Deriving Visualisations from Data Cubes to Support Structured Visual Exploration of Spatiotemporal Data

Susanne Bleisch *, Michael Zwick*

* FHNW University of Applied Sciences and Arts Northwestern Switzerland, Institute of Geomatics Engineering, Muttenz, Switzerland

Abstract. Visual exploration of spatiotemporal data is still a very manual task often requiring a high level of domain knowledge and at least working knowledge of visualisation tools and methods. A set of skills and knowledge that may not be combined in one person (Van Wijk, 2006). Knowing the specific data set characteristics together with pre-defined or developing tasks and questions allows creating suitable visualisations to explore the data and find insights and answers. However, there is a risk that exploratory data analysis remains ad hoc and may miss insights that are not covered by the initial questions posed, the visualisations employed, or the interactions applied. Such insight may be revealed by systematic exploration that ensures coverage of all aspects of the data (Perer & Shneiderman, 2008). It is thus helpful to guide the data analyst through the process. Additionally, there are several cognitive biases, for example, cognitive inertia, which may impede our ability to explore all facets of a data set independently for insights.

To overcome these limitations, we are exploring the use of data cubes as a spatiotemporal data structure for visualisation generation. Mapping the data cube axes to different visual variables generates overview visualisations of the data set. Additionally, standard data cube functionality is used to systematically derive subsets of the data and to generate a range of visualisations for exploration. This allows guiding a data analyst through the exploration of all facets of a spatiotemporal data set. Furthermore, being creative with mapping data dimensions to visual variables allows for random or surprising visualizations that may keep the data analyst from getting tired and allow for different or unexpected views on the data set (Bleisch, Duckham, & Lyon, 2013). Over time, this system will also enable the collection of information on data characteristics as well as insights gained through the visualisations. Combining knowledge about data set

characteristics with measures of representation usefulness, such as number of insights gained, allows recommending the most promising representations for visual data exploration of specific data sets to less experienced data analysts.

Data cubes are used as data structure supporting multidimensional data with spatial and temporal references. The implementation uses the Python Data Analysis Library pandas (McKinney, 2012) which offers functionality for data cube preparation, handling as well as a number of predefined functions, such as slice, dice, restriction and aggregation, allowing for simple selection and sub setting of the data set. The resulting data views are then mapped to different visual variables in predefined or random orders to create a large variety of visualisations. For the visualisations the Python Interactive Visualization Library Bokeh (ContinuumAnalytics, 2013) is employed.

Creating a range of visualisations of the same data is a key strength of the implemented system. As the goal is spatiotemporal data exploration, without previously knowing what exactly we are looking for, it is difficult to optimise the visualizations for a specific purpose. Thus the interpretation of visual artefacts has to be hindered by exploring multiple views of the same data set.

The process from data preparation to the generation of visualisations has been implemented and applied to two different spatiotemporal data sets. One data set combines fish movement data from fish in an Australian river over several years with environmental context data, such as water temperature, river flow or moon phases (Bleisch, Duckham, Galton, Laube, & Lyon, 2014). The other data set details traffic accidents in different areas over several months and in relation to context information such as the weather situation. While the two data sets show commonalities in their data structures they are from two very different application domains posing different questions and hypotheses for exploration. The prototype system proved able to generate a variety of, often data dense, visualisations for both data sets (see for example Figure 1).

Keywords: visualization, exploratory, spatiotemporal, data cubes, structured guidance

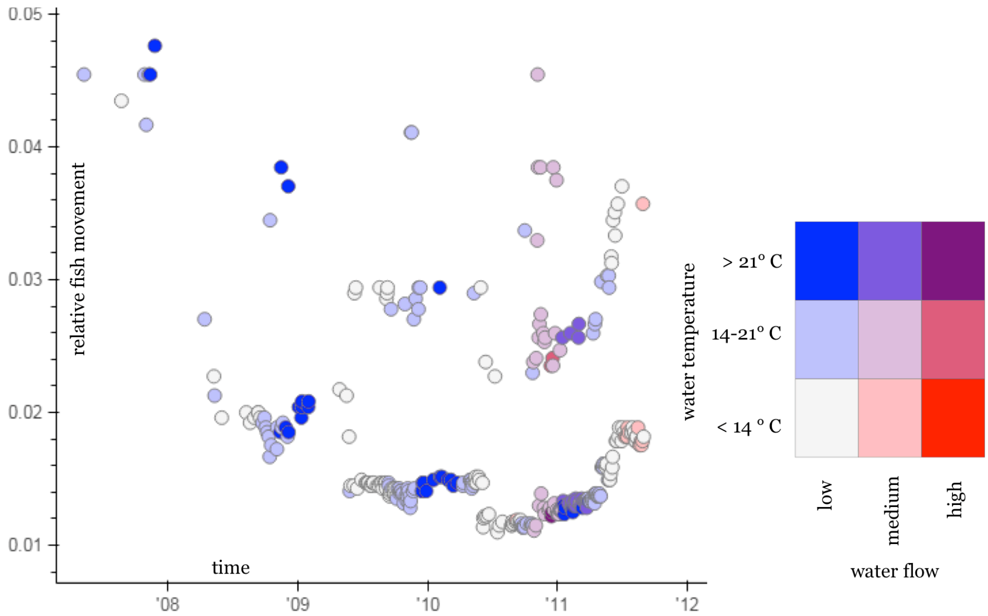

Figure 1. Example visualization of relative fish movement in a section of the river in relation to water temperature and water flow over time.

References

Bleisch, S., Duckham, M., Galton, A., Laube, P., & Lyon, J. (2014). Mining candidate causal relationships in movement patterns. International Journal of Geographical Information Science, 28(2), 363–382.

Bleisch, S., Duckham, M., & Lyon, J. (2013). Evaluating the Usefulness of Overview Visualizations for Users with Varying Levels of Domain Knowledge. In 26th Int. Cartographic Conference ICC 2013, 25-30 August 2013. Dresden, Germany.

ContinuumAnalytics. (2013). Welcome to Bokeh. Retrieved April 5, 2015, from bokeh.pydata.org/

McKinney, W. (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.

Perer, A., & Shneiderman, B. (2008). Systematic Yet Flexible Discovery: Guiding Domain Experts through Exploratory Data Analysis. In IUI'08 (pp. 109–118). Maspalomas, Gran Canaria, Spain: ACM.

Van Wijk, J. J. Van. (2006). Bridging the Gaps. IEEE Computer Graphics and Applications, 26(6), 6–9.