VASTvis - Visual Analytics with Multiple Coordinated Views

VAST 2010 Mini Challenge 2 Award: Good Analytic Process and Explanation

Rick Walker† Llyr ap Cenydd‡ Serban Pop‡ Jonathan C. Roberts§

School of Computer Science
Bangor University

ABSTRACT
We present a multiple coordinated view system VASTvis with a central view and several mini-views around the side. Each view is coordinated together, where operations are linked between windows. Furthermore, the user can perform sequential queries on the data through a series of brushing operations. The tool was used to analyse the VAST 2010 Mini Challenge 2 dataset of Hospitalisation Records. In this article we describe how the tool was used to characterise the Pandemic Spread.

Keywords: Visual Analytics, Coordinated Multiple Views, VAST 2010, information visualization, pandemic visualization

1 INTRODUCTION
The scenario of the VAST Challenge 2010 dataset was to identify and characterise a pandemic spread over several countries given hospitalisation record data. The datasets for the challenge were synthetically created, and included hospital admittance and death records for specific cities. There were two overarching tasks: to 'analyse the records you have been given to characterize the spread of the disease' and 'compare the outbreak across cities'.

The software developed to help in our analysis was written from scratch over the competition period and the direction of its development was strongly influenced by the corresponding progress of our analytical inquiry. Our aim was to develop a highly interactive visualization tool that would enable an analyst to not only answer the given questions but to allow them to explore the data. It was important that every visualization was simultaneously visible. Thus, the overarching principle of VASTvis was to use the concept of multiple coordinated views, where interaction in one view is automatically shared with that of other views [4]. Our final design provides a central focus-view with several smaller visualizations around the side.

2 ANALYTICAL PROCESS
As with many visual analytics problems, in our work the detailed questions evolved throughout the project as new factors were discovered. For the development of VASTvis we acted as both analysts and tool builders. Weekly in-person meetings occurred plus remote discussions via Google Wave [1], which is an online communication and collaboration tool. Both forms of communication were used to discuss analytical results and to determine the additional questions that we needed to answer to progress in the analysis. The focus of tool development over the following week was then to develop methods to answer those questions. This development cycle followed a Scrum methodology. We found the use of Google Wave important. It provided us with a way to collate developing questions, screen-shots of the code and other correspondence in one place. It also provided an archive of our investigations and development.

While this was a labour-intensive process it had the effect of keeping a tight focus on the end goals of the development. Success or failure each week was a matter of whether the tool would now allow us to answer the questions posed previously. If so, then the process could advance, and if not, then the code sprint would continue. The rapid integration of new functionality meant that the final tool bore little resemblance to early designs, but is undoubtedly more appropriate for the task. In the remainder of this paper, we highlight some of the features developed in this manner.

VASTvis was developed using Processing [3], which allows the visualizations produced to be readily stored in a vector format. Thus, we designed VASTvis to output any view as a document in vector format (which can be stored in a PDF or SVG document). Consequently snapshots of the visual display can be stored and easily embedded into publications.

3 MULTIPLE VIEWS
Our first questions were those most easily answered, such as "what is the distribution of deaths in Barcelona over the given time period?". Several static graphs were created of this data. These static views gave us an idea of the scale of the epidemic and provided some insight into the death distribution. However, as our analysis progressed, it became important to see multiple variables simultaneously. It was thus necessary to show age, mortality rate, symptoms and days from admission to death. Depicting only one view at a time risked missing some visual connections between these variables. However, displaying all of these graphs together generated a cluttered screen and it became hard to determine the focus of the visualization. This led us to move to a focus+context design solution.

VASTvis is shown in Figure 1. Any view can be selected to be the centre-view, with the remaining views tiled around the edge of the display and kept up to date. The centre view is larger and enables the user to focus on the information in one view while seeing how the information relates to others. When a view is selected to be the centre view it is smoothly animated back to its edge position. Opacity is used as a perceptual cue to indicate if the view is in the process of updating through a query. Together, these help to maintain visual continuity when each view is swapped in or out, and when new query results arrive.

4 VISUAL QUERYING
The data was stored in an SQLite database, which uses a self contained database file that can be readily exchanged between users. Performing frequency analysis revealed that there were duplicates, misspellings and abbreviations in the syndrome data. We cleaned the data and grouped like symptoms together by creating a synonym list. This data was then stored in the database.

The clicking, dragging and brushing operations create SQL queries that are run on the database and subsequently visualized.
Figure 1: VASTvis is shown demonstrating several views: (a) geographical view for city selection (b) deaths by age (c) cumulative deaths by date (d) deaths by date (e) statistical information on the centre view (f) deaths by time from admission to death (g) mortality rate by date (h) deaths by symptom. The picture depicts data from patients with one or more of the six principle symptoms, admitted between 26th April and 11th June, who died on the eighth day after admission, with this filtering performed exclusively through selection on the available views.

Figure 2: Statistical summary data on deaths due to the pandemic is shown on the box plot view, which updates as the focused view is changed.

5 Statistical Overlays

The visual queries and linked brushing enables the user to achieve side-by-side comparison, but we needed a way to compare the different views more accurately. We used a percentile-based approach: since some of the distributions were bimodal, using mean and standard deviation would be misleading. This statistical information was shown through box plots as in Figure 2 and in overlays on each view.

6 Conclusions

VASTvis was intensely developed through a Scrum-like collaborative design process. This proved to be an effective methodology and the answers thus obtained were accurate. The design ideas (especially the side-views and interaction) were successful and these techniques could be adapted for other datasets. Further implementation details are available at: http://cvev.bangor.ac.uk/VAST2010/

Acknowledgements

The authors wish to thank Liam Kurmos and Jess Rowlands for their input in early discussions.

References