Evaluating the Effect of Visualizing Dimensions Search Space on Exploratory Visual Data Analysis

ABEER ALSAIARI ANDREW JOHNSON

INTRODUCTION

In exploratory visual data analysis, analysts constantly investigate different subsets of data. The cost of deciding what to explore next “Gulf of Goal Formation” [1] is a major component of interaction costs in information visualization. When a team of analysts collaborates using multiple devices to work on an analysis task, the decision cost can be higher due to short-term memory and the recency effect. Analysts in collaborative settings need to understand what was investigated by the team and what was left. Visualizing the dimensions search space can communicate to the team what dimensions have been investigated (and in what combination) and what were left. We conducted a between-groups study to evaluate the effect of visualizing the dimensions search space.

[1] Lam, Heidi. "A framework of interaction costs in information visualization "IEEE transactions on visualization and computer graphics 14.6 (2008): 1149-1156

VISUAL DESIGN

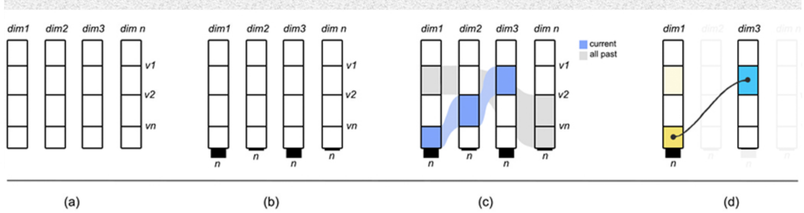
We selected the parallel set as the visual representation for visualizing the dimensions’ search space:

(a) Each dimension is represented by a line- set divided into several blocks representing its values distribution (categorical or numerical)

(b) Each dimension is attached with a bar representing the frequency of investigation of this dimension

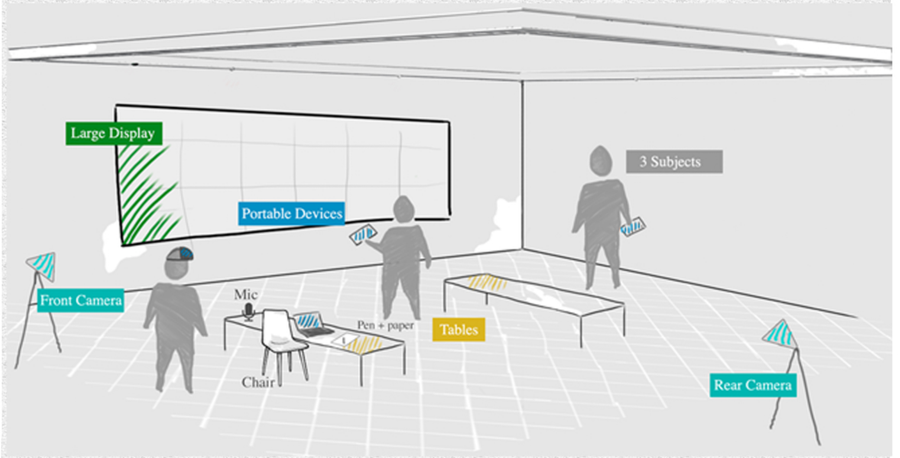
(c) Current dimensions that the analysis is pivoted to will be stacked to the left showing the current coverage of the data space (in blue) and all past data spaces will be combined into one category (in grey) and send to the back

(d) Clicking on a dimension shows co-investigation of data spaces (appearing together in a chart)



STUDY DESIGN

We conducted a between-groups study to evaluate the effect of visualizing dimensions’ data space coverage and co-investigation. The study contained two conditions: baseline and full versions where half of the groups used a baseline visual analysis tool and the other half of the groups used a full version of the tool enhanced with a visualization of the dimensions search space.



Participants: 30 participants as 10 groups of 3

Apparatus: in baseline version, participants used PolyVis, a visual data analysis tool designed for cross-device collaboration. In full version, participants used PolyVis integrated with a visualization of the search space in a separate window on the large display.

Setup: the study was conducted in a room approximately 10.61 by 5.59 meters, equipped with a high-resolution large display. Other devices were placed on a table in the middle for use during the study: one MacBook Pro one 8” Samsung - Galaxy Tab A, one 10” Samsung - Galaxy Tab A, and one Microsoft HoloLens 1. Systems usage logs were collected. The study was video and audio recorded.

Task: In the first task, participants were given focus questions that can be answered by creating one or two visualizations. We opted for the focus questions to be a practical tutorial on how to use the system. Participants were then asked to explore the earthquake events and wells injection activities and identify trends/observations in the data using two geoscience datasets.

FINDINGS

The visualization of the dimensions search space played a central role in data selection and attributes co-investigation. Visualizing each dimension’s data space coverage facilitated the selection of the next course of analysis. In addition, forming attributes co-investigation mostly took place by referring to the list of dimensions visualized on the big display.

To measure the effect of visualizing the dimensions search space on reducing the decision cost, we calculated the rate of producing views for each condition. The higher the rate of views generation, the less the cost of goal formation.

First, we counted the number of created visualizations by each group. Full version groups created an average of 43.6 views (SD = 18.15), versus 20.4 (SD = 13.07) for baseline version groups. A tow-tail independent t-test showed that full version groups generated more views than baseline groups ($t = 2.3198$, $df = 8$, $p = 0.0489$ at $p < .05$).

To eliminate the effect of sessions’ time and calculate the views’ generating rate, we divided the number of created views by the session’s time. Full version groups created an average of 0.7880 views per minute (SD = 0.1987), versus 0.4360 views per minute (SD = 0.1781) for baseline version groups. Tow-tail independent t-test showed that full version groups generated views at a higher rate than baseline groups ($t = 2.9498$, $df = 8$, $p = 0.0184$ at $p < .05$) which indicates a reduction in decision cost.

FUTURE WORK

Supporting the breadth of exploratory visual analysis has been a central goal in the visualization community.In future work, we aim to study how visualization of dimensions search space can increase the breadth of the analysis.

In addition, when working on an analysis task, participants ask questions and come up with observations about the data. Furthur analysis will reveal if visualization of dimensions search space can help analysts ask more questions and find more observations.

Contact info: aalsai3@uic.edu