Evaluating the Usefulness of Overview Visualizations for Users with Varying Levels of Domain Knowledge

Susanne Bleisch*, Matt Duckham*, Jarod Lyon**

* Department of Infrastructure Engineering, University of Melbourne, Australia, susanne.bleisch@unimelb.edu.au, matt@duckham.org
** Arthur Rylah Institute, Melbourne, Australia, jarod.lyon@dse.vic.gov.au

Abstract. Exploratory data analysis may be supported by suitable visualizations. This study aims to evaluate how useful overview visualizations are to generate hypotheses for further analysis. Focus groups with domain experts and lay users participated in the evaluation of four different overview visualizations. The results show that all visualizations have their value but are of varying difficulty to interpret. Lay users were able to learn about the data set and data collection process. Both domain experts and lay users could use the visualizations to develop ideas and hypotheses about the represented data set. While experts were able to use their context knowledge to report findings not picked up by lay users they also seemed somewhat restricted in openness and creativity as findings not related to domain knowledge were generally doubted.

Keywords: overview visualization, exploratory data analysis, usefulness, focus groups, evaluation, context knowledge

1. Introduction

Suitable visualizations are assumed to support exploratory data analysis (EDA, Tukey 1977) where the users want to familiarize themselves with the data, search for interesting patterns or outliers, gain ideas for further investigation, or formulate hypotheses for detailed analysis. The information seeking mantra (Shneiderman 1996) suggests starting with an overview before offering tools for filtering the data or examining details. While the information seeking mantra is often used as a guiding principle for visualization design (Craft & Cairns 2005) there is scant literature regarding the

importance and usefulness of using visual overviews for generating ideas and hypotheses.

Spatio-temporal location based movement data is one type of available movement data (Andrienko et al. 2011). Here, we utilized location-based movement data from a radio-tagging network in the Murray River, Australia (Figure 1). In total more than 1000 fish were radio-tagged and 18 logging towers recorded when fish moved past them on a daily basis between 2006 and 2011. Originally, the data set was collected for evaluating the influence of management interventions in specific river sections on the native fish population. However, the rich data set may additionally be able to give further insight into fish behavior in the regulated Murray River.

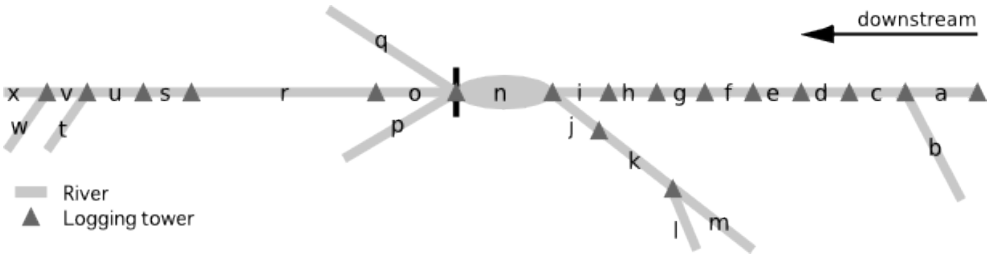

Figure 1. Schematic drawing of the monitored part of Murray River, Australia.

This study aims to evaluate different overview visualizations in regard to their usefulness for generating ideas or hypotheses for further detailed data analysis. Tukey (1977) stresses the importance of domain or context knowledge for exploratory data analysis. We evaluated different overview visualizations with users interested in data analysis but possessing different levels of domain knowledge to analyze the influence of context knowledge on the usefulness.

2. Methods

2.1. Overview visualizations

The data set was reviewed for its structure and four different overview visualizations were created. We define overview visualizations as representations showing all the data in a single display. Generally, highlighting specific aspects or subsets of the data (in this case variables such as species or seasons) is avoided. The only assumption made is that fish movement is more relevant for data analysis than stationary fish. Different techniques

for displaying all data points concurrently are used. However, overview visualizations can also be achieved by using different aggregations based on the structure of the data (for example, summarizing all fish movements based on predetermined criteria). Two of the created visualizations are adaptations of overview visualizations already in use or suggested by the data owners. Two additional overview visualizations were devised based on the data characteristics. Basing the visualizations on the data characteristics ensures that the same visualization types could also be used for other location based movement data sets.

2.2. Data collection

The four different overview visualizations were discussed in three focus group sessions with participants having varying levels of domain knowledge. Focus groups are small groups moderated to ensure that specific topics are discussed by the group (Grumbein & Lowe 2010). One focus group consisted of three domain experts (river ecologists), who were familiar with the test data set. They are assumed to have the best possible knowledge of the data set that is to be analyzed. Two other focus groups consisted of four PhD students in each group who are interested in data analysis but have only lay knowledge of fish behavior in rivers.

The focus group sessions were moderated to ensure similar treatment of the different groups. All groups were first introduced to the data set and its collection and then to the four different visualizations of the data set. The introduction took about 10-12 minutes. The rest of the 60-minute sessions were used to discuss the visualizations and the data they show. Participants were encouraged to report what they see in the visualizations, to ask questions and to generate hypotheses and ideas about how the would like to further explore the data set. Participants were asked for their preferences when the discussions were ebbing away.

With the participants' consent, video recording was used to capture statements and to allow for reconstruction of the knowledge built up through discussion and the use of the representations, i.e. including participants sketching on the paper printouts of the representations.

2.3. Data analysis

The video recordings were transcribed and the collected data was qualitatively analyzed for insights (North 2006), insights being the unexpected in the data or the generated hypotheses for further exploration. As insights were rarely reported directly, the data was subsequently qualitatively analyzed for its content according to an emerging coding scheme. The scheme comprises problems with the visualization, findings in a broader sense,

questions asked, improvements suggested, and participant's preferences. The coded and summarized content was then compared between the different visualizations and between the focus groups with different levels of context knowledge to gain an understanding of the usefulness of different visualization types and the influence of domain knowledge.

The data was not analyzed quantitatively. Focus groups have the advantage that the participants can build upon each other's statements and the discussions are thus likely to go deeper and be more varied. However, statements or findings are often reported only once, as participants tend not to repeat. Participants may nod or give another indication of their agreement with statements of others but this is difficult to interpret as it could mean that they agree but have not thought of that aspect themselves or it could mean that they had similar thoughts. Thus, quantitative analysis of statements is difficult.

3. Results

All four overview visualizations have their value in helping the participants of the focus groups to develop hypotheses and see specific aspects in the data. Also all focus groups pointed out problems with the visualizations and suggested improvements. Insights or findings were rarely reported directly, more often they were phrased as questions or even as problems, i.e. reporting not being able to see something could indicate the interpretation of the visible in a certain way.

The most basic visualization (showing all fish over time and indicating the different river zones (space, cf. Figure 1) where a fish stayed on a specific day as color value) helped the lay users most in gaining a better understanding of the data set and the data collection process. The domain experts used the basic visualization only briefly for confirming their knowledge of the data set by relating findings to specific aspects of the data collection process they knew about. All groups commented on the less than optimal use of color to convey spatial information. Overall, all overview visualizations greatly helped the lay users gain a deeper understanding of the data set, measured as the total number of questions asked. The visualizations prompted the domain experts to discuss some novel aspects of the data set and the data collection process, and to reconsider some of the original hypotheses.

Generally, the lay users reported similar findings as the domain experts did. Lay users often phrased findings as questions but were more open in accepting a finding as interesting and were creative in suggesting ways of how findings could be analyzed further to, for example, prove the existence of a

pattern seen in the visualizations. This may be related to their more general interest in data analysis and knowledge about data analysis techniques. The domain experts on the other hand often instantly related the findings to their context knowledge. Findings that did not match up previous knowledge were phrased as questions and often doubted. However, in a few instances the domain expert group pointed out interesting patterns, which they would like to explore further, that the lay user groups did not report.

Statements of preferences for the different visualization types are difficult to interpret. Generally, all focus groups agreed that the visualizations are more valuable if more time is spent with them and that most cannot be understood and analyzed at a first glance or rather are more useful after some discussion of what they show. All groups commented upon the underlying spatial representation of the river network (cf. Figure 1) in one of the visualization positively as it allows the instant localization of the different river zones. However, it was also mentioned that the easy look may be deceptive. All the other visualizations use a more abstract encoding of the spatial information through ordering or color.

4. Conclusion and Outlook

The overview visualizations evaluated in this study are useful for domain experts and lay users. The latter can use them for gaining a better understanding of the data set in general but also for identifying patterns and generating ideas for further analysis. In most cases, domain experts reverted to their context knowledge rather than perusing the visualizations and examining their findings in more depth. The context knowledge may even override some of the creativity or openness that is assumedly needed when trying to learn something new or additional from a well-known data set.

Evaluating the same visualizations with a focus group of domain experts who do not have knowledge about the specific data set used but rather the domain will extend this study. Additionally, another focus group with people interested in data analysis but with more research experience than PhD students will be included. Those results should allow expanding on some of the findings reported here.

Acknowledgement

We like to thank the focus group participants. This study would not have been possible without your input and time commitment. This research is funded by the Australian Research Council Discovery Project DP120100072

‘From environmental monitoring to management: Extracting knowledge about environmental events from sensor data’.

References

Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., & Wrobel, S. (2011). A conceptual framework and taxonomy of techniques for analyzing movement. Journal of Visual Languages & Computing, 22(3), 213–232.

Craft, B., & Cairns, P. (2005). Beyond Guidelines: What Can We Learn from the Visual Information Seeking Mantra? Ninth International Conference on Information Visualisation (IV’05) (pp. 110–118). IEEE.

Grumbein, M. J., & Lowe, P. A. (2010). Focus Group. In N. J. Salkind (Ed.), Encyclopedia of Research Design (pp. 501–502). SAGE Publications, Inc.

North, C. (2006). Toward Measuring Visualization Insight. IEEE Computer Graphics and Applications, 26(3), 6–9.

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proc. Visual Languages 96.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley Publishing Company.