figure 3). The linear interpolation of the data shows smooth data changes over time. As all data dimensions (item values) change over time, we have to interpolate both width and height of each segment of the graph. In addition, in case that a data dimension is represented by the color, also segment colors change smoothly over time. The main advantage of animation is that users can see the time dependent changes in all dimensions, i.e structure (item weights) and values (inflation rates), simultaneously. Based on a user suggestion, we added a visualization of time development of selected data items in a separate view (see Figure 4).

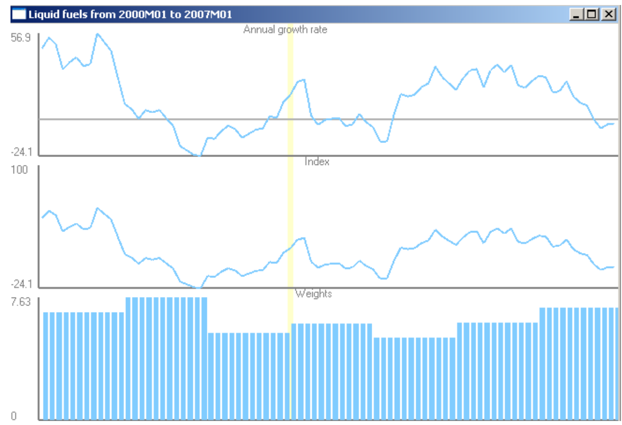


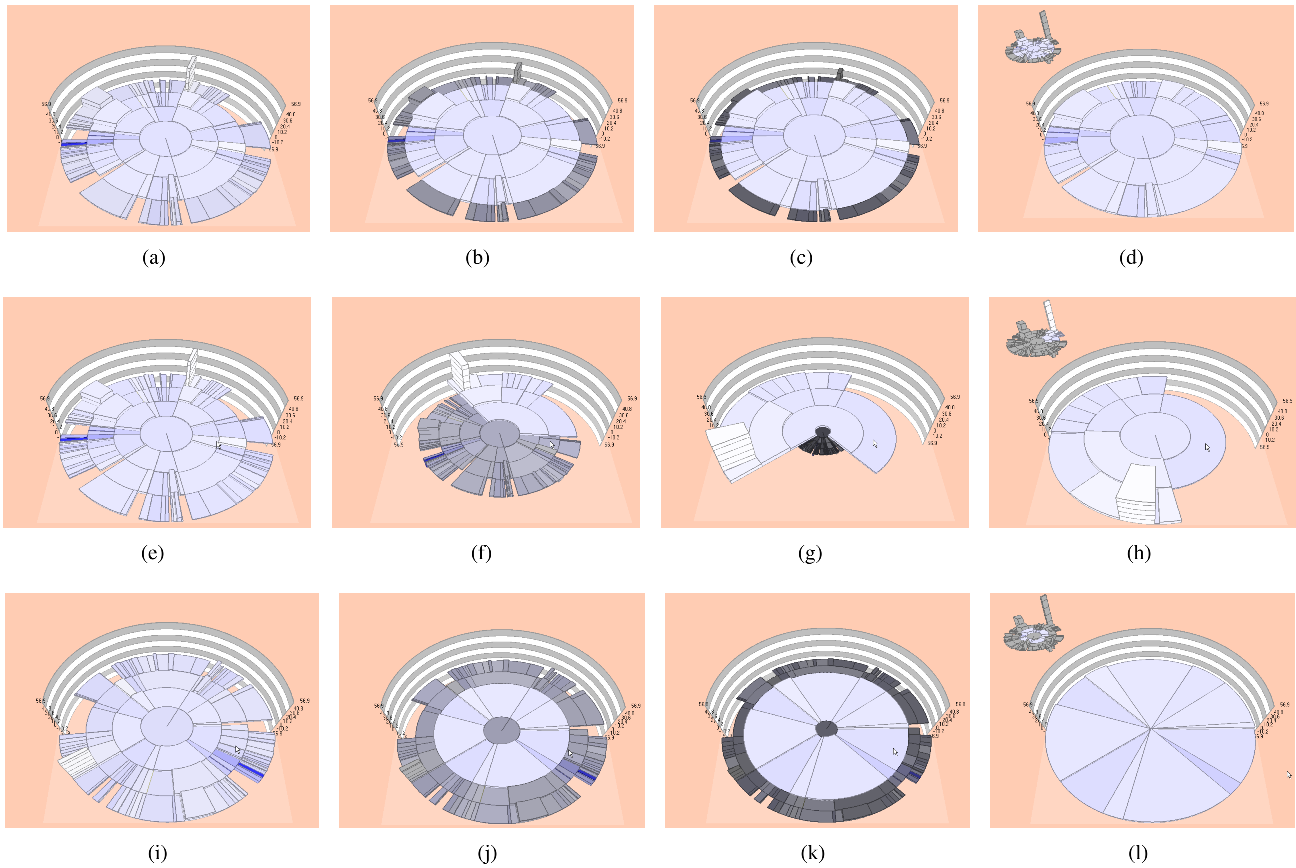
Figure 4. Time development of a selected data item. The yellow highlight in the background indicates the current time period displayed in the main hierarchic plot.

4.2 User Interaction Techniques

The system functionalities have been developed to support the analytical needs of economic and financial analysts, who need to examine multi-dimensional structural data over time. In addition to the data visualization and data animation described above, interactive features such as focus and zoom, threshold and details on demand are also incorporated in the system. To better recognize structures in the data, users are given the possibility to zoom in and out, to rescale the Y axis, and to rotate the plot. As users may wish to receive more information on the visualized data, data legends and further details on demand are provided in a pop-up window. The layout of data legends has been inspired by [3] - the size of the label text is determined by the weight of the item in the basket. A number of item selection methods are implemented:

- Select items up to a specified depth in the hierarchy;
- Selected items (and thereby, also its sub-items) in the basket;
- Select a given hierarchy level.

These selection methods allow for an explorative data analysis in particular if an analyst wishes to examine only a certain category of the consumer basket, e.g., food. For focusing, we have developed an algorithm for animated transition between different foci which improves the work of Pula [16]. He uses different techniques for setting focus in column and pie charts and finds animation the most suitable method. We have enhanced the animation technique for hierarchic column and circle plots in three dimensions. The algorithm concurrently blends out the items of the graph which are out of focus and then iteratively minimizes the width and height of those items while enlarging the area of the focused part of the graph. Additionally, the user can



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 5. (a-d) Transition steps during change of focus on selected maximum hierarchy depth. (e-h) Transition steps during change of focus on selected part of the category hierarchy. (i-l) Transition steps during change of focus on selected hierarchy level.

see the full graph with highlighted focused area in the left upper corner of the display. This feature allows the user to better orientate in the data space. Figure 5 illustrates the implemented selection and animation methods.

As analysts also often need to answer the question ‘Which items have higher inflation as item X?’, we implemented a threshold function. A benchmark can be set to a selected item, then a semi-transparent threshold plane is shown (see Figure 6). This allows a quick identification of the items which are above the set benchmark, and support possible discovery of unknown patterns in the data, especially in combination with the animation feature. This also helps to mitigate a common problem of three dimensional plots, namely, that users have difficulties to assess or compare precisely the height of the segments in the 3D environment.

5 User Survey and Evaluation

We have also conducted a qualitative user survey to evaluate the techniques for visualization of hierarchic multivariate data with time dimension as presented in this paper (see Section 4). We have asked ten potential users, all professionals in the financial data analysis domain, to asses the usability of the proposed visualization system for their analytical work. In the detailed survey, the participants were asked to evaluate the different options of data display - compare 2D with 3D display, and assess the advantage of encoding additional variables using color, and using animation to reflect time-dependent data changes. Next, we informally present the main findings we obtained from our survey.

User Background. The analysis background of the surveyed users covers various fields of economic and financial analysis including internal financial data, macro-economic data, monetary data and financial market data. In general,

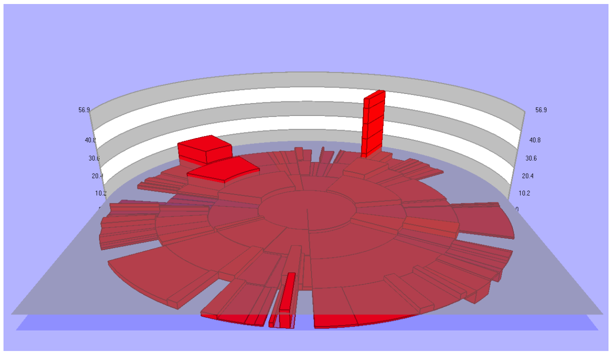


Figure 6. Threshold visualization

they analyze and visualize data on a daily or weekly basis. *Time series data* is analyzed on a very frequent basis. The analysis and visualization of *hierarchic data* is evenly distributed from frequent (daily) to seldom (several times per year). For their analysis, the participants regularly employ *pie and stacked column charts*. None of the participants routinely uses *hierarchic plots* for multiple levels of aggregation. The reason for this, according to the survey, is that the users are not sufficiently familiar with the latter techniques. The major disadvantages of the currently used visual analysis tools were stated to be (a) for pie charts: the lack of showing data over time and (b) for stacked column charts: the lack of showing multiple aggregation levels, and a tendency to over plotting on large data sets.

Assessment of the Visualization. Most participants preferred the *3D display* because it immediately allows to visually filter different magnitudes of developments in addition to the color. Additionally, it is a "more interesting visualization of data". Participants who prefer *2D display* argued that the 3D tends to be more complex and therefore may be difficult to interpret for users unfamiliar with the display. The 3D display is preferred by experienced users for more complicated tasks with multiple data dimensions.

Comparing the *column and circular plot*, most users preferred the circular display. However, the suitability of circular or column plot was stated to depend on the type of data analyzed and the task at hand. Experienced users, who spend more time with data analysis claimed that the column plot is better suitable for more accurate data analysis as the amount of data distribution (e.g. weights) is better readable and comparable in a column plot. As a consequence, the option to switch between layouts is deemed necessary.

The usage of *color* for encoding a variable in the display is preferred by all users. Color was stated to be beneficial in enhancing the differences in the variables, as the values are easily estimated and exceptions in data values are apparent immediately. The possibility to change color schemes (different color scales) or to color-code the data by category was deemed necessary for some analytical tasks.

In general, *animation* can be a good communication tool for data development over time [22] and therefore, was expected to be useful for many tasks, e.g., identification of volatile subcomponents, or sub-components with recurring patterns. According to the surveyed users, animation is useful to get a rough assessment of the data development, while for more accurate analysis of time series, line charts are better suited. We therefore included an option to show selected time series in an additional detail view. The users also voiced that interaction facilities to select time periods, and control the speed of the animation would be desirable.

Assessment of the Interaction Facilities. The participants were also asked to assess usefulness of the interactive features in the visualization for their analytic work. In general, the interactive functionality is deemed very useful allowing the users to analyze the data more efficiently from different view points. *Focussing functions* (selection of maximum hierarchy depth, focus on the subcategories of a selected item and focus on a single hierarchy level) help to undertake a detailed analysis of the composition of the data and reduce data complexity. *Setting up a threshold* permits users to evaluate the data according to a criterion. *Highlighting of selected data items* is good for localizing the outcome of the visual search, for example the identified outliers.

Summary. The main practical advantage of the proposed visualization tool was seen in its support to quickly get a complete overview of developments in the total data and its components, allowing a summary over all levels of aggregation and potential interrelationships. Moreover, it was recognized by the survey participants that the display of multidimensional data allows for a more comprehensive data analysis in a single view. It was stated that the tool can more adequately present data along structure and time dimension, than common tools currently in usage by the survey participants. The animation feature was judged useful for assessing general time dependent developments in the data.

6 Conclusions and Future Work

We presented an effective visualization of time-dependent structural data. To reflect the analytic requirements, our system displays the data in a three dimensional hierarchic circular or column plot. The time-dimension of the data is represented by animation. Our system provides interaction facilities for explorative data analysis. The system allows to examine the data from various perspectives, thus enhancing the analytic process. We applied our system to real European inflation data, showing its usefulness for effective visual analysis. An evaluation of the system by expert users has shown that our approach significantly improves over the functionality offered by many tools currently in use in the economic data analysis domain. Specifically, the visualization of time-variation in addition to the structural information over multiple aggregation levels has been deemed very valuable. An animation feature has been implemented which shows very useful for assessing general time dependent developments in the data.

As an improvement of the system, we are considering the incorporation of time-varying data visualization into a static view similar to the one presented in [10]. This feature would allow for better comparison of the developments in time-series data. Following user suggestions, we like to extend the analytic functionalities of the system, e.g., by visual comparison of data from various countries, or with additional economic data analysis features.

The visualization and analysis techniques described in this paper can be applied to various fields of economic and financial analysis. E.g., it is also expected to be useful for the analysis of portfolio composition and analysis of financial markets (bonds, yields, or exchange rates). In business management, it can be used for analysis of sales data (turnover and profits of various product (sub-)categories). In macroeconomic analysis, for example, development of the gross domestic product, or money demand composition and growth can be studied with the help of our system. Potential users also suggested applications in data warehousing (e.g. data from management information systems) or in the socio-economic domain.

Acknowledgements

The authors thank all participants of the user study for their valuable feedback and suggestions provided.

References

[1] K. Andrews and H. Heidegger. Information slices: Visualizing and exploring large hierarchies using cascading, semicircular discs. In *IEEE Symposium on Information Visualization*, 1998. Late Breaking Hot Topic Paper.

[2] M. Ankerst, D. A. Keim, and H.-P. Kriegel. Circle segments: A technique for visually exploring large multidimensional data sets. In *Proceedings of the IEEE Visualization*, 1996.

[3] C. Collins. Docuburst: Document content visualization using language structure. In *IEEE Symposium on Information Visualization*, pages 112–113, 2006. Poster Paper.

[4] C. Csallner, M. Handte, O. Lehmann, and J. Stasko. Fundexplorer: Supporting the diversification of mutual fund portfolios using context treemaps. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 203–208, 2003.

[5] European Central Bank. *Annual Report 2006*.

[6] Eurostat. *Harmonized Indices of Consumer Prices (HICPs) - a short guide for users*. European Commission, 2004.

[7] Eurostat. *http://epp.eurostat.ec.europa.eu*. European Commission, 2007.

[8] M. Ghoniem and J.-D. Fekete. Animating treemaps. In *Proceedings of 18th HCIL Symposium - Workshop on Treemap Implementations and Applications*, 2001.

[9] M. Hao, D. A. Keim, U. Dayal, and J. Schneidewind. Business process impact visualization and anomaly detection. *Information Visualization*, 5(1):15–27, 2006.

[10] M. C. Hao, U. Dayal, D. A. Keim, and T. Schreck. Importance-driven visualization layouts for large time series data. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 203– 210, 2005.

[11] I. Herman. Graph visualisation and navigation in information visualisation: A survey. *IEEE Transactions On Visualisation And Computer Graphics*, 6(1):24– 43, Jan. 2000.

[12] D. A. Keim, H. Barro, C. Panse, J. Schneidewind, and M. Sips. Exploring and visualizing the history of information visualization. In *IEEE Symposium on Information Visualization*, 2004. Contest Poster.

[13] D. A. Keim, M. C. Hao, and U. Dayal. Hierarchical pixel bar charts. *IEEE Transactions On Visualisation And Computer Graphics*, 8(3):255–269, July 2002.

[14] P.-Y. Koenig, G. Melancon, C. Bohan, and B. Gautier. Combining dagmaps and sugiyama layout for the navigation of hierarchical data. In *Proceedings of the International Conference Information Visualization*, pages 447–452, 2007.

[15] S. Mansmann, F. Mansmann, M. H. Scholl, and D. A. Keim. Hierarchy-driven exploration of multidimensional data cubes. In *GI-Fachtagung fr Datenbanksysteme in Business, Technologie und Web*, 2007.

[16] K. Pulo. Navani: Navigating large-scale visualisations with animated transitions. In *Proceedings of the International Conference Information Visualization*, pages 271–276, 2007.

[17] M. Schedl, P. Knees, G. Widmer, K. Seyerlehner, and T. Pohle. Browsing the web using stacked three-dimensional sunbursts to visualize term co-occurrences and multimedia content. In *Poster Compendium of IEEE Visualization*, 2007.

[18] T. Schreck, D. A. Keim, and S. Mansmann. Regular treemap layouts for visual analysis of hierarchical data. In *Proceedings of the Spring Conference on Computer Graphics*, 2006.

[19] H.-J. Schulz, M. Luboschik, and H. Schumann. Exploration of the 3D treemap design space. In *Poster Compendium of IEEE Information Visualization Conference*, 2007.

[20] SmartMoney. Map of the market. www.smartmoney.-com/marketmap/.

[21] Y. Tu and H.-W. Shen. Visualizing changes of hierarchical data using treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1286–1293, 2007.

[22] B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, Oct. 2002.