# Cognitive Style and Zooming in Interactive Visual Data Analyses

Peter Bak and Joachim Meyer

*Ben Gurion University of the Negev / Israel*

**Abstract**

*The study deals with the choice of resolution levels and task performance in interactive graphs. An experiment aimed to assess the effects of task properties and users' cognitive style on performance of a visual data analysis task. Participants had to determine whether specific data-points are above or below a threshold value. To do so, they needed to adjust the resolution of the display to make correct decisions. Using a higher level of resolution limited the possible gains from a correct decision. Results showed that users with high analytic cognitive styles had more correct answers and used higher resolution levels than low analytic participants. Highly intuitive participants used lower resolution levels and made fewer correct decisions than lower intuitive participants. Cognitive style was shown to be related to the use of interactive tools in visual data analysis.*

## 1. Introduction

Visual data mining is a rapidly growing research field, in which humans' perceptual abilities are combined with the computer's algorithmic capabilities. Earlier applications presented the results of a data mining process to the user who usually had limited information about the way the algorithm performed the analysis and had little control over it. More recently systems incorporate active interaction between users and algorithms, providing users with the ability to control the process. This generally leads to more effective exploration [4, 6]. In addition, users tend to have more confidence in results that were obtained from active interaction with the system [4].

The interactive and visual exploration of large data sets is a difficult task, in which users have to encode the visual elements of the display, manipulate the display and the data so information will be available, combine information from different sources, and interpret the findings for decision making [8]. Performance in such environments can be a great challenge, and can be affected by factors like the visibility of properties of the displayed data, the users' cognitive information processing style, and properties of the task that needs to be performed.

Data sets for visual data analysis are often very large, and they can be viewed at different levels of resolution. Lower resolution displays limit the number and complexity of the graphics and tend to be more readable for users and decision makers. But low resolution also means that less detail is shown, and decisions have to be based on inaccurate information. The identification of trends and patterns is more difficult, and users can loose track of global dependencies and structures [7]. Several techniques have been developed to address this dilemma of context and detail in visualizations. For instance, detail-and-overview techniques show both a global overview and a detailed representation of a selected area. This technique is further developed in focus-and-context methods where the detailed information is spatially embedded in the context [5].

Visualization environments today generally provide users with tools to manipulate the display or the underlying data in order to perform a task. When data are available in different resolution levels, users need to find an appropriate (and preferably the optimal) resolution level for understanding the data and its underlying structures [8]. The relative efficiency of different levels of resolution depends on properties of the data, such as the variability and periodicity of the data, or the number and size of changes in the data that are of interest [1].

Performance in information extraction tasks is also affected by characteristics of the viewers [1]. One particular type of individual differences that received much attention in recent years is cognitive style [2]. This approach distinguishes between experiential and rational cognitive styles by which people process information. Rational cognitive style is described as analytic, reason oriented, evidence based and context dependent. Its processing is effortful and involves computations, comparisons, and integration of information. Experiential cognitive style is described as holistic, feeling oriented and changes only with repetitive and intense experience. It is often associated with biases in decision making and judgment [2, 3]. People can have any combination of scores in experiential and analytical cognitive styles, although the two tend to be inversely correlated.

Previous research found cognitive style to be a major determinant of performance in a visual data analysis task that required the detection of cause-and-effect relations in time series datasets [1]. Users with higher scores on the rationality measure performed slower but more accurately and used tools more strategically than users with lower scores on this measure. Users with higher scores on the experientiality measure performed tasks faster, but less accurately. They were also less systematic in their tool usage than users with lower scores on this measure [1].

## 2. Empirical Research

The empirical study investigates users' performance in an interactive visual data analysis task. The task was designed as a decision making task for which participants could view data at different resolution levels. Users should aim to determine the optimal resolution level depending on properties of the data and the task in order to optimize performance. The experiment allows us to assess the effects of cognitive style and task properties on the choice of the levels of resolution, performance accuracy and speed.

### 2.1. Data

The displayed data were generated by summing values of five sine functions with different amplitudes and frequencies, as shown in Table 1.

**Table 1.** *Parameters of the sine-functions, used to generate the displayed data.*

| Variables | Amplitude | Frequency |
|---|---|---|
| A | 23 | 0.114 |
| B | 11 | 0.043 |
| C | 19 | 0.093 |
| D | 5 | 0.091 |
| E | 3 | 0.051 |

The data could be viewed at seven different levels of resolution. For each level the average of a certain number of points was computed and was shown as a single data point. In addition, a pay-off matrix for correct and incorrect answers was defined. The possible reward for a correct response decreased, the higher the resolution level the person used. Penalties for incorrect responses were always the same. The specific resolution levels and payoffs are shown in Table 2.

**Table 2.** *Number of data points shown at each level, and pay-off matrix.*

<table>
<tr><th>Level</th><th>Points to Average</th><th>Points to display</th><th>Correct Answer</th><th>Incorrect Answer</th></tr>
<tr><td colspan="5">Lowest Accuracy Level</td></tr>
<tr><td>7</td><td>71</td><td>9</td><td>6</td><td rowspan="7">- 6</td></tr>
<tr><td>6</td><td>57</td><td>11</td><td>5</td></tr>
<tr><td>5</td><td>47</td><td>14</td><td>4</td></tr>
<tr><td>4</td><td>33</td><td>20</td><td>3</td></tr>
<tr><td>3</td><td>23</td><td>28</td><td>2</td></tr>
<tr><td>2</td><td>13</td><td>50</td><td>1</td></tr>
<tr><td>1</td><td>No average</td><td>650</td><td>0</td></tr>
<tr><td colspan="5">Highest Accuracy Level</td></tr>
</table>

### 2.2. Environment

The experimental system consisted of a graphical environment, which allowed user-interaction, experimental task definition and the recording of user actions. The display had four major areas. The first area was the canvas that displayed the graph in a resolution level. Only one resolution level could be seen at a time. Figure 1 shows views of one data set with 650 data points at different resolution levels. The first graph shows level 6, each bar representing the average of 57 data points. The bar containing the target data point 750, located in the range between 742 and 799, has the value 97. The second graph shows level 2, where each bar represents the average of 13 data points. The bar containing data point 750 now shows the average of points 741 to 754 and has a value of 115. Finally, the most detailed graph at level 1 shows a line graph on which data point 750 has the exact value of 122.

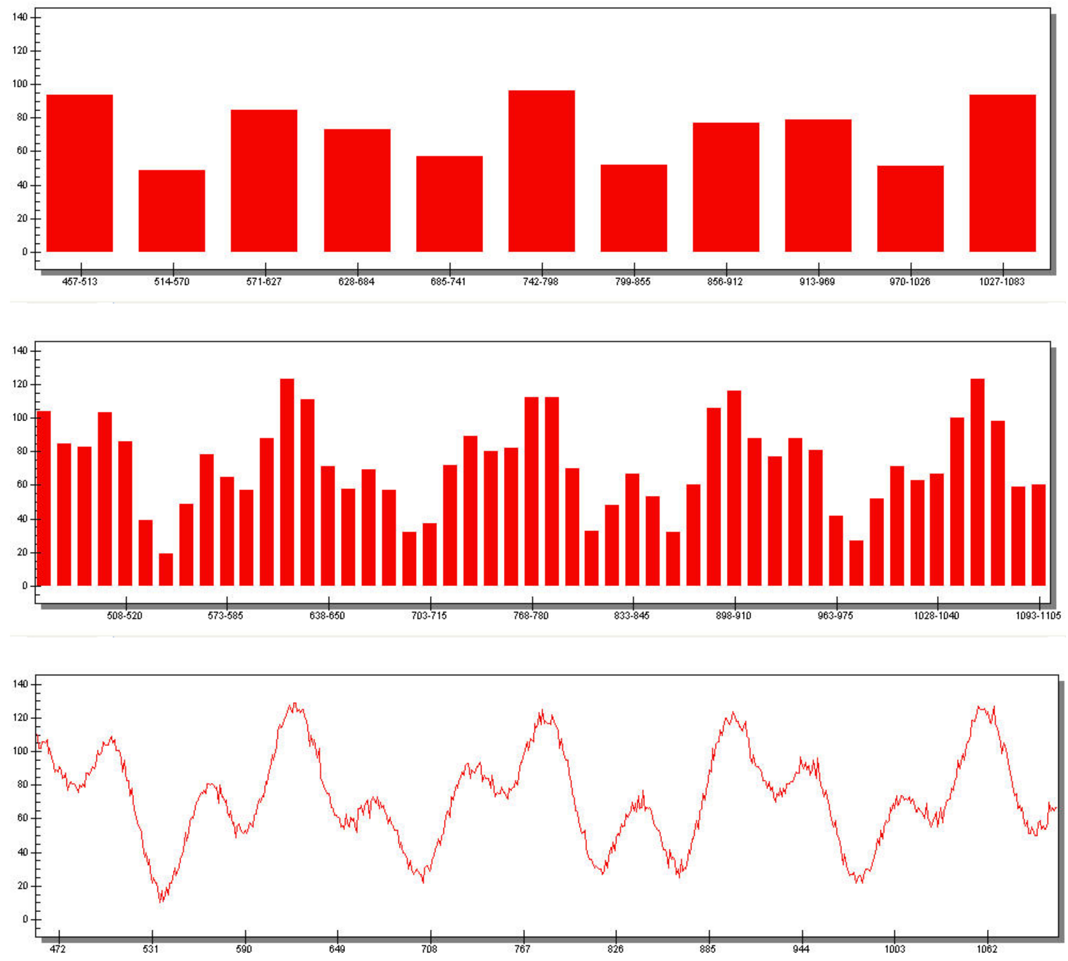


**Figure 1.** *Graphs with different resolution levels*.

The second area was the tool-area, which allowed the user to change the resolution level. Every task started at level seven, and users could decide to go to higher resolution level, if they felt not sufficiently certain to make a decision. With higher levels of resolution they saw more detailed information on the display and gained greater certainty about the behavior of the graphs. However, when moving to higher resolution levels, users had to take into account that the pay-off matrix got worse by every step, as presented in Table 2. Users could decide to go back to lower resolution levels for learning purposes, but they always received the score of the highest level reached during a task. The third area was the pen-tool, which provided the user with numeric information about selected data points. On each data bar or point the mouse-cursor changed its shape to a "pointing hand". By clicking at this position, the pen-tool area showed the range that was averaged and the value of the data bar or point. The fourth area was the task-area. In each task a data point and a threshold value were presented. The user had to indicate whether the actual value of the data point is greater or smaller than a given threshold value. The feedback-area displayed information whether the decision was correct or incorrect, the score value of the decision, and the total score. While the feedback-area was shown, the task-area was hidden to make sure that participants observe their feedback.

**2.3. Experimental Task**

The experiment involved a series of decision making tasks. Each task referred to a data point and to a threshold value. The user had to indicate whether the actual value of the data point was greater or smaller than the threshold value. The value of the data point 750 was 122, as shown in the Figure 2. An example of a task could be to decide whether the true value of the data point 750 is below or above the threshold value 115.

At the highest resolution level (level 1), where all data point were presented without any transformation, the task could be performed with perfect accuracy by reading off the data-point value from the graph. Therefore, at this level, the user was not rewarded for a correct answer, but was still punished for an incorrect answer. However, on all other resolution level data points were averaged and the actual data point requested in the task was hidden in a group of neighboring data points. The lower the resolution level, the less the displayed value resembled the value of the target point. At the lowest level, where all tasks started, a correct answer received the highest reward and an incorrect answer was punished as in all other resolution levels.

The user had to develop an analytic strategy in order to solve the tasks. One possibility was to compare the height of the bar, in which the data point was hidden, with the threshold value. The user had to take into account whether the bar was growing or shrinking when going to higher and more accurate resolution levels. The user could also see whether the data point was located towards the side of the bar or in the middle. If the data point was at the side of a bar, the user could expect that the data point's value will be close to the values of the neighboring bar. Users had to consider all these types of information in their decisions.

### 2.4. Participants

Seventy-two 3rd and 4th year students from the department of Industrial Engineering and Management participated in the experiment. All participants had gained experience with computers and with interactive graphs during their studies. Participation in the experiment was rewarded with one point as course credit, and with a score-dependent financial reward.

### 2.5. Experimental Design

The experiment had three parts. The first part was an on-line tutorial, providing some background information about the data and the task. The tutorial also explained how to change the resolution level and how to understand the pay-off matrix. In the tutorial users were also informed about payment and credit point conditions.

The second part of the experiment was the actual study. Users had to perform 38 tasks. The first 6 tasks were for learning the system. They were the same for all participants, and were not included in the analysis of the results. The order of the 32 experimental tasks was randomized for each participant separately. Each task asked the user to compare a data point to a threshold value. The graph display area always showed the lowest resolution level (level 7) at the beginning of each task. Users pressed the “greater” or “smaller” button according to their decision on the relative value of the point compared to the threshold. If the user was not certain enough about a decision, he/she could change the resolution level to the next level that was more accurate. If the user was not sure about the height of the bar he/she could also use the pen-tool. By clicking on a bar the user received numeric information about the actual range that was averaged for the bar and its value. The system always displayed the data point requested in the task in the middle of the screen. Therefore there was no need to scroll and look for the data point. After submitting a decision, the feed-back area was shown and the task area was hidden for three seconds, so that users received information about their decision-accuracy and score.

The last part of the experiment was a questionnaire. The questionnaire was a translation of the REI questionnaire suggested by Epstein and Pacini [2, 3]. Users had to answer all questions in order to finish the experiment successfully. The system created a time-stamped log file for every user-action that allowed the calculation of the variables required for the statistical analysis.

**2.6. Variables**

Dependent variables were accuracy of performance, the resolution level at which decisions were made, and the performance time for each task. Accuracy was measured as the number of correct answers. Resolution level was measured as the highest accuracy level the user reached in each task. Performance time was measured in seconds as the duration of each task.

Independent variables were the task properties and the users' cognitive style. The first task property was defined as Close/Far question type, referring to the distance of the threshold value from the true value of the data point. In case the threshold value was "close" to the true value, the distance was 8 pixels. When it was "far", the distance was 20 pixels between the threshold value and the true value of the data point. The second task property referred to the expected correct answer, which could be "greater" or "smaller".

Cognitive style measures of rationality and experientiality were collected with a scale developed by Epstein and Pacini [2, 3]. Rationality and Experientiality are two independent measures, and higher scores on either of them indicate higher ability on a measure. Rationality and experientiality have two sub-measures. We focused on the cognitive-ability scores, as suggested by Epstein and Pacini [2]. Each measure can have any value between 6 and 30. There are some suggestions in the literature how to use this measure in statistical analyses [2, 3]. One way is the median-split method that allows a categorization into high and low categories. This method has the advantage that a categorical measure can be easily used in a regular ANOVA. The second way maintains these measures as continuous variables. In this case these measures have to enter the statistical analyses as covariates.  This provides a more accurate estimate of the effects of the variables and their interactions with other variables.

## 3. Results

For each of the dependent variables a four-way repeated-measures ANOVA was conducted. In all analyses the independent variables were Close/Far, Greater/Smaller as categorical variables. The cognitive style variables of rationality and experientiality were continuous measures and entered the analysis as covariates. Therefore the analyses can be seen as a combined analysis of variance and co-variance. The analyses were carried out in SPSS which allows this combination of categorical and continuous measures in an ANOVA.

### 3.1 Number of correct answers

The analysis of performance accuracy (measured as the number of correct answers) showed a significant main effect of rationality, $F(1,41)=7.591$, $p<.05$, $MSe=.155$. Participants with higher rationality scores tended to have more correct responses in the task, as shown in Figure 2.

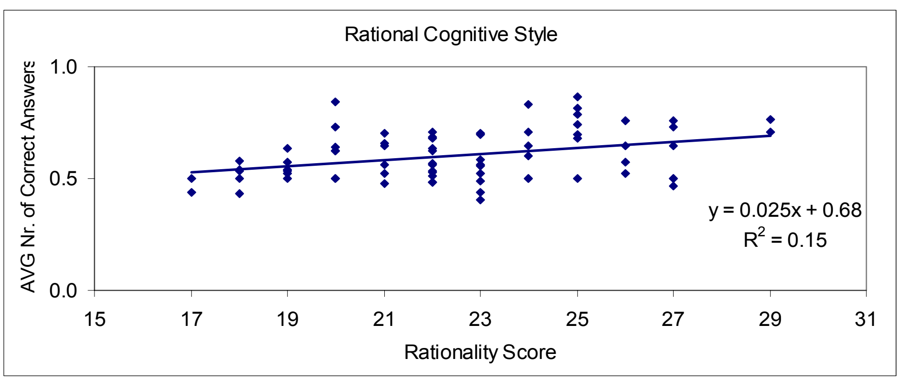


**Figure 2.** *Accuracy as a function of rationality.*

The interaction Close/Far and Experientiality was also significant, $F(1,41)=5.573$, $p<.05$, $MSe=.07$, as shown in Figure 3. The number of correct answers was inversely correlated with experientiality when the threshold value was “far”, requiring less accuracy. There was no such correlation when the threshold value was “close” and more accuracy was required.

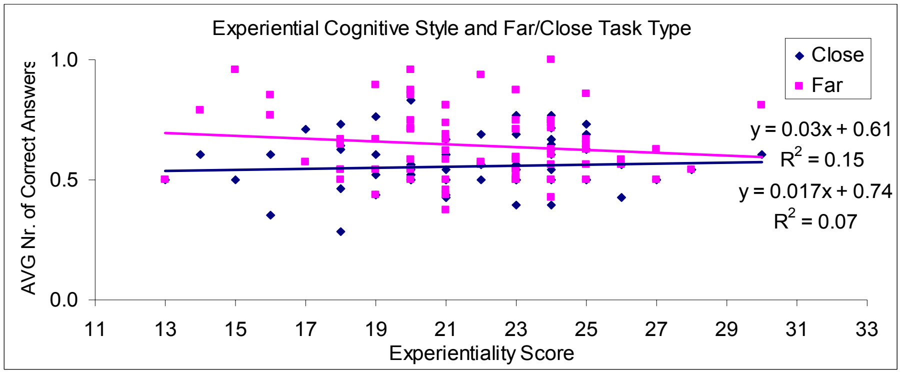


**Figure 3.** *Performance accuracy as a function of experientiality and Close/Far task property.*

### 3.2 Performance time

None of the effects in the analysis of performance times were significant. Thus cognitive style did not affect the time participants took to perform the tasks.

### 3.3 Resolution level

The analysis of the resolution level chosen for each task used the highest resolution level reached in a task as the dependent measure. There was a significant main effect of rationality, $F(1,41)=8.453$, $p<.05$, $MSe=10.9$. Participants with higher rationality scores chose more often higher resolution levels than participants with lower rationality scores, as shown in Figure 4.

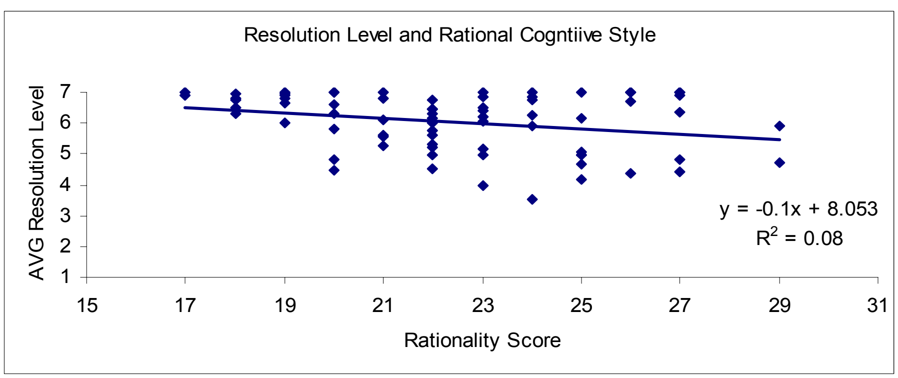


**Figure 4.** *Average resolution level as a function of rationality.*

There was also a significant main effect of experientiality, $F(1,41)=5.405$, $p<.05$, $MSe=10.9$. Participants with higher experientiality scores tended to use lower resolution levels than participants with lower experientiality scores, as shown in Figure 5.

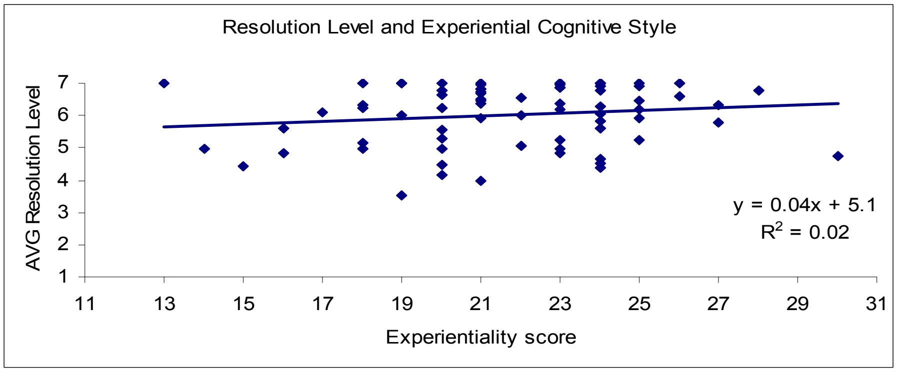


**Figure 5.** *Average resolution level as a function of experientiality.*

### 3.4 Pen-Tool usage

The analysis of the frequency of pen-tool usage showed a significant main effect of rationality, $F(1,28)=6.078$, $p<.05$, $MSe=32.4$. Higher rationality participants used the pen tool more often than low rationality participants, as shown in figure 6.

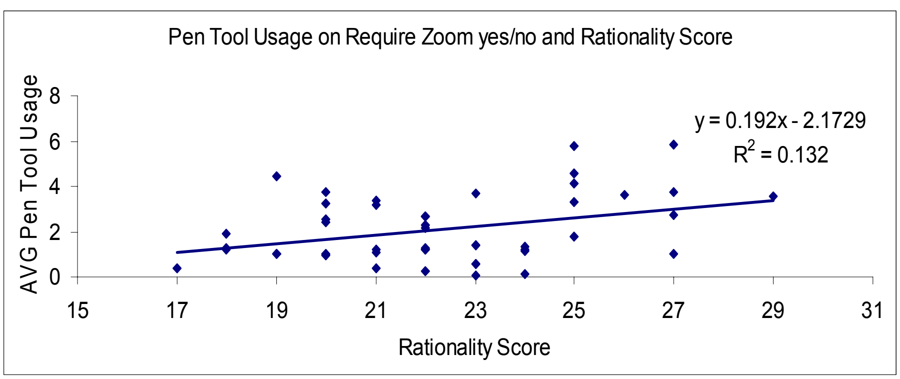


**Figure 6.** *Pen-Tool usage as a function of rationality.*

## 4. Summary and Discussion

The effects of cognitive style and properties of the tasks were analyzed in this empirical study. The task involved the simple decision whether the value of a data point exceeds a threshold value or not. To perform this task, users could change the resolution level at which the data were displayed.

Results point to a clear advantage of participants with higher scores on the rational-analytic cognitive style measure. These participants made more correct decisions than participants with lower scores in this measure. Performance differences seem to be due to the differences in finding an appropriate resolution level. Participants with higher rationality scores, or lower experientiality scores, choose higher resolution-accuracy levels than their counterparts.

There were a number of interactions of cognitive style with properties of the task. These results show that a higher score in rationality results in a performance that can be characterized as more visual and graph oriented. These participants used the zoom and pen-tool more often and showed good performance in all tasks conditions.

Experiential-intuitive cognitive style showed mostly an interaction with Close/Far task types. Participants with higher experientiality scores preferred to make their decisions on lower resolution-accuracy levels than participants with lower scores on this measure. Participants with high experiential scores also solved the tasks less accurately when the task required more accuracy than participants with low experientiality scores. Since the pay-off matrix punished participants for zooming to higher resolution levels that allowed more accurate performance, these results point towards a pay-off matrix orientation.

The two cognitive styles seem to derive from different systems, but they seem to have similar effects on task performance (although in opposite directions). This study is a step towards the understanding of the process by which users visually and interactively analyze graphically presented data sets. Further research is required to assess the effects of cognitive styles and task properties in visual data analysis in more complex environments and in real-world-data.

## 5. References

[1] Bak, P., and Meyer, J., "Effects of Cognitive Styles and Data Characteristics on Visual Data Mining", *Proceedings of SPIE International Society of Optical Engineering* 5669, 2005, pp. 77-86.

[2] Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H., "Individual Differences in Intuitive-Experiential and Analytic-Rational Thinking Styles", *Journal of personality and Social Psychology* 71 (2), 1996, pp. 390 – 405.

[3] Pacini, P., and Epstein, S., "The Relation of Rational and Experiential Information Processing Styles to Personality, Basic Beliefs, and the Ratio-Bias Phenomenon", *Journal of Personality and Social Psychology*, Vol. 76 (6), 1999, pp. 972-987.

[4] Keim, D.A., "Information Visualization and Visual Data Mining", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 7 (1), 2002, pp. 100-107.

[5] Oliveira, M.C.F., and Levkowitz, H., "From Visual Data Exploration to Visual Data Mining: A Survey", *IEEE Transactions on Visualization and Computer Graphics*, Vol 9 (3), 2003, pp. 387-394.

[6] Shneiderman, B., "Inventing Discovery Tools: Combining Information Visualization with Data Mining", *Information Visualization*, Vol. 1, 2002, pp. 5-17.

[7] Tory, M. and Moller, T., "Human Factors in Visualization Research", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 10 (1), 2004, pp. 1-13.

[8] Trafton, J.G., Kirschbaum, S.S., Tsui, T.L., Miyamoto, R.T., Ballas, J.A., and Raymond, P.D., "Turning Pictures into Numbers: Extracting and Generating Information from Complex Visualization", *Int. Journal of Human Computer Studies* **53 (**5), 2000, pp. 827-850.