BinX: Dynamic Exploration of Time Series Datasets Across Aggregation Levels

Lior Berry*
University of British Columbia

Tamara Munzner†
University of British Columbia

ABSTRACT

Many fields of study produce time series datasets, and both the size and number of theses datasets are increasing rapidly due to the improvement of data accumulation methods such as small, cheap sensors and routine logging of events. Humans often fail to comprehend the structure of a long time series dataset because of the overwhelming amount of data and the range of different time scales at which there may be meaningful patterns. BinX is an interactive tool that provides dynamic visualization and manipulation of long time series datasets. The dataset is visualized through user controlled aggregation, augmented by various information visualization techniques.

CR Categories:
I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

Keywords: time series, aggregation, detail, overview, marking

1 INTRODUCTION

BinX was developed to provide dynamic visualization and manipulation of long time series datasets. It assists a human in gaining better understanding of the data and underlying patterns, without requiring many views or applying complex transformations. The dataset is handled using the aggregation technique of binning. Aggregation has been shown to be an effective method for dealing with large data sets [1]. Still, it can often be misleading, disorienting or limited by the display size. BinX alleviates these problems.

Data: The example time series dataset that we show here comprises the daily exchange rates of several currencies sampled daily over a period of 15 years, so the size is slightly beyond 5000 data points. Despite its medium size, it cannot fit the screen in a single view.

Tasks: The primary task that BinX was designed to support is the analysis of a single currency exchange rate behavior over time. The secondary task is comparing two currencies to detect correlations. These tasks build on several subtasks: obtaining a global view for aggregation at different levels of detail, detecting time periods with interesting features that should be explored further, finding global and local minimma and maxima, and monitoring a specific period at different levels of detail.

2 DYNAMIC TIME SERIES VISUALIZATION COMPONENT

The modular *dynamic time series visualization component*, or DTVC, is the cardinal building block in the BinX system. A DTVC, as shown in Figure 1, presents one or two time series datasets drawn as a line graph at a controllable aggregation level, along with a choice of scale for the time axis and the ability to show other miscellaneous information on the binned data. The component was designed to expose a clean public API so that it can be integrated as a black box into applications. The rest of this section contains a short description of its functionality and the visualization techniques it incorporates.

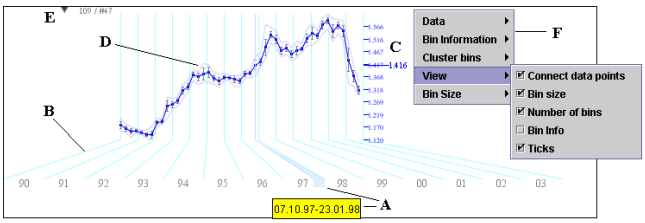

Figure 1: A dynamic time series visualization component, or DTVC, is dynamically controlled through the menu (F). Several visual cues are visible: A) bin period highlight, B) trapeze timelines, C) bin value, D) bin minima, maxima and standard deviation, and E) aggregation level control.

Aggregation and Time Scale Controls The main way to control the DTVC is by dynamically changing the aggregation level; that is, changing the number of the bins used to process the data. The lower section of the DTVC is a horizontal axis that shows the entire range of time for the currently loaded time series dataset, with labelled vertical "timelines" that act as tick marks to demarcate bin boundaries. When the user chooses to use a small number of bins for higher level overview, the changing upper section will narrow, as in Figures 1 and 2 Top. The anchoring timelines always continue to the upper section for a trapeze-like effect, where the width of the trapeze and the inward or outward slant of the lines connecting the top and bottom DTVC sections shows the aggregation level. Figure 2 Bottom shows that the trapeze widens when the user chooses to see more lower level details by increasing the number of bins. At low levels of aggregation, lines slant outward and many end outside the window, giving a visual indication that more bins could be seen by panning left or right. The number of bins corresponds to the time scale, as shown by the horizontal spacing between timelines, is also controllable.

Animation and Navigation We always provide smooth transitions to help the user visually track elements during changes. The aggregation level changes when the user changes the control knob or selects a predefined aggregation level (week, month etc.). Selecting a time period to be the focus of attention also triggers a smooth transition.

Marking Any time period can be marked by dragging out a rectangular box that remains visible as a colored translucent box, as in Figure 2 Top. Timeboxes were introduced by Hochheiser [2, 3] to support querying. With the DTVC, these boxes dynamically update as the aggregation changes, supporting the exploration of large datasets. The user can also turn on marks that show the minimum value, the maximum value, and the standard deviation for each bin. When the mouse is over a bin, the corresponding section of the

*e-mail: berry@cs.ubc.ca
†e-mail:tmm@cs.ubc.ca

bottom scale is highlighted and the full date range information is shown in a yellow box.

3 BinX Framework

The BinX application, shown in Figure 2, contains two DTVC components that are integrated via the DTVC API with other views that support exploration of the dataset via several other visualization options. The software was written in Java and is available at `http://www.cs.ubc.ca/~tmm/courses/cpsc533c-04-spr/projects.html#berry`.

Detail and overview The two DTVCs can be independent, so that each one has a different focus, aggregation level and information display. Another important navigation mode is linked navigation and marking across the two DTVCs, controllable via a toggle switch. When linking is selected, navigating or marking in one DTVC will result in a matching navigation or marking in the other, but each DTVC can still control its own aggregation level and information display. The simultaneous highlighting of the binned period under the mouse in both displays provides useful context, as shown in Figure 2. The possibility of distinct levels of aggregation in two linked views allows a flexible detail + overview exploration.

Clustering BinX offers a clustering mechanism that classifies the time series dataset, as introduced by van Wijk [4]. The classification can be dynamically updated as the aggregation level changes, and classification of each bin is shown in the visualization by superimposing a small color-coded box directly on the line graph, as shown in Figure 2 Bottom. A separate cluster viewer window can be opened to show a plot of the values of all clusters.

3.1 Scatter Plot View and Brushing

Figure 2 also shows that scatter plot views are available that can present the direct correlation between two binned currencies, regardless of time. This view is linked to the active DTVC view, so that changing the aggregation level will update the scatter plot. The user can also use the marking mechanism to depict correlation only in selected time periods.

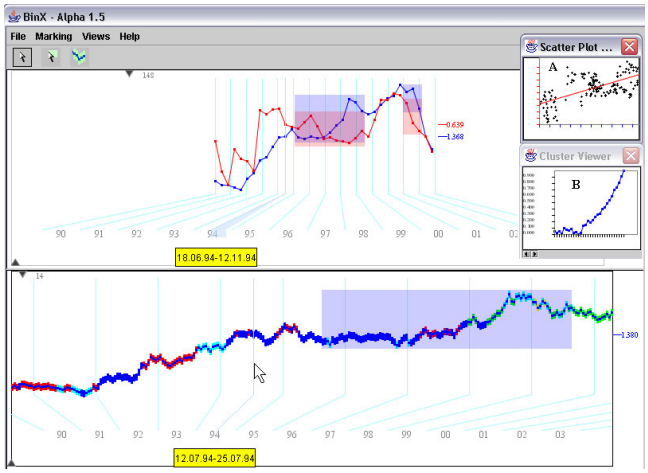

Figure 2: Two DTVCs support various exploration techniques. The views are linked, so marking, navigating and mouseover highlighting in one DTVC will be reflected in the second. The top DTVC shows two currencies as superimposed line graphs, using a small number of bins for a high level of aggregation. The bottom DTVC uses a low level of aggregation so that more detailed information is visible, and it shows the results of clustering the bins by color coded boxes of red, blue, green, or cyan directly superimposed on the line graph. Other view windows: A. Scatter plot view, B. Cluster viewer

4 Discussion

The dynamic aggregated "trapeze" view has proved to be very useful when used on the currency data sets. The animated navigation options and indication of off-screen data that can be easily shown enable an effective and smooth exploration, focusing on the interesting periods. BinX enables a flexible mix and match of exploration techniques: linked vs. unlinked navigation, displaying one vs. two currencies, the display of different bin information in each DTVC, and more. These can be effectively adapted to fit the nature of the data as well as the user preferences and needs. The marking mechanism, combined with the aggregation level control, results in a powerful capability to manipulate long periods. Finally, integrating visual cues on the binned data greatly helps in preventing false conclusions, because "noisy" bins can be detected by using the min/max or standard deviation cues. Finally, clustering assists in comparing periods.

An early prototype tested a spiral view where changing the size of the spiral altered the aggregation level. The fundamental problem with this approach was that when the display space is insufficient for all data points, there was no natural visual extension to the model. In our final version, we solve this problem with the "trapeze" view that provides navigation and cues for off screen data. Nevertheless, further investigation of how to combine dynamic aggregation level with period length control may better support the detection of cyclic patterns.

Future Work BinX has not yet been tested with data sets larger than 50,000 data points. Although we conjecture that the visualization principles will still hold for larger datasets, further optimization may be required to improve performance. It would also be interesting to explore whether these techniques are suitable for use with small display areas by adapting DTVC for handheld use.

5 Conclusion

We have introduced BinX, a visualization tool for dynamic exploration of long time series datasets. Its fundamental navigation metaphor is to change the number of bins used to create the display for smooth transitions between levels of aggregation. We integrate several visual cues in the display to provide information without clutter, and also provide clustering support and scatterplot views. Our experiments with analyzing currency exchange rates datasets have shown the strength of the flexibility to have two dynamic time series visualization components that can be linked or unlinked in several ways to accomodate the needs of the user and the nature of the data. The modular visualization components were designed to be reusable by any other applications where there is a need for analyzing time series datasets.

References

[1] Mei C. Chuah and Steven F. Roth. Dynamic aggregation with circular visual designs. In *Proceedings of Information Visualization, IEEE*, volume 4, pages 124–125, 1998.

[2] H. Hochheiser and B. Shneiderman. Visual specification of queries for finding patterns in time-series data. In *Proc. Discovery Science*, 2001. Univ. Maryland CS Tech. Report CS-TR-4326, UMIACS-TR-2001-25.

[3] H. Hochheiser and B Shneiderman. Visual queries for finding patterns in time series data. University of Maryland, Computer Science Dept. Tech Report #CS-TR-4365, UMIACS-TR-2002-45, May 2002.

[4] J. J. van Wijk and E. R. van Selow. Cluster and calendar-based visualization of time series data. In *Proc. IEEE Symposium on Information Visualization*, pages 4–9, 1999.