EuroRV[3]: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2013)
P. Rosenthal, R. S. Laramee, M. Kirby, and G. L. Kindlmann (Editors)

Gaze into Hierarchy: A Practice-oriented Eye Tracking Study

N. H. Müller B. Liebold D. Pietschmann P. Ohler P. Rosenthal

Chemnitz University of Technology, Germany

1. Introduction

The visualization of hierarchical data is a wide field and plenty of different approaches have been proposed for various applications and purposes. A comprehensive survey on hierarchy visualizations was recently presented by Schulz et al. [SHS11]. Although every approach has its own claimed advantages, for practitioners it is often unclear what these mean in the specific context and which method to use.

Consequently, several user studies concerning different scopes and methods for evaluating visualizations of hierarchical data exist. In a traditional user study, measuring performance in terms of correctness and time, Stasko et al. [SCGM00] evaluated the treemap and sunburst approach regarding performance for standard file-management tasks. The results show that the more explicit representation of hierarchy relations in sunburst outperforms the treemap representation. In a similar setting, Wang et al. [WTM06] evaluated the effectiveness of the standard file browser, rings, and treemaps. Recently, Burch et al. [BKH*11] investigated different node-link layouts in an eye-tracking study. Using one question type and also an exploratory task, the authors discovered that circular layouts perform worse compared to axis-aligned layouts due to the fact that participants are more confident when using the latter. However, to our knowledge there exists no study to systematically investigate dependencies between type of visualization and type of task that had to be performed.

We present a user study measuring the ability of different hierarchy visualization techniques to facilitate a rapid overview of the structure and intuitive impression of proportions between nodes. For this purpose we compare the three most popular top-down representations: nodelink, icicle plot, and squarified treemap. We compare these types of visualizations with respect to four tasks, from which three are supposed to favor one visualization respectively and one can be seen as equally difficult for all visualizations:

- Count all nodes of the hierarchy. (nodelink)
- Count leaf nodes of the hierarchy. (treemap)
- Compare the combined area of two pairs of nodes within one level of the hierarchy. (icicle plot)
- Compare the combined area of two pairs of nodes across different levels of the hierarchy. (equal)

2. User Study

We conducted an experiment at a university computer laboratory measuring participants' performance (accuracy), reaction times and eye movements. We focus on performance measures and their suitability to evaluate task-specific differences between visualizations. We employed a $3\times4\times2$ within-subjects design with visualization type, task, and hierarchy complexity (low, high) as independent variables. The task levels differed in difficulty: Counting leaves or nodes is less cognitive exertive than comparing node areas. We recruited $N = 69$ (f = 53; m = 16; age: $M = 21.09$; SD = 2.40) university students of the local communication studies program. Each participant had to complete all of the 24 item combinations. Items were pre-randomized in two sequences, such that two items of the same factor level could not be in succession. The sequences did not differ in their performance, $t(67) = 0.238$, n.s.. For each hierarchy, the participants had to choose the correct answer out of three possibilities.

The eye tracker (SMI RED), used to track the participants' visual problem solving strategies, was a contact-free binocular infrared camera system measuring a corneal reflex. The camera was mounted below a 19" monitor with a 1280×1024 resolution, presenting the stimulus. All participants were calibrated using a five point matrix.

3. Results

Alpha levels for all calculations were set to $p < .05$. To test our hypotheses we used a $3\times4\times2$ repeated measures ANOVA with participants performance as dependent variable. We found significant main effects for the type of visualization, $F(2,136) = 53.77$, $p < .001$, $\eta^2_{\text{part}} = .442$, and the type of task, $F(3,204) = 28.471$, $p < .001$, $\eta^2_{\text{part}} = .295$, indicating that both type of visualization and task significantly influenced the participants performance. Participants' performance was significantly lower when using treemap compared to nodelink and icicle plot, whereas the latter two did not differ significantly. Participants performed well above chance level with both nodelink ($M = 0.55$, $t(68) = 8.73$, $p < .001$) and icicle plot visualizations ($M = 0.54$, $t(68) = 9.03$, $p < .001$). However, participants did not perform above chance when presented the treemap ($M = 0.33$, $t(68) = 0.332$, n.s.). Contrasts for the main effect of task revealed that participants performed better for counting tasks than for comparison tasks, $F(1,68) = 70.78$, $p < .001$, $\eta^2_{\text{part}} = .510$. They further performed better, when counting nodes compared to leaves, $F(1,68) = 14.05$, $p < .001$, $\eta^2_{\text{part}} = .171$, but did not perform differently in comparison tasks within or across hierarchy levels, $F(1,68) = 0.05$, n.s.. Performance at counting tasks was significantly above chance level ($M = 0.57$, $t(68) \geq 5.69$) but did not exceed chance level at comparison tasks ($M = 0.37$, $t(68) \leq 1.79$). Results did not show a significant main effect for complexity of hierarchies, $F(1,68) = 2.607$, n.s..

In terms of 2-way interactions, participants' performance differences between tasks did depend neither on the presented visualization (visualization×task), $F(5.16, 350.65) = 1.81$, n.s., nor on the complexity of the visualization (task×complexity), $F(3,204) = 1.17$, n.s.. Nonetheless, we performed simple effects analyses for the visualization×task interaction. While the icicle plot yielded significantly better comparison performance in contrast to treemaps, $F(1,68) = 12.41$, $p < .01$, $\eta^2_{\text{part}} = .154$, nodelinks did not yield a different performance than icicle plots, $F(1,68) = 3.26$, n.s.. Differences in performance between both complexities marginally reached significance (visualization×complexity), $F(2,136) = 3.08$, $p < .05$, $\eta^2_{\text{part}} = .043$. Moreover, nodelink and icicle plot differed significantly in their performance according to the respective complexity, indicating that participants performed better when presented an icicle plot in the low-complexity condition and a nodelink in the high-complexity hierarchy.

The RM-ANOVA also revealed a significant 3-way interaction (visualization×task×complexity), $F(6,408) = 13.37$, $p < .001$, $\eta^2_{\text{part}} = .164$, indicating that individual visualizations in combination with different tasks yielded a difference in performance when comparing low and high complexity conditions. Participants' performance with nodelinks compared to icicle plots differed at all levels of independent variables combinations, $F(1,68) > 9.96$, $p < .01$, $\eta^2_{\text{part}} \in [.128, .222]$. Interestingly, participants were able to perform above chance level in only one of the tasks when using a treemap, i.e. counting nodes at low complexity ($M = 0.48$, $t(68) = 2.40$, $p < .05$).

For the analysis of eye-movements, the fixation detection was set to 80ms. Visual inspection of the eye-tracking data suggests a focused strategy of participants during the identification and guessing efforts when performing the area comparison task. This behavior was found for both, low as well as high complexity of the hierarchy. A heatmap for the comparison of the sum of areas across different levels of the hierarchy when using a simple nodelink can be seen in Figure 1. Similar results are visible during the counting tasks for both, nodes and leaves. The first of the three answer-choices at the left-hand side of the visualization has been used most frequently as a reminding reference regarding the task. Our ongoing analysis will combine the accuracy and reaction time measures with the findings of the eye-tracking setup.

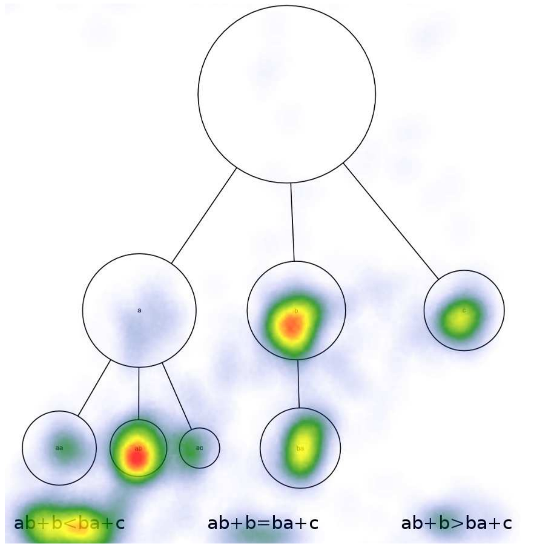

Figure 1: *Heatmap of the fixation durations during the task of comparing the sum of areas of two nodes respectively.*

4. Conclusions

We have presented first results of our user study of the most well-known and practically relevant visualizations for hierarchical data. Regarding our hypotheses we found mixed results. The treemap visualization underperformed, probably due to the relatively low experience within the group of participants. Also, the counting of nodes seems to be easier than counting leaves, probably because the former does not require any knowledge about the tree structure. Further insights of our performance analysis as well as an enhanced and substantiated discussion of our hypotheses will be presented in the extended version by a combined analysis of the eye-tracking data and the reaction times.

References

[BKH*11] Burch M., Konevtsova N., Heinrich J., Hoeferlin M., Weiskopf D.: Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Trans. Vis. Comput. Graphics 17*, 12 (2011), 2440–2448. 1

[SCGM00] Stasko J., Catrambone R., Guzdial M., McDonald K.: An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies 53*, 5 (2000), 663–694. 1

[SHS11] Schulz H.-J., Hadlak S., Schumann H.: The design space of implicit hierarchy visualization: A survey. *IEEE Trans. Vis. Comput. Graphics 17* (2011), 393–411. 1

[WTM06] Wang Y., Teoh S. T., Ma K.-L.: Evaluating the effectiveness of tree visualization systems for knowledge discovery. In *Proceedings of EuroVis* (2006), pp. 67–74. 1