Comprehending City Economics from Heterogeneous Data

Rainer Splechtna*
VRVis Research Center, Vienna, Austria
Thomas Hulka†
VRVis Research Center, Vienna, Austria
Disha Sardana‡
Virginia Tech, Blacksburg, Virginia, United States
Nikitha Donekal Chandrashekar§
Virginia Tech, Blacksburg, Virginia, United States
Denis Gračanin¶
Virginia Tech, Blacksburg, Virginia, United States
Krešimir Matković‖
VRVis Research Center, Vienna, Austria

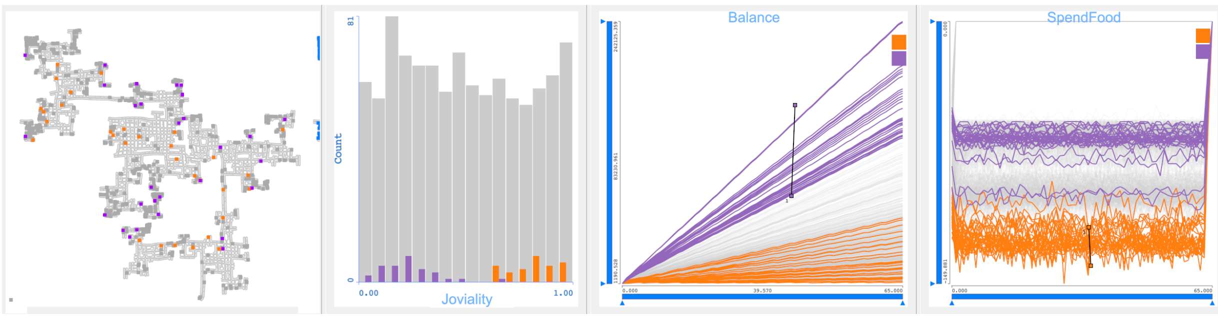

Figure 1: The user brushes 31 participants with the highest amount of money spent on food (orange brush) and compares them with the second selection of 30 participants with the highest money balance (purple brush). Those who spend lot on food have low account balance and high joviality, and those who have a high balance, have low joviality and do not spend a lot on food. Both groups of particiopants live all across the city.

Abstract

Visualizing complex, heterogeneous data that includes spatial information, tracking data of citizens, as well as rich data about business and citizens over a long period of time is a difficult problem. We describe how we deployed a coordinated multiple view system to explore and comprehend such data. The system is developed as a response to the IEEE VAST 2022 Challenge 3. We briefly demonstrate the power of the developed system using three examples that use a citizen-centric and a pub-centric data set.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Visual Analytics

1 Introduction

In this paper we describe how we approach the IEEE VAST Challenge Three: *Economic*. Since the tasks are of exploratory nature, we rely on an interactive solution. We extend our existing coordinated multiple views solution—ComVis [1]—in order to provide the necessary views and interaction for the predefined analysis tasks. ComVis is a stand-alone application developed in C++ which uses QT for the user interface and OpenGL for the visualization. In addition we use several F# and Python scripts to preprocess the data. The data is provided in several files and we need different data sets depending on the analysis tasks. As we are analyzing spatial data, the newly developed map view is used in all configurations. In the following, we briefly describe how the newly developed map view combined with standard views, such as histograms or curve views, for example, is used to solve different tasks.

*e-mail: splechtna@vrvis.at
†e-mail: hulka@vrvis.at
‡e-mail: dishas9@vt.edu
§e-mail: nikitha@vt.edu
¶e-mail: gracanin@vt.edu
‖e-mail: matkovic@vrvis.at

2 Data and Views

The data for the VAST challenge is divided in several files. We preprocess the data to generate table like structures, where each row represents a record and each column an attribute. In contrast to a common data structure where attributes can be numeric or categorical values, we allow also curves, i.e. a sequence of (x, y) pairs to be an attribute as well. Depending on the analysis task we create different data tables.

If we are interested in the individual citizens we create a record for each citizen. The citizen-centric data set contains attributes, such as, apartment location, household size, education level, interest group.... In addition to these scalar attributes which are easily extracted from the provided data files, we also create time series attributes, such as spending for education, recreation, or account balance, for example. This spatial attribute has a sequence of pairs of time and amount spent for education.

For the business-centric tasks we create a data set where each record corresponds to a business and provides scalar attributes (number of guests, e.g.) and time series attributes which show how income, number of visits, or anything else that changes over time.

We use a coordinated multiple views system which supports basic views (histogram, scatter plot, curve view, etc.). In order to support the analysis of spatial data we have added the map view. It shows buildings, apartments, pubs, ..., as provided in the data. Each business type can be represented with its own color and symbol. The user can choose which data is shown. Additionally, in order to better support exploratory analysis tasks, we also show aggregated statistics on the map.

3 Interactive Exploration and Analysis

We briefly illustrate the analysis on three examples in this section. Figure 1 shows four views and two different brushes used to explore the citizen-centric data. We start by selecting 31 participants, who

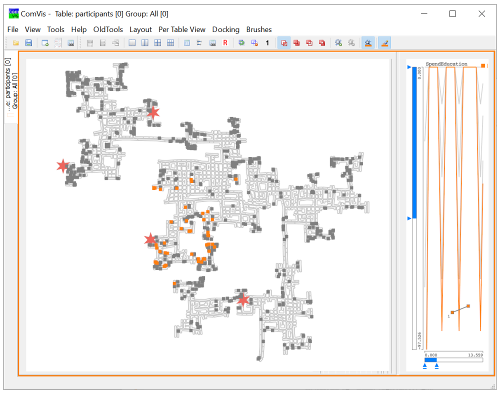

Figure 2: The map on the left shows four schools as stars, and apartments as gray and orange dots. The curve view on the right shows payments for education of all citizens. More than 1000 curves are superimposed. There are five clusters of curves, citizens that pay no education, and four different fees that are paid regularly. The most expensive fee is brushed. We see where corresponding citizens live.

spend a lot of money on food, by a simple line brush in the right most view (orange brush). Interestingly their account balance is not high and they do have a high joviality value. We then activate another brush (purple) and select 30 participants with the highest account balance. They do not spend a lot on food and have low joviality values. There is no clear spatial clustering of the two groups.

Figure 2 shows another snapshot from the citizen-centric analysis. On the right we can see the curve view which shows education expenses over time. We can see a regular pattern, the participants pay the same amount once a month. There are four possible amounts that are paid, and there is also a large group of participants who do not have education costs at all. It is plausible to assume that they do not have school-going aged children. If we brush the highest education fees in the curve view by simply drawing a line that crosses the corresponding curves, the map view highlights the apartment locations of the brushed citizens. We see the apartments in orange and all other apartments in gray (see the map in the Figure 2). Interestingly, all the apartments are relatively close to one of the schools. If we brush other amounts they are also clustered around other schools. In this way we can deduce the school fees of the individual schools from this correlation.

Finally, Figure 3 shows the map view with all pub locations depicted using the green stars. In addition there is a histogram for each pub. It shows distributions of the distance the pub guests have traveled in order to reach the pub (top image) or go back from the pub (the bottom image). We can see that most pubs are visited by customers who do not travel far. The circled pubs are exceptions. The pubs with IDs 1344 and 1342 have guests who travel a larger distance in order to get to the pub. There are even more of the customers who travel larger distances when leaving the pub. From the remaining pubs, only pubs 442 and 434 have the customers who travel a longer distance on their way home. This can indicate a high popularity of these four pubs. The pub-centric data set is used in this example. Each row in the table represents a pub.

4 Conclusion

A holistic analysis of complex data as given in this challenge can be successfully conducted by means of visual analytics. Due to the complexity and heterogeneity of the data and tasks, an extensive data processing that results in several data sets is needed. Custom views are also often needed when analyzing such complex data. Coordinated multiple views with support for complex brushing make it possible to efficiently explore various hypothesis and comprehend the data.

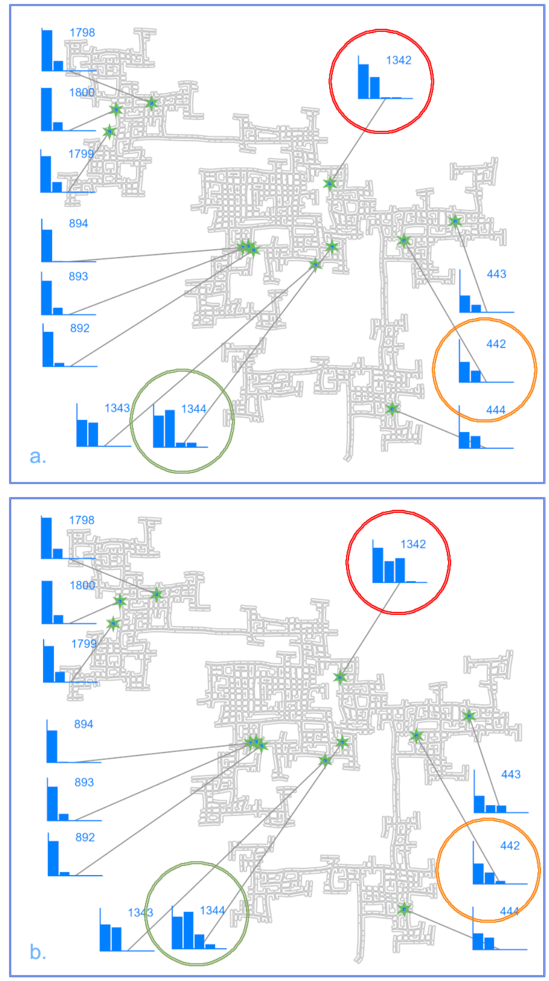

Figure 3: The map view shows all buildings and pubs as green stars. Additional categories (schools, restaurants. . . can be switched on). For each pub we visualize the distribution of distances the guests traveled towards the pub (**a.**) and when leaving the pub (**b.**). The circled pubs show larger distances when leaving the pub. These are also the pubs where the guests are willing to travel longer towards the pub.

We plan to further improve the newly proposed map view and combine the interactive approach with automatic analysis methods in the future. It seems that pure automatic or pure interactive methods are not sufficient for such unstructured data combined with exploratory analysis tasks.

Acknowledgments

VRVis is funded by BMK, BMDW, Styria, SFG, Tyrol and Vienna Business Agency in the scope of COMET – Competence Centers for Excellent Technologies (879730) which is managed by FFG. Parts of this work have been supported by Virginia Tech Institute for Creativity, Arts, and Technology.

References

[1] K. Matković, W. Freiler, D. Gracanin, and H. Hauser. ComVis: a coordinated multiple views system for prototyping new visualization technology. In *Proceedings of the International Conference on Information Visualisation (IV 2008)*, pp. 215–220. IEEE, Washington, DC, USA, 2008.