Forest or slope? Comparing 2d and 3d visualisations through the wording of task answers

Susanne Bleisch[1], Jason Dykes[2], Stephan Nebiker[1]

[1] FHNW University of Applied Sciences Northwestern Switzerland, Institute of Geomatics Engineering, Gründenstrasse 40, CH-4132 Muttenz, Switzerland
Email: (susanne.bleisch, stephan.nebiker)@fhnw.ch

[2] School of Informatics, giCentre, City University London, London, EC1V 0HB, UK
Email: j.dykes@soi.city.ac.uk

1. Background and motivation

It's now easy to create applications displaying abstract numeric data in earth browsers (Sandvik 2010) such as Google Earth (Google 2010). However, guidance on how, when and even where to use abstract symbolism are limited, as is empirical evidence upon which such advice can be based. The size of any data object displayed in a desktop based 3d virtual environment varies, in addition to the varying data values, also according to perspective, one of the monocular depth cues enabling depth perception on a 2d display (Ware 2004). Additionally, we know that judging the size of non-aligned symbols is less effective in 2d (Cleveland and McGill 1984). Initial experiments (Bleisch et al. 2008) have demonstrated that displaying 2d bars on billboards in desktop based 3d virtual environments is suitable for simple tasks involving comparison of two bars, irrespective of depth cues. Additional observations have shown that, for example, displaying numeric values as circles of varying size in a virtual environment makes it almost impossible to compare two values. Changing the viewpoint in the virtual environment changes the size of the circles without any reference of size or depth cue. Using bars the constant width may serve as reference enabling the interpretation of the varying height in a perspective environment. Nevertheless the analysis of abstract data that have a relation to the landscape in which they were collected may benefit from a combined visualisation of data and surrounding landscape in 3d.

2. Experiments and analysis

Here we present the results of two further experiments that enable us to determine whether the display method is also suitable for more complex tasks with denser data sets (deer tracking data from the Swiss National Park) which are more related to real world applications. In doing so we are trying to 'bridge' between in-vitro (psychophysical) experiments and in-vivo case studies (Bleisch et al. 2009). Hypothesising that data displays in virtual environments may especially benefit analysis tasks relating data to altitude and landform we expect to find a more frequent use of altitude/landform related terms in the participant's responses when they explore data sets within 3d virtual environments as compared to 2d displays.

Experiment I consists of eight data sets showing deer activity at different times of day. Each displayed as several single bars of varying height in two different study areas (A and B) in 2d and 3d settings (Figure 1). In a balanced within-subject design the participants were asked to answer seven questions (see Figure 3, right) referring either to altitude, location or both in two 2d and two 3d settings. The questions were based upon an established set of task definitions (Andrienko and Andrienko 2006).

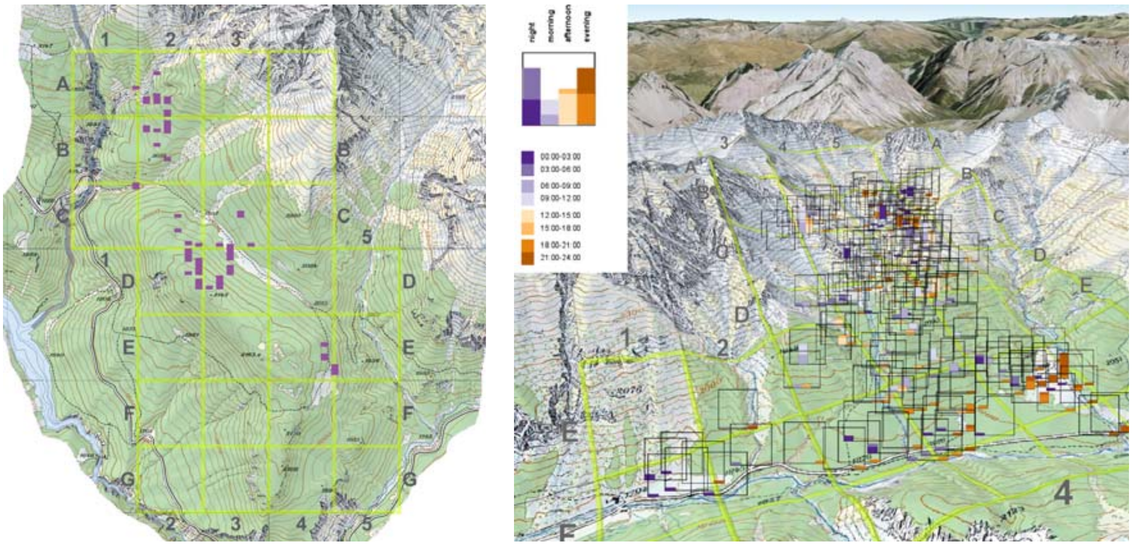


Figure 1. Screenshots: (left) area A, 2d, with a data set from exp. I (navigable SVG); (right) area B, 3d, with a data set from experiment II (navigable in Google Earth).

In experiment II bar charts showing the aggregation of the eight data sets of experiment I (thus showing deer activity of a whole day) were used in the same two areas (A and B) and were displayed in 2d and 3d (Figure 1). The participants, students and staff with GI background, were asked to "Analyse the deer data and describe the deer's habits regarding location, altitude and time of the day." We use insight reporting (Rester et al. 2007) whereas the participants report their findings in two half hour periods for one 2d and one 3d setting assigned in a balanced within-subject design. Similarities between the 2d and 3d settings are maintained with the isolated difference being the 3d representation of altitude and landform (digital elevation model draped with map/ortho imagery) and possible oblique views of it in the 3d settings compared to the single top view of the map/ortho imagery in the 2d representation (Figure 1).

Question answers (experiment I) and the reported insights (experiment II) are analysed in terms of word usage to evaluate whether altitude/landform or location words are more frequent in the 2d or the 3d settings and to identify any other aspects that influence the wording of the collected data. The different categories of words with examples are listed in Table 1. The two main categories altitude/landform (A) and location (L) were predefined by the study goals while the subcategories emerged as being datum (1), object (2) and relation/description (3) in both main categories. Additionally, the participant's comments are analysed using content analysis to better understand some of the findings of the word count analysis. We plan to analyse the data qualitatively regarding the complexity and plausibility of the given answers and reported insights at a later stage.

Table 1. Word categories and respective example words.

Category	**example words**	**category short name** (used in the figures)
altitude	1950m, 2400m	A1
landform	mountain, slope, ridge	A2
form words	steep, lower, highest	A3
grid	D3, F4	L1
land cover	forest, scree, grassy	L2
location words	north, left, south-east	L3

3. Results

In experiment I, 34 participants spent a total of 45h 11' answering the questions using a total of 15’455 words. In experiment II, 36 participants spent about 36 hours reporting insights using 5’929 words.

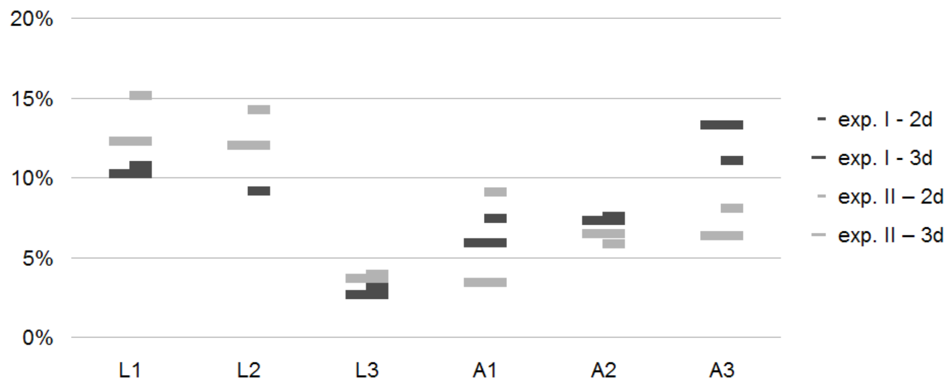
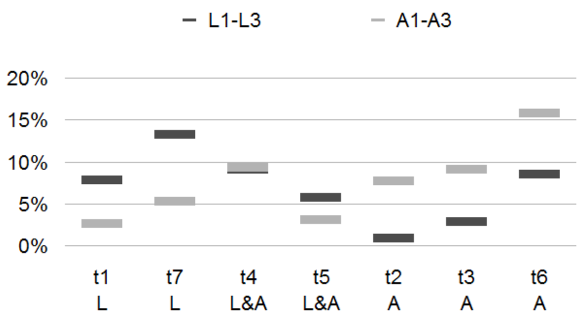


Figure 2. Percentage word counts per experiment (I and II) and visualisation dimension (2d and 3d) for different word categories (see Table 1 for explanation of L1 – A3).

Chi^2 tests comparing responses to the two settings in both experiment I ($X^2 = 25.8402$, df = 5, p-value = 9.583e-05) and experiment II ($X^2 = 14.7808$, df = 5, p-value = 0.01134) reveal that the word counts do vary with dimension of the display at a significance level of α=0.05 in both cases (see Figure 2).

Testing each category shows that category A1 varies significantly for both experiments I and II. Comments state that altitude values (A1) as found on the north oriented map are difficult to read in the potentially rotated 3d view. This may account for the fact that altitude values are more often used in task answers for 2d visualisations. Categories A3 and L2 vary significantly for experiment I where words from these categories are more often used in the 3d visualisations. Interestingly the trends shown in the same categories for experiment II are reversed. This may partly be explained by Figure 3 and the explanations below. In experiment II only one general task asks for reporting insights regarding altitude and location and thus seems to allow more freedom in word choice.



t1: simple task, asks for information regarding location (L)
t2: simple task, asks for information regarding altitude (A)
t3: asks for A
t4: asks for L&A concurrently
t5: asks for L first then A
t6: comprehensive task asks for A
t7: comprehensive task asks for L

Figure 3. Percentage of word counts in the aggregated word categories (L1-L3 and A1-A3) per task (task description on the right) for experiment I.

Figure 3 shows how the wording of the answers in experiment I is highly dependent on the wording of the tasks. For example, tasks t1 and t7 ask for information regarding location and the answers contain mainly location words (categories L1, L2 and L3).

Additionally, the wording of the answers is also dependent on the data sets and the area (A and B) they are displayed in (Figure 4). No comparable trends can be found for data set and display area dependency between the experiments I and II.

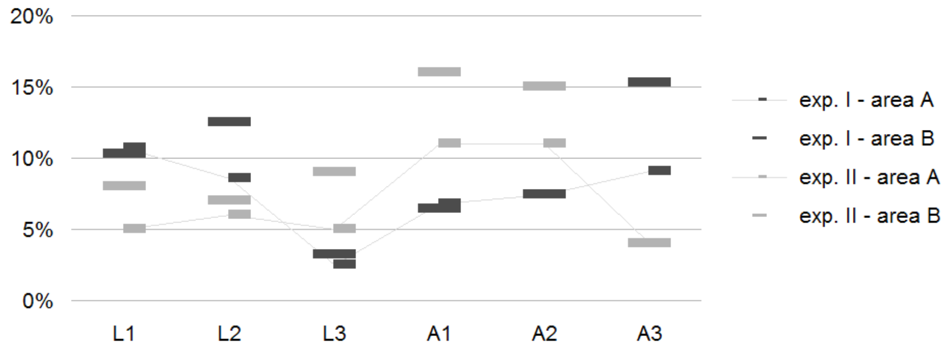


Figure 4. Percentage of word counts per word category in the areas A and B in
- left: experiment I (four data sets in A and B, seven questions),
- right: experiment II (one data set in A and B, one general question).

4. Conclusions and outlook

The anticipated trend that words relating to altitude or landform (A1-A3) are used more often in the 3d visualisations and, vice versa, words relating to location (L1-L3) are used more often in the 2d visualisations is not detected in the data from the two experiments described here. But the comparison of the word counts shows that the wording of the question influences the usage of words in the answers more strongly than the dimension of the display (2d or 3d). Additionally, the word counts also vary strongly between the different areas and data sets and also other aspects such as exact altitude values being difficult to read from a potentially rotated map in the 3d views have a strong influence on the use of them in the answers.

These results suggest that when geovisualization tasks that involve relating georeferenced values to generate insights, the types of terms used in describing insights derived from different visualisation displays are more dependent on the task phrasing and study areas or data sets used for the evaluation than on the variation in display type. Employing complex relational tasks and diverse data sets is required for geovisualization studies when trying to ‘bridge’ between in-vitro experiments and in-vivo case studies (Bleisch et al. 2009). However, those undertaking such studies need to account for ways in which the selection of the data sets, tasks and phrasing of the tasks may influence the outcomes of the evaluation and may make them potentially less generalisable. These findings are in line with the suggestion that information in a visualisation can be significantly influenced by context, task demands or verbal instructions (Ziemkiewicz and Kosara 2008).

The importance of aligning tasks with objectives (e.g. Board 1978) is well known and there is also literature available about how to phrase questions/tasks from areas such as usability testing or asking questions for surveys or interviews. However, while these guidelines offer valuable but sometimes conflicting advice for asking questions they rarely consider the influence of the question wording on response. As Ziemkiewicz and Kosara (2008) demonstrate, using different metaphors in wording the questions can have an effect and demonstrating, understanding, predicting and ultimately accounting for such effects can help develop theory for visualization and modern cartography. Our plans to conduct further evaluation of the data collected during experiments I and II for plausibility and response complexity will contribute to this ongoing process.

Acknowledgements

We thank the Swiss National Park for making the deer tracking data available. 'Thank you' also to the many participants in our experiments. We appreciate their efforts in supporting our studies.

References

Andrienko, N. and G. Andrienko (2006). *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Berlin, Springer.

Bleisch, S., J. Dykes and S. Nebiker (2008). "Evaluating the Effectiveness of Representing Numeric Information Through Abstract Graphics in 3D Desktop Virtual Environments." *The Cartographic Journal* 45(3): 216-226.

Bleisch, S., J. Dykes and S. Nebiker (2009). A meta-framework of methodological approaches exemplified by 3D geovisualization research. Workshop Human Aspects of Visualization, INTERACT 2009, Uppsala, Schweden.

Board, C. (1978). "Map reading tasks appropriate in experimental studies in cartographic communication." *The Canadian Cartographer* 15(1): 1-12.

Cleveland, W. S. and R. McGill (1984). "Graphical Percpetion: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79(387): 531-554.

Google. (2010). "Google Earth." Retrieved 23 April, 2010, from http://earth.google.com/.

Rester, M., M. Pohl, S. Wiltner, K. Hinum, S. Miksch, C. Popow and S. Ohmann (2007). Evaluating an InfoVis Technique Using Insight Reports. 11th International Conference Information Visualization, IV2007, Zürich, Switzerland, IEEE Computer Society.

Sandvik, B. (2010). "Using Geobrowsers for Thematic Mapping." Retrieved 23 April, 2010, from http://thematicmapping.org/.

Ware, C. (2004). *Information Visualization - Perception for Design*. The Morgan Kaufmann Series in Interactive Technologies. San Francisco, Elsevier.

Ziemkiewicz, C. and R. Kosara (2008). "The Shaping of Information by Visual Metaphors." *Transactions on Visualization and Computer Graphics (Proceedings InfoVis)* 14(6): 1269-1276.