A Comparative Study of Three Sorting Techniques in Performing Cognitive Tasks on a Tabular Representation

Inkyoung Hur, Sung-Hee Kim, Anya Samak, and Ji Soo Yi

July 17, 2012

Abstract

We investigated the impacts of three sorting techniques on various cognitive tasks performed on a tabular representation. The tasks under study were a multi-attribute object selection task and selected low-level analytic tasks. Three sorting techniques, including sorting by a column (Typical Sort: TS), sorting by all columns simultaneously (SimulSort: SS), and sorting by all columns with faithful vertical locations (ParallelTable: PT), were compared with a static table without the sorting feature (Baseline: B). An incentivized controlled laboratory study with 80 participants and a preliminary eye-tracker study were conducted to better understand the strengths and weaknesses of the four different approaches. We found that SimulSort and ParallelTable significantly improved the performance of multi-attribute object selection. ParallelTable, however, suffers from an occlusion problem, so it is not an appropriate support for some low-level analytic tasks. We used the findings to propose appropriate sorting techniques for specific tasks performed on a table.

Keywords: Visualized Decision Making, tabular visualization, SimulSort, ParallelTable, sorting

1 Introduction

[Figure 1 about here.]

A table is a powerful and pervasive means of representing multidimensional data. By presenting data in a grid, a table allows a user to scan and compare values in different cells horizontally and vertically as shown in Figure 1(a). In addition, interaction techniques, such as sorting by column, could make a table an even more effective means of data representation. As shown in Figure 1(b), sorting a table by one of the columns helps the user find a certain value more quickly without visually scanning every value in the column. Sorting also helps find the maximum and minimum values of a column. Thus, tables with or without the sorting feature have been widely adopted in various software applications, such as spreadsheets, word processors, statistical packages, and even email clients.

However, there is room for improvement in sorting features. While a table shown in Figure 1(b) could be useful for various cognitive tasks along one dimension, it is less effective for comparisons along multiple dimensions. For example, as shown in Figure 1(b), when the table is sorted by column C, it is difficult to quickly grasp where item 10 is ranked in terms of column B. In order to see the rank in column B, one needs to re-sort the table in column B, which will rearrange the entire table. Then, the insight gained from the previous sorting order (in column C) may be lost. We call this problem the "one-column sorting problem."

In order to resolve this problem, a visualization technique, called "SimulSort," was proposed in our previous study (Hur & Yi, 2009). As shown in Figure 1(c), SimulSort sorts all of the columns simultaneously, which allows the user to see ranks of an item in different columns without changing column orders. SimulSort virtually eliminates the one-column sorting problem at the expense of losing some benefits of typical tables (e.g., one cannot see which cells belong to a same item without highlighting.).

A potential drawback of SimulSort is that relatively similar values may be further apart, which may cause unnecessary bias. For example, in column B of Figure 1(c), a cell with a value of 73, highlighted in yellow, and a cell with a value of 72, highlighted in green, are two cell-heights apart even though the value difference between the two cells is just 1. This may cause unnecessary visual distortion because visual representation is unfaithful to the data that it represents. In order to resolve this issue, we created a new technique, called "ParallelTable." In Figure 1(d), the two cells are now partially overlapping to faithfully visualize their true values. However, the drawback of ParallelTable is that some cells are fully or partially occluded, even though the occlusion may be a faithful representation of underlying values.

The scenarios discussed above illustrate the potential strengths and weaknesses that exist in different sorting techniques (or different tabular visualizations). These strengths and weaknesses would matter more or less depending on what kinds of tasks the user needs to perform on a table. For example, while basic sorting may be sufficient for simple tasks, such as finding the maximum value in a column, SimulSort or ParallelTable may be more suitable for complex tasks, such as selecting an object while considering multiple attributes simultaneously. In addition, when SimulSort and ParallelTable are compared, it is not clear which of the following two problems is more serious: the visual distortion in SimulSort or the occlusion in ParallelTable. To date, no empirical evidence exists that provides a concrete answer to these questions.

The goal of this study is to empirically investigate how the four tabular visualizations described above (a static table, a table with a typical sorting feature, SimulSort, and ParallelTable) are perceived, learned, and employed by users for different tasks. To achieve this goal, we conducted two controlled experiments in which users participated in a multi-attribute object selection task and several lower level analytic activities, which will be described in more detail.

The contributions of this paper are as follows:

- We provide quantitative empirical results clearly showing which of three different sorting techniques are more appropriate for various cognitive tasks.
- We introduce a new visualization technique, called "ParallelTable," and test its effectiveness as an alternative to SimulSort.
- We provide other researchers and designers with guidelines for designing proper sorting techniques for tabular representations.

The paper is structured in the following manner. First, the related work is discussed, including various tabular visualization techniques and corresponding evaluation studies. Second, the four tabular visualizations used in the experiment are described. Third, the design of experiment, including the tasks used in the experiment, is described. Fourth, results of the experiment and their implications are discussed with results of a small eye-tracker study. Finally, conclusions and future studies are described.

2 Related Work

2.1 Tabular Visualizations

Tables are widely used visual representations that arrange numerical, textual, and even symbolic data in a grid, so that each datum can have either a column-wise attribute, a row-wise attribute, or both. Even in non-interactive form, tabular representations provide readers with the ability to retrieve information and see some patterns easily.

Various interaction and visualization techniques, such as sorting, filtering, highlighting, and zooming, amplify the effectiveness of these tables. One commonly used simple interaction technique in tables is sorting (mostly sorting by column). Using sorting, cells under a selected column are sorted in ascending or descending order. A spreadsheet is an interactive extension of a static table, and commercial spreadsheet implementations (e.g., Microsoft Office Excel®) have been widely successful in the market. Bertin suggests that sorted information is helpful for gaining better understanding of information and its retention (Bertin, 1983). Table Lens (Rao & Card, 1994) is another exemplar interactive tabular visualization technique that broadened the boundary of a table. However, the one-column sorting problem previously discussed could be a drawback of this sorting technique.

Some of the tabular visualization techniques inspired by parallel coordinates could be a solution for this one-column sorting problem. Parallel Bargrams (Wittenburg, Lanning, Heinrichs, & Stanton, 2001) and SimulSort (Hur & Yi, 2009) are examples of such variations. By sorting values in all columns, a user can see overall trends more easily without changing sorting orders. In particular, SimulSort was designed to share the same user interface components as a typical interactive table as shown in Figure 1(c). This commonality allows SimulSort to easily be switched to a table with the typical sorting feature if necessary. In addition, users may have less difficulty in understanding how to use SimulSort due to the familiar look and feel.

Other tabular visualization tools support decision-making by employing subjective expectation utility models, such as ValueCharts (Carenini & Loyd, 2004), which sorts items based on weighted-additive utility values. Finally, Table Lens is a visualization spreadsheet with dynamically resizable rows to support dynamic aggregation and details-on-demand, so that much more data can be presented in a single screen without relying on excessive panning or scrolling.

2.2 Evaluation of Tabular Visualizations

There is a tremendous amount of literature regarding the effects of graphs and tables on user performance in various cognitive tasks. Several frameworks have been suggested to explain why visual representation could enhance performance (Jun, Landry, & Salvendy, 2011; Ntuen, Park, & Gwang-Myung, 2010). However, studies comparing graphs and tables reported mixed results: Some reported that graphs are better than tables, but the others reported opposite results. In order to resolve these conflicts, Vessey (1991) proposed cognitive fit theory, which suggests that the mixed results are caused by the match and mismatch between presentations (charts or tables) and task types (spatial or symbolic tasks). In other words, when tasks are more spatial, graphs work better. When tasks are more symbolic, tables work better. Many researchers have used cognitive fit theory to explain the effectiveness of information presentation for various tasks (Vessey, 2006). It is less clear how to apply this theory to our given problem because the distinction between spatial and symbolic representations is blurred for tabular visualizations, which contain both graphs and tables in some sense.

Several studies have been conducted to evaluate tabular visualization techniques. Wittenburg et al. (2001) reported that the participants selecting an item from a menu of multiple items generally preferred EZChooser, an implementation of Parallel Bargrams, over a spreadsheet application. Table Lens was also compared with S-PLUS in terms of exploratory data analysis (EDA) using the GOMS (Goals, Operators, Methods, and Selection rules) method (Pirolli & Rao, 1996). The researchers reported that Table Lens demonstrated comparable performance to Splus. Expanding heatmap views on tabular interfaces were also found to be effective on understanding the distribution of columns (Sopan et al., 2012). Value Charts was also evaluated, and Bautista and Carenini (2008) demonstrated that participants successfully made choices using this method.

Even though these evaluation studies provide some insights about how people use tabular visualizations, these results do not help answer our research questions directly. Wittenburg et al. (2001) only reported participants' subjective ratings, so it is difficult to know whether Parallel Bargrams actually improved decision quality, or if it was highly rated because of novelty effects. The evaluation should be designed to capture if the visualization is actually influencing the information process of individuals (Carswell, 1992). Pirolli and Rao (1996) used the GOMS method, which is based on predefined sets of tasks, but it is difficult to say that these predefined tasks are what people do in realistic situations. Some of low-level tasks or heuristics that people use to unburden cognitive loads may not be directly applicable because visual information provided by SimulSort and ParellelTable might influence participants' decision making procedure in an unconventional way.

3 Four Tabular Visualizations

Four different tabular visualization techniques shown in Figure 1 were used for this study. Except for the differences in cell arrangement, the four tabular visualizations shared a common look and feel. A highlighting feature was provided for all four tabular visualization techniques so that users could identify a row by hovering a mouse cursor over a cell (highlighted in yellow) or selecting a row (highlighted in green).

Baseline (B). A static table without any interaction techniques, except for the highlighting feature, was included as the "Baseline" (see Figure 1(a)).

Typical Sorting (TS). The next tabular representation that we used for our study was a table with the typical sorting feature, which allows the user to sort each column by clicking on the header of the column (see Figure 1(b)). Having this simple sorting feature could enhance the effectiveness of various cognitive activities, such as finding extreme values, finding a particular value, identifying general trends, and finding relationships. This table with the sorting feature is referred to as "Typical Sorting." As discussed earlier, the effectiveness of the Typical Sorting table could be limited when multiple attributes are considered and compared simultaneously.

SimulSort (SS). SimulSort was developed in order to overcome the limitations of Typical Sorting (Hur & Yi, 2009). Unlike Typical Sorting, SimulSort can mitigate the one-column sorting problem by sorting all of the columns at the same time (see Figure 1(c)). In order to find the corresponding attribute of each item, the user should hover the mouse cursor over an item or select it, and corresponding cells will be highlighted. Since all columns are sorted in descending order, cells near the top of the screen will have higher numbers, and cells near the bottom of the screen will have lower numbers within that column. Since cells in a column are sorted in the table format, two cells with the same value would be located in different rows, which may confuse some users. This suggests that if the values of two cells in one column happen to be the same, the vertical locations of two cells could provide a user with an incorrect perception of those values. The distortion may impact the simple tasks that Typical Sorting is more suitable for. Note that other features in SimulSort, such as horizontal bars and zooming, were suppressed in this study in order to clearly observe the effects of the simultaneous sorting feature.

ParallelTable (PT). ParallelTable is a variant of SimulSort that was designed to overcome the distortion problem of SimulSort. In contrast to SimulSort, the vertical location of a cell is precisely in line with the relative value of the cell (see Figure 1(d)). For example, if the relative value of a cell is 1.0 (or the maximum in the column), the cell will be located at the top of the column. If the relative value of a cell is 0.0 (or the minimum in the column), the cell will be located at the bottom of the column. This representation provides instant and visual understanding of the values of cells. However, we readily recognized some potential drawbacks of ParallelTable. First, since it may be perceived visually as cluttered due to overlaps among cells, users might not easily understand how to interpret ParallelTable. Second, when there is a tie, two cells will be perfectly overlapped with each other. Some interaction techniques (e.g., jittering) might help work around this issue, but this may introduce additional cognitive burden or visual distortion. In this paper, we intentionally did not introduce these work-around techniques in order to more clearly measure the effects of this occlusion problem.

4 Methods

4.1 Participants

A total of 80 undergraduate university students volunteered to participate in the study. Subjects earned $26 on average for their two-hour-long participations, where their earnings depended on their task performance during the experiments. Participants were randomly assigned to one of four experimental treatments, each of which is identical except for the tabular visualization employed. Thus, the four groups will be also referred to by the visualization techniques used (i.e., B, TS, SS, and PT). The four groups were homogenous based on self-reported demographic information, including gender, age, school year, major, grade point average (GPA), personal traits (e.g., openness to experience and political opinions), and the level of computer experience (e.g., comfort of using computer and years to use computers) (we used an electronic survey system, “z-Tree” (Fischbacher, 2007) to gather data).

We conducted an ANOVA test (for numerical measures) or Chi-Square test (for categorical measures) for each demographic variable separately. We did not find significant differences across treatment for gender ($\chi^2(3, N = 160) = 2.8605, p = 0.4136$), age ($F(3,74) = 0.6896, p = 0.5613$)), school year ($\chi^2(15, N = 160) = 20.7895, p = 0.1437$), major ($\chi^2(21, N = 160) = 24.39, p = 0.2745$), GPA($F(3,73) = 0.4717, p = 0.703$), openness to experience ($F(3,74) = 0.7101, p = 0.549$), political viewpoint ($F(3,74) = 1.1662, p = 0.3284$), or comfort and experience with computer use ($F(3,74) = 1.3194, p = 0.2745$). Degrees of freedom vary because some participants skipped some of the questions (e.g., GPA).

4.2 Procedures

The computer lab where the experiment was conducted can accommodate 20 students simultaneously. When participants arrived, they were randomly assigned to one of the 20 computer stations. Each station had a standard personal computer with a 19” monitor, a keyboard, and a mouse. A partition between stations prevented participants from seeing the computer screen of any other participant. Required documents including the instructions were provided at each station. An experimenter read the instructions out loud while participants followed along. Prior to beginning the experiment, a four-question quiz was given to the participants to verify their understanding of the tasks. After the quiz, participants were asked to participate in Tasks 1 and 2 (see Section 4.3) using the interface assigned for each treatment group. Figure 2 shows a screenshot of the experimental system in the SS treatment. As shown in the figure, the top portion of the screen displayed the currently selected item (item 09 highlighted in green), and the bottom left of the screen displayed the remaining time. For each round, we displayed different artificial datasets of 15 objects (labeled from “item 01” to “item 15” in the rows) with 7 attributes.

[Figure 2 about here.]

4.3 Tasks

Our goal in the experiment was to measure the impact of tabular visual representations for different types of tasks; therefore, we conducted an experiment with two types of tasks. Task 1 consisted of a series of object selection tasks, in which the user needed to select the most valuable item after considering multiple columns. While designing Task

1, we intended to mimic everyday decision making situations, in which a person may need to compare many options while considering multiple attributes. At the same time, we do not want participants' performance to be influenced by individual differences in knowledge and preferences, so we did not use a specific context and kept the task as abstract as possible. In contrast, Task 2 was designed to mimic simpler cognitive tasks that people often do on a table, such as retrieving values and finding extreme values in a column.

4.3.1 Task 1

More specifically, Task 1 was an incentivized experiment in which subjects were instructed to select the highest-valued item from the list of 15 items while considering all 7 columns, a task that we refer to as "multi-attribute object selection." The value of an item is equal to the sum of its column-wise values, and a column-wise value is valued from \$0 to \$1, depending on the attribute's numerical value relative to the other numerical values in the column. For example, when the value range of a column is between 50 and 70, the column-wise values for 50, 60, and 70 are \$0, \$0.5, and \$1.0, respectively. If a participant selects an item which has maximum column-wise values in all columns, the value of the item would be \$7. However, note that the probability of earning \$7 in this experiment is 0 because the average inter-attribute correlation (AIAC) was controlled at around 0.01 (this will be detailed in Section 4.4). More specifically, the value was calculated as follows:

$$value_i = \sum_{j=0}^{7} \frac{T_{ij} - \min T_{\cdot j}}{\max T_{\cdot j} - \min T_{\cdot j}} \tag{1}$$

where $value_i$ is the value of the i^{th} item, T is the whole data set (15×7), and T_{ij} is the value of i^{th} item and j^{th} attribute. Though this equation appears to be overly complex, we propose that this closely mimics consumer decision making, such as purchasing a used car while considering multiple attributes. Subjects participated in 20 rounds of Task 1 with different datasets in each round. Subjects had 3 minutes in each round to select the highest-valued item. They could either calculate the value of the item in working memory, or approximate the value by using the features specific to the treatment (for example, by sorting columns in TS, or by relying on visual cues in SS and PT).

4.3.2 Task 2

Task 2 was a series of seven different low-level analytic activities originated from Amar and Stasko's ten types of multi-attribute tasks (Amar, Eagan, & Stasko, 2005). Out of Amar and Stasko's (Amar et al., 2005) different tasks, three tasks - sorting, determining range, and computing derived value - were not included in our study for several reasons. First, sorting and determining range is very similar to a task we did use - finding extremum. Second, computing derived value can be easily supported directly (e.g., by having another row showing computed values). The seven tasks we used were as follows:

- *Retrieving value* (RV): Given a set of specific cases, find attributes of those cases (Please write down the value of Column A of item 12 in the following blank).
- *Finding extremum* (FE): Find data cases possessing an extreme value of an attribute over its range within the data set (Please write down the highest value of column B in the following blank).
- *Characterizing distribution* (CD): Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attributes values over the set (Please select the rank of the value of 63 in column C)[1].
- *Correlating* (Cor): Given a set of data cases and two attributes, determine useful relationships between the values of those attributes (Please select the column that is most highly correlated with column D? In other words, as the values of column D increase, which values of column increase most?).
- *Finding anomalies* (FA): Identify any anomalies within a given set of data cases with respect to a given relationship or expectation (There is an exception to the strong correlation between column B and column C. Please write down the item as the exception in the following blank).
- *Clustering* (C): Given a set of data cases, find clusters of similar attribute values (Please select how many groups of items with similar values in column E and column F).
- *Filtering* (F): Given some concrete conditions on attribute values, find data cases satisfying those conditions (Please select how many items that have less than 50 in column D, less than 30 in column E, and more than 70 in column F).

[1] We asked the rank of a certain value in a column instead of asking a shape of distribution because asking a shape of distribution become an open-ended question, which is difficult to be quantitatively evaluated.

We believe that the low-level tasks in Task 2 are relatively simpler than the multi-attribute object selection tasks in Task 1 because the former require the user to consider fewer columns than the latter. For example, in Task 2, the user is guided to consider one column for retrieving value, finding extremum, characterizing distribution, and correlating tasks, two columns for finding anomalies and clustering, and three columns for filtering; however, in Task 1, the user should (ideally) consider all seven columns to maximize the compensation received.

Subjects participated in two trials for each question, and each trial used different data sets. Subjects received \$0.50 for each correct answer, so that if a participant responded to all of the questions correctly, he or she was rewarded with \$7.00 (= 14 × \$0.50).

4.4 Data Sets

A total of 34 different data sets (20 for Task 1 and 14 for Task 2) were randomly generated in order for participants to see a different data set in any round of the experiment. Each dataset had fifteen items (rows) and seven attributes (columns), and each cell contained a two-digit numerical value from 10 to 99. In order to maintain the same task difficulty across datasets, we made sure that the average of inter-attribute correlations (AIAC) (Lurie, 2004) was controlled. The AIAC is the average of correlations between all combinations of two columns out of all columns. Based on our initial pilot trials, the dataset with the AIAC value of 0.01 made the problem an appropriate level of difficulty for the study. For example, we decided to use 15 objects, or rows, in Task 1 based on the finding from the pilot study that people in the baseline condition cannot calculate the values of more than 15 items within the allotted time limit.

4.5 Software

The experimental system was developed using Adobe Flex[2] with Flare[3] and Ruby on Rails[4] to make an experimental system that can be accessed by multiple participants simultaneously. The web-based system also collected detailed user interaction (e.g., mouse movements) and saved the collected data in the back-end MySQL database.

[2]http://www.adobe.com/products/flex/
[3]http://flare.prefuse.org/
[4]http://rubyonrails.org/

5 Results

5.1 Results: Task 1

[Table 1 about here.]

5.1.1 Decision Quality

[Figure 3 about here.]

Because the highest valued item in each set of objects changes in each round depending on the dataset, we use a "decision quality" measure, calculated as follows, to determine performance in Task 1:

$$Decision\ quality = \frac{value_i - \min(value.)}{\max(value.) - \min(value.)} \tag{2}$$

where $value_i$ is derived from Equation 1. Table 1 summarizes the decision quality of items selected and time (in seconds) spent in each round, where 1.00 is the highest decision quality and 0.00 is the lowest decision quality of an item in any given round. Figures 3 and 4 display the decision qualities of items chosen and time spent in each round for all treatments. We found that on average, subjects in B and TS attained decision qualities of 0.84 and 0.85, respectively, while subjects in SS and PT attained decision quality of 0.94 and 0.92, respectively.

A mixed model ANOVA with repeated measures was employed with type of visualization as the between-subjects factor and the number of a round as the within-subjects factor. The test shows the main effect of visualizations ($F(3,76) = 6.26, p = 0.0007$). We also conducted pair-wise comparisons with adjusted p values using the SIMULATE option in SAS for each visualization (Westfall, Tobias, & Wolfinger, 2011; Edwards & Berry, 1987). We expected that the ability to sort (TS), which is ubiquitous in many applications, would significantly improve decision quality relative to the Baseline (B). However, there is no statistically significant difference in decision quality between TS and B ($t(76) = -0.64, p = 0.9191$). This result necessitates the development of new systems to aid decision-making in this task, which is the goal of SimulSort and ParallelTable. Both SimulSort (SS) and ParallelTable (PT) increase decision quality as compared to the Baseline ($t(76) = 3.64, p = 0.0028$ and $t(76) = 2.99, p = 0.0194$, respectively).

Compared to Typical Sorting (TS), SimulSort (SS) showed significant decision quality differences ($t(76) = 3.00, p = 0.0189$). The difference between Parallel Table (PT) and Typical Sorting (TS) is marginal ($t(76) = 2.35, p = 0.0972$). These results suggest that visual analytics methods are effective for improving decisions in the multi-attribute object selection task. We did not find statistically significant differences between SimulSort (SS) and ParallelTable (PT) ($t(76) = -0.65, p = 0.9156$).

5.1.2 Time Spent

[Figure 4 about here.]

The time spent (in seconds) in each round was compared using mixed model ANOVA with repeated measures. On average, the participants in B and TS spent more time than the participants in SS and PT (94.72 and 97.88 seconds for B and TS as compared to 76.43 and 68.35 seconds for SS and PT). The visualization had a significant main effect ($F(3,76) = 3.64, p = 0.0163$). However, when we conducted pair-wise comparisons with adjusted p values using simulation, the only significant difference in time spent was observed between Parallel Table (PT) and Typical Sorting (TS) ($t(76) = -0.87, p = 0.0206$), while the other five pair-wise comparisons did not show statistically significant differences (all p values > 0.10). As can be seen in figure 4, there is a general trend that SimulSort and Parallel Tables may help users make more efficient choices. Some of participants tend to use full 3 minutes, and we believe that this aspect may dilute the differences between visualization techniques in terms of efficiency of decision making.[5]

In summary, we found that Typical Sorting (TS) did not result in choices that were significantly different from the Baseline (B). However, both SimulSort (SS) and Parallel Table (PT) significantly increased the decision quality of the final choice selected and marginally decreased the amount of time spent compared to Baseline and Typical Sorting.

5.1.3 Elicited Confidence

At the end of every round and before the outcome was revealed to subjects, we also asked subjects how confident they were that their choice was the best one. [6] Previous work has suggested that graphical displays improve confidence in

[5] We also employed various transformation techniques (logarithmic, square-root, and box-cox transformations) to alleviate the skewness of the data, but these transformations did not reveal any new findings.

[6] We asked, “How confident are you that you made the best choices in this round?” Participants answered using a 7-point Likert scale from “Very confident” to “Not at all confident.”

various decision-making tasks. Related studies have used a similar approach for rating confidence (Kamis & Stohr, 2006; Adidam & Bingi, 2000), and previous research has suggested that the use of an interactive visual interface may improve confidence (Sharkey, Acton, & Conboy, 2009). Increased confidence may result in better outcomes from decision-making, for example, increased confidence can result in following through with a financial plan (Lusardi & Mitchell, 2005).

We find that confidence is increased with the use of PT and SS, but not with TS. We find statistically significant differences comparing B with SS or PT (Z-scores = 2.42, 2.78; all p-values < 0.01). However, we do not find statistically significant differences between TS and B (Z-score = 0.89 p-value > 0.10).

5.1.4 Regression of Influencing Factors on Decision Quality

[Table 2 about here.]

We also conducted a series of tobit random effects panel regressions for each treatment (with subject random effects upper-censored at 100 and lower-censored at 0), regressing time spent, confidence, a round trend, gender, and comfort using the computer, new software or sorting variables on decision quality measured from 0% to 100%, the results of which are reported in Table 2.[7] Our primary finding was that learning was present in all rounds, so that over time individuals learned to perform better in the task through use of the interface (1/round is negative for all treatments and statistically significant in TS and SS). We found that coefficients on time spent in each round were low in magnitude. Comfort in computer use was associated with improved performance in the B treatment, but not in other treatments. Comfort with new software was actually associated with worse performance in the TS treatment, and we do not have an intuitive reason for this result. Comfort with sorting was not statistically significantly linked in any treatment.

5.2 Results: Task 2

Since we posit that performance of the visualizations would vary depending on different tasks, we conducted Task 2 to see the impacts of different tasks. We employed a logistic regression analysis with accuracy of each activity because

[7] (†) represents p values below 0.10, (*) Represents p-values below 0.05, and (**) represents p-values below 0.01).

outcomes are binary (whether they got it right or wrong). We also employed linear mixed models to analyze response time because response time data do not satisfy the assumption of homogeneity of variance.

First, we found that the second trial took statistically significantly less time than the first trial in all the activities except for filtering extremum (p-value < 0.01 for retrieving value, < 0.01, p-value < 0.05 for correlating, < 0.01 for filtering anomalies, < 0.01 for clustering, and < 0.01 for filtering), indicating that learning effects exist.

However, the effects of visualization techniques varied depending on the characteristics of the low-level tasks. For three activities, including retrieving value, finding extremum, and finding anomalies, we find no significant differences in time and accuracy between visualizations. One of potential explanations for the lack of differences between visualizations is that these tasks are too easy for subjects. These tasks required participants to consider only one column and produced the highest average accuracy among the seven tasks, which may mean that these tasks were too easy to accomplish and subjects did not get many benefits from more extensive visualization methods.

For the other four tasks, including characterizing distribution, correlating, clustering, and filtering, we observed the advantage and disadvantage of four different tabular visualizations:

For the characterizing distribution task, we found that subjects in PT performed significantly less accurately than ones using the other techniques (p-values for PT vs. B, TS, and SS are < 0.05, 0.01, and 0.01, respectively) as shown in Figure 5, and subjects in B performed significantly slower than subjects using other techniques (p-values for B vs. TS, SS, and PT are < 0.001, 0.001, and 0.05, respectively) as shown in Figure 6. In both figures, the significant differences are shown using arrows. Interestingly, not having any sorting technique (B) only slowed down participants, but their accuracy in conducting the task was not degraded. However, since we asked participants to identify the rank of a particular value, the occlusion that occurred in the PT condition appeared to give participants difficulties. Having the sorting features without occlusion in TS and SS helped participants perform faster and more accurately.

[Figure 5 about here.]

For the correlating task, participants in PT performed significantly less accurately than ones in B, TS, and SS (p-value < 0.01, < 0.05, and < 0.01, respectively). Again, we believe that the occlusion in PT makes it difficult to interpret the relationship between two columns. We do not find any statistically significant differences in time spent among four conditions.

The clustering task, which involves two attributes, turned out to be one of the most difficult tasks out of seven. Participants in all four conditions performed very poorly. No statistically significant differences in accuracy are found among the four groups. However, participants in TS performed more slowly than ones in B and PT (p-value < 0.05 and < 0.01, respectively) as shown in Figure 6. One potential explanation for the slow performance in TS is that participants switched sorting column multiple times since the clustering task involved two columns. When using the other tools, participants did not have the sorting feature (as in B) or did not need to sort because the two columns were already sorted (as in in SS and PT).

For the filtering task, which involved three attributes, we found that participants in TS performed more accurately than participants in B and PT (p-value < 0.01 and < 0.05, respectively); participants in SS performed more accurately than participants in B (p-value < 0.05) as shown in Figure 5. Interestingly, although this task involved multiple columns, TS seemed to be the most appropriate technique. No statistically significant difference was found between TS and SS. One difficulty that we observed in conducting the filtering task using SS and PT is that one needs to highlight a cell to see which item the cell belongs to, which may result in some errors due to ruling out filtered items. Participants in PT performed even less accurately, probably due to the occlusion.

[Table 3 about here.]

[Figure 6 about here.]

6 Discussion

In summary, we found that no single sorting technique (or tabular visualization) was the definite winner. Each visualization technique showed advantages and disadvantages depending on task, which are summarized in Table 3.

B vs. TS. To our surprise, subjects in TS did not significantly outperform ones in B in the most of tasks. Subjects in B and TS demonstrated compatible performance both in time and decision quality in most of the tasks except for the following three cases: time in characterizing distribution (TS $>$ B), accuracy in filtering (TS $>$ B), and time in clustering (B $>$ TS). One explanation that we have is that the number of items (15) used in this experiment is not large enough to penalize B, especially for tasks in Task 2. We expect that the differences between the two techniques will become more salient when the data set becomes larger.

SS vs. PT. We were initially interested in comparisons between SS and PT because we introduced PT to overcome the distortion problem of SS. However, PT also has the potential occlusion problem. The results of our experiments generally showed that the occlusion problem turns out to be more serious than the distortion problem, especially in the low-level analytic tasks in Task 2. In Task 1, in multi-attribute object selection tasks, subjects in SS and PT showed comparable results both in decision quality and time. Subjects in PT performed slightly better than subjects in SS, but the differences are not statistically significant. However, in Task 2, occlusion caused some serious issues. More specifically, in characterizing distribution, correlating, and filtering tasks, subjects in PT performed significantly less accurately. Thus, if one tabular representation needs to support various cognitive tasks including the three latter tasks, PT may not be a strong design candidate. The costs of using PT outweigh the benefits of PT, and SS could be a better alternative. Interestingly, the distortion problem of SS was less salient than expected, so subjects in SS could performed well both in multi-attribute object selection tasks and low-level analytic tasks. Of course, the occlusion problem of PT could be overcome by employing other visualization (encoding the density of elements using transparency or split a cell horizontally into multiple columns) or interaction (jittering) techniques. The effectiveness of these approaches should be tested in future studies.

SS vs. TS. In Task 1, subjects in SS outperformed subjects in TS in object selection, and these two are generally comparable in Task 2. This shows that applying the idea of parallel coordinate on a tabular visualization could be effective on supporting a multi-attribute object selection task. However, when we analyzed the decision strategies that subjects reported, it was unclear how subjects in SS performed differently from subjects in TS.

Since the strategy responses were open-ended, the level of detail in the description varied and was often unclear. In order to better understand strategies employed by subjects, we conducted an eye tracker study with five subjects. The five subjects recruited had already participated in the lab-controlled study.Two of the subjects were from the TS condition, and the other three were from SS. The subjects completed the same tasks, and the only difference was that their eye-gazes were recorded using an eye tracker (Tobii X60) to investigate strategies.

[Figure 7 about here.]

Figure 7 show heat maps generated from two representative participants (one in TS and one in SS). The other three participants' heat maps showed generally similar patterns. Figure 7(a) shows the location of the subject's gaze while using TS. The red area indicates that fixation occurred mostly on the item number column. We assume that participants might have constantly checked the position of the item number for comparison as the table is shuffled due to the one-column sorting problem. Moreover, the attraction of fixations gradually decreased towards the columns on the right side of the table, implying that people may consider the left most column more than other columns. This might indicate that the subject is limited by the number of items that he or she can mentally note while using TS. Since subjects are not allowed to use any external memory (such as a pen and paper), this result could simply show the limitation of short-term memory (Miller, 1956), which should be verified in future studies.

In contrast, for SS (Figure 7(b)), the fixations are concentrated on the center of the table. The interesting part is that the fixation was not spread out evenly over multiple columns, but was concentrated at the central region of the screen. We do not believe that this result implies that subjects in SS only look at the columns in the centers (the fourth column) because this strategy would lower the overall decision quality. A more compelling explanation would be that subjects see overall visual patterns of selected or hovered items using both central and peripheral vision. These heat maps show that a visualization tool, SimulSort in this case, changed the nature of decision-making from an information processing task (e.g., reading a value, memorizing the value in short-term memory, comparing values, and making a decision) to a visual perception task.

7 Conclusions

The major contribution of our study was to provide empirical evidence of how different tabular visualizations actually affect performance in object selection tasks and in various low-level data analysis tasks. In addition, this study utilized several novel multidimensional visualizations, which were designed to support multi-variate decision-making. First, SS sorts all of the attributes in a table format simultaneously and also uses visual cues. Second, PT provides sorted data in a table view with absolute, rather than relative, graphical distance between attributes. These visualizations were developed to solve the limitations of typical sorting for multi-variate decision-making.

The strength of this study was to conduct a carefully controlled evaluation of the effect on performance of using different tabular visualizations. A controlled environment with incentive rewards and carefully designed datasets provided us with the chance to observe object selection and multi-variate data analysis.

There are several limitations in our study. We incentivized participants only on their response accuracy, so we noticed that some participants spent the whole three minutes that were allowed to them. For some tasks, these three minutes could be much more than sufficient. In addition, we only used 15 items for our study after adjusting time and task difficulty with two digit numbers in B which provides no interactive function. 15 items may be too small to show the effectiveness of sorting features. In our study, some activities asking for single-attribute exploration could be performed with very high accuracy even when not using typical sorting features, which produced no significant differences between treatments. More items also make the table overflow the screen, which caused some of the visual highlights for a row fall outside the screen in the cases of SimulSort and ParallelTable. Horizontal bars and zooming features that are disabled for this experiment may help overcome this problem, but we do not know the impacts of these additional features.

In spite of these limitations, we were able to provide clear empirical findings outlining the strengths and weaknesses of four tabular visualizations. While Typical Sorting generally supports information processing activities, SimulSort and ParallelTable successfully support for the multi-attribute object selection task. In the future, we plan to conduct additional studies with different data settings (bigger datasets, different average inter-correlation, and different distribution) and additional interaction techniques, such as filtering.

8 Future Work

This study is one of early studies of a larger research theme, called "visualized decision making," which promotes synergy between information visualization and decision science. This study showed *whether* a specific visualization technique helps decision making and other cognitive tasks, but it barely explained *how much* and *why*. The results from the eye tracking study provide some hints about what kinds of cognitive processes research participants experienced, but the results are not yet sufficient to provide a detailed picture of the phenomenon. In addition, it is also unclear whether such benefits of visualization can be obtained in more realistic situations. The experiment was conducted in a very controlled environment without any realistic decision making context, so the impacts of realistic contexts are worthwhile to investigate. In addition, since some of the match and mismatch between task and display modality can be modulated by time constraints, future research should also investigate the effect of incentivizing rapid decisions. Finally, there are many other visualization techniques that have potential to promote decision making, such as Dynamic Query (Ahlberg & Shneiderman, 1994) and Dust & Magnet (Yi, Melton, Stasko, & Jacko, 2005). These techniques could be surveyed more comprehensively and investigated thoroughly to deepen our understanding of "visualized decision making."

9 Acknowledgements

This work has been partially funded by the U.S. National Science Foundation (Grant No. 09064963), and we also thank the students workers who participated in this study.

References

Adidam, P., & Bingi, R. (2000). The importance of decision confidence to strategy outcomes. *Journal of Applied Business Research*, *16*(2), 35–50.

Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI conference on human factors in computing systems: celebrating interdependence* (pp. 313–317). Boston, Massachusetts, United States: ACM.

Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. In *Proceedings of ieee symposium on information visualization* (pp. 111–117).

Bautista, J., & Carenini, G. (2008). An empirical evaluation of interactive visualizations for preferential choice. In *Proceedings of the working acm conference on advanced visual interfaces* (pp. 207–214).

Bertin, J. (1983). *Semiology of graphics*. University of Wisconsin Press.

Carenini, G., & Loyd, J. (2004). ValueCharts: analyzing linear models expressing preferences and evaluations. In *Proceedings of the working acm conference on advanced visual interfaces* (pp. 150–157).

Carswell, C. M. (1992, October). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(5), 535–554.

Edwards, D., & Berry, J. J. (1987). The efficiency of simulation-based multiple comparisons. *Biometrics*, *43*, 913-923.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Hur, I., & Yi, J. S. (2009). SimulSort: Multivariate Data Exploration through an Enhanced Sorting Technique. In *Proceedings of the 13th international conference on human-computer interaction* (Vol. LNCS 5611, pp. 684–693).

Jun, E., Landry, S., & Salvendy, G. (2011). A visual information processing model to characterize interactive visualization environments. *International Journal of Human-Computer Interaction*, *27*(4), 348–363.

Kamis, A., & Stohr, E. (2006). Parametric search engines: What makes them effective when shopping online for differentiated products? *Information & Management*, *43*(7), 904–918.

Lurie, N. (2004). Decision making in information-rich environments: The role of information structure. *Journal of Consumer Research*, *30*(4), 473–486.

Lusardi, A., & Mitchell, O. (2005). Financial literacy and planning: Implications for retirement wellbeing. In *the retirement research consortium towards a secure retirement system.*

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81–97.

Ntuen, C. A., Park, E. H., & Gwang-Myung, K. (2010). Designing an information visualization tool for sensemaking. *International Journal of Human-Computer Interaction*, *26*(2-3), 189–205.

Pirolli, P., & Rao, R. (1996). Table lens as a tool for making sense of data. In *Proceedings of the acm workshop on advanced visual interfaces* (p. 80).

Rao, R., & Card, S. (1994). The table lens: Merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the acm sigchi conference on human factors in computing systems* (p. 322).

Sharkey, U., Acton, T., & Conboy, K. (2009). Modelling the Effects of Decision Tools in Online Shopping. *BLED*

2009 Proceedings, 10.

Sopan, A., Freire, M., Taieb-Maimon, M., Plaisant, C., Golbeck, J., & Shneiderman, B. (2012). Exploring data distributions: Visual design and evaluation. *International Journal of Human-Computer Interaction*. (Advance online publication: DOI: 10.1080/10447318.2012.687676)

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*(2), 219–240.

Vessey, I. (2006). The theory of cognitive fit: One aspect of a general theory of problem solving? In P. Zhang & D. Galletta (Eds.), *Human-computer interaction and management information systems: Foundations, advances in management information systems series* (pp. 141–183). Armonk, NY: ME Sharpe.

Westfall, P. H., Tobias, R. D., & Wolfinger, R. D. (2011). *Multiple comparisons and multiple tests using SAS*. SAS Institute.

Wittenburg, K., Lanning, T., Heinrichs, M., & Stanton, M. (2001). Parallel bargrams for consumer-based information exploration and choice. In *Proceedings of the acm symposium on user interface software & technology* (pp. 51–60).

Yi, J. S., Melton, R., Stasko, J., & Jacko, J. A. (2005). Dust & magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, *4*, 239–256.

Inkyoung Hur is a Ph.D. student in the department of Decision Sciences and Information Systems at Florida International University. Her research is aimed at information visualization for helping in decision-making with multi-attributes. Hur received her masters degree in Industrial Engineering from Purdue University.

Sung-Hee Kim is a Ph.D. student in the School of Industrial Engineering at Purdue University. The main focus of her research is applying information visualization techniques to improve individuals decision-making processes.

Anya (Savikhin) Samak is an assistant professor at the University of Wisconsin-Madison School of Human Ecology. Savikhin is an applied microeconomist specializing in experimental and behavioral economics. Savikhin uses the methodology of experimental economics to explore decision-making in different contexts. Savikhin received her Ph.D. in Economics from Purdue University in 2010.

Ji Soo Yi is an assistant professor in the School of Industrial Engineering at Purdue University. Yis research interests are mainly on applying information visualization techniques to various decision making and building empirically tested theories about how visualization techniques could be used to support decisions. Yi received his Ph.D. in Industrial and Systems Engineering from Georgia Institute of Technology in 2008.

List of Figures

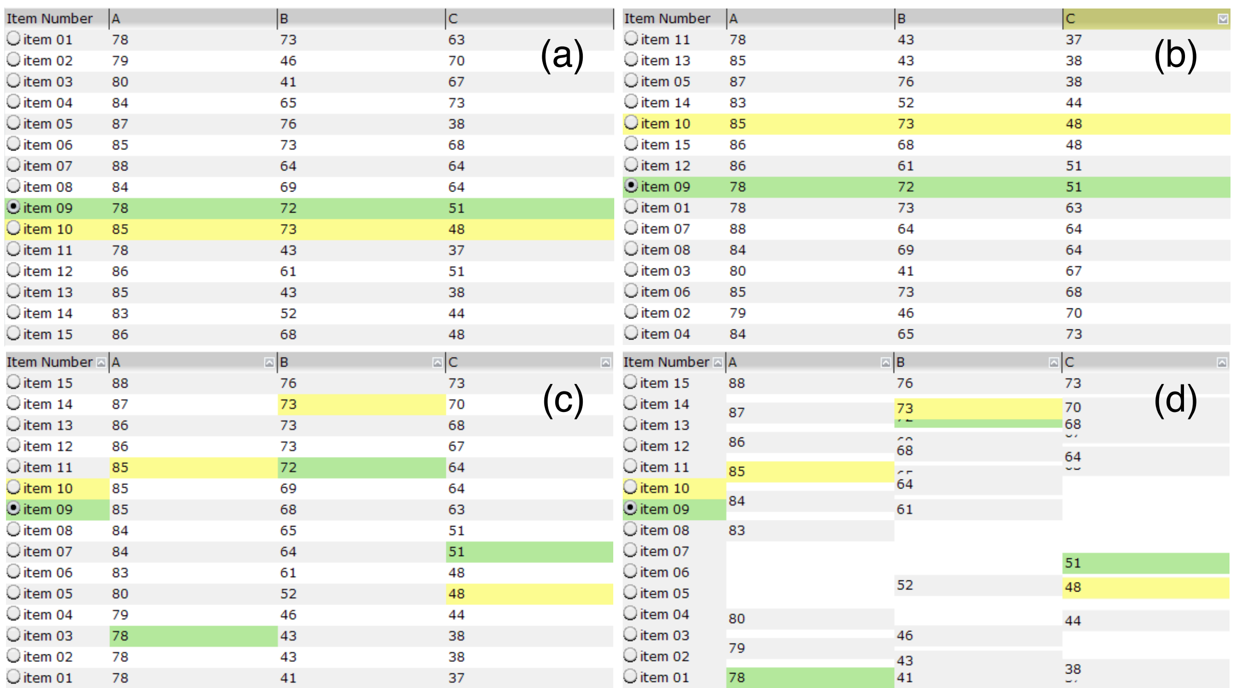


Figure 1: Screen shots of the four visualizations: (a) a static table (Baseline: B); (b) a table with one-column sorting (Typical Sorting: TS); (c) a table with all columns simultaneously sorted (SimulSort: SS); and (d) a table with all column sorted and faithful vertical locations (ParallelTable: PT).

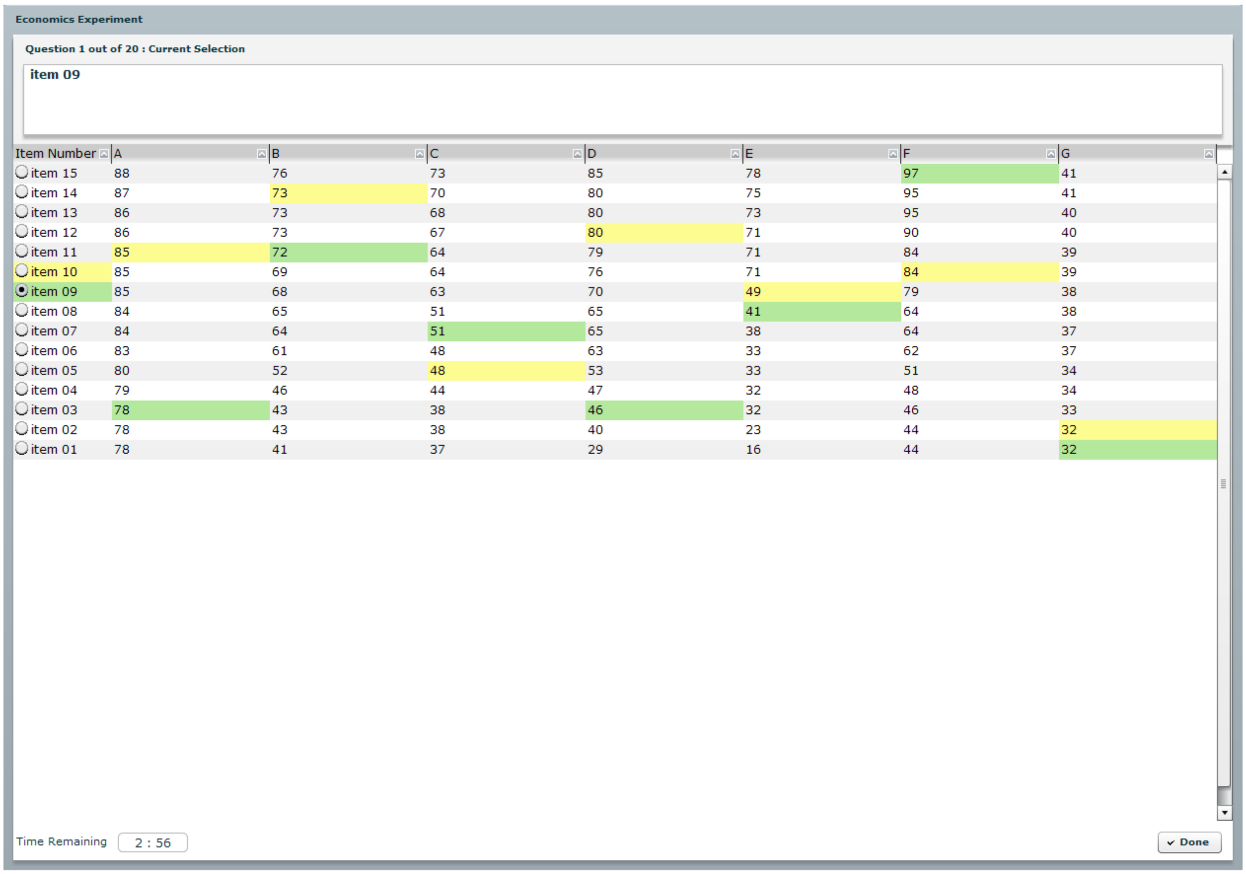


Figure 2: A screenshot of the interface for research participants in the SimulSort treatment.

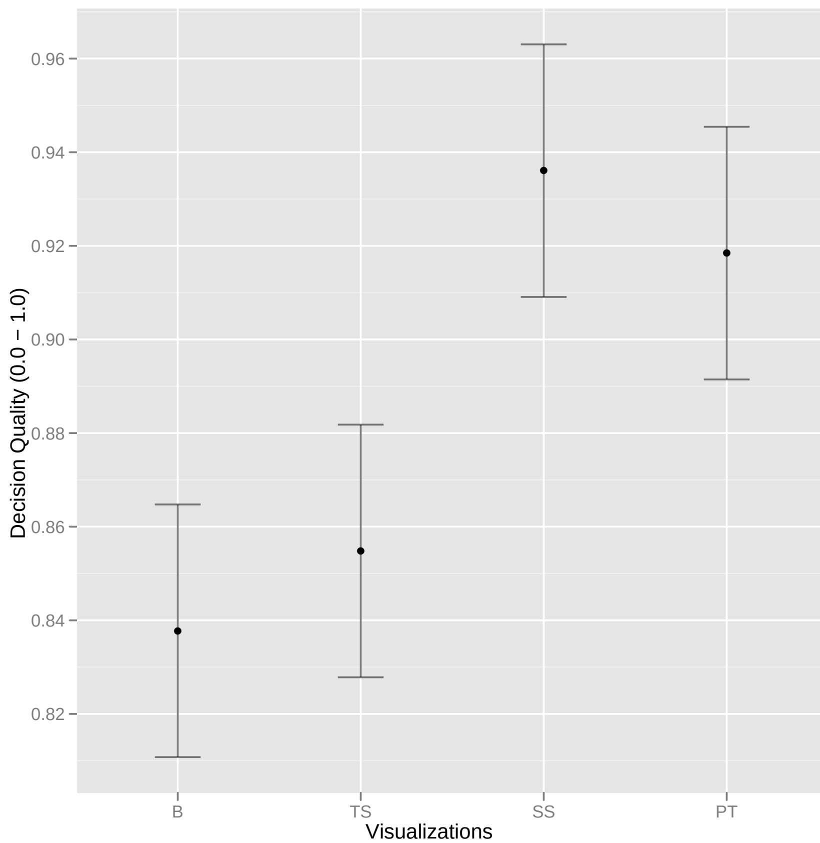


Figure 3: Decision qualities shown in four visualizations in Task 1.

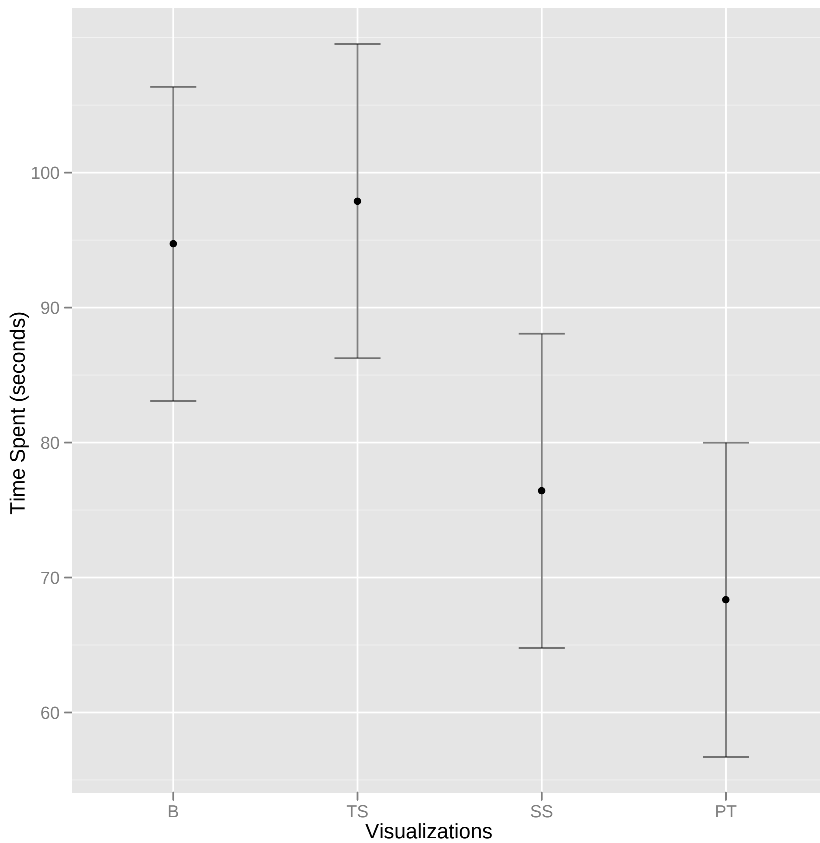


Figure 4: Time spent for the four visualizations in Task 1.

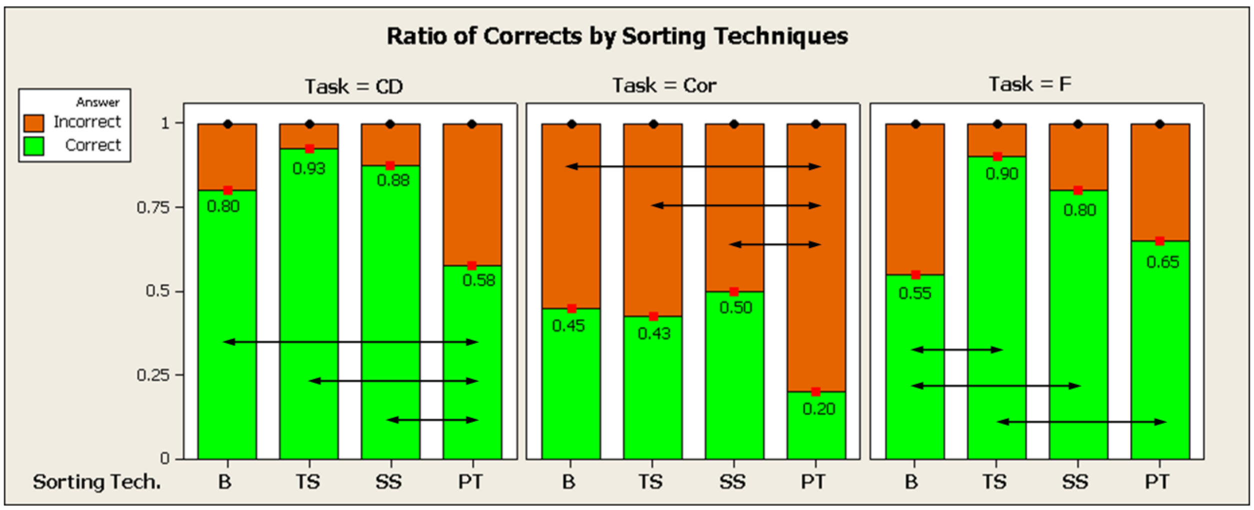


Figure 5: Ratio of correct answers of three low-level analytic tasks: characterizing distribution (CD), correlating (Cor), and filtering (F). The arrows indicate statistically significant differences at the error level of 0.05.

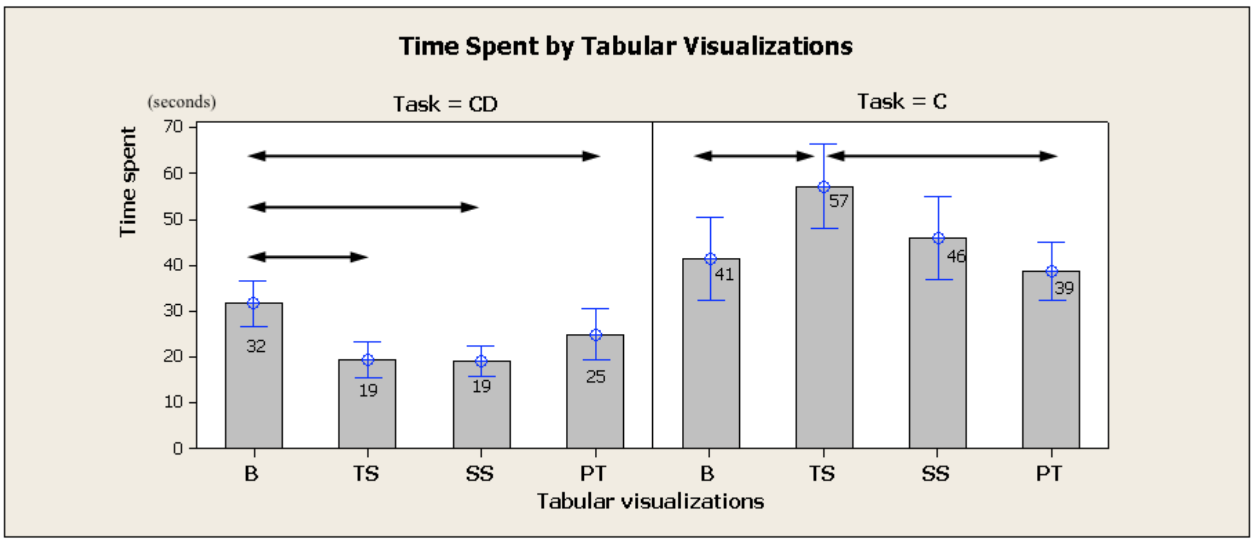
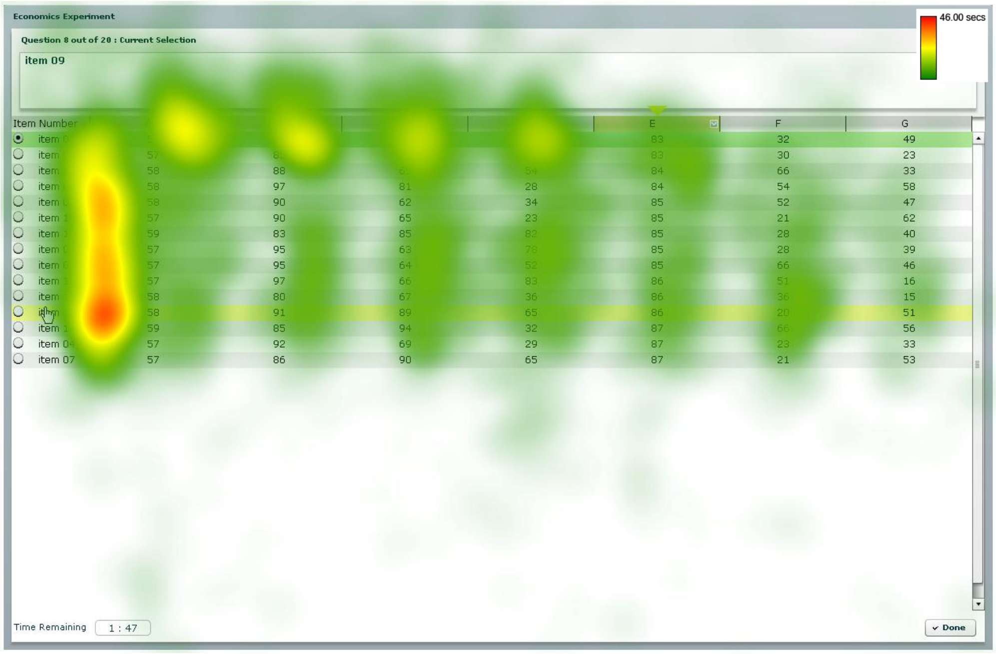
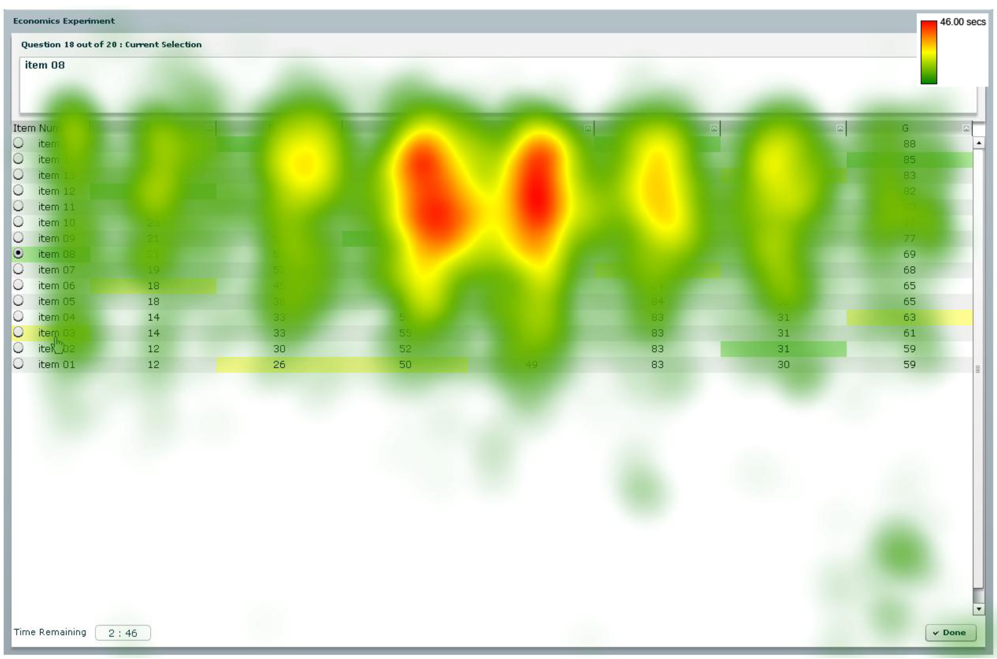


Figure 6: Time spent of two low-level analytic tasks: characterizing distribution (CD) and clustering (C). The arrows indicate statistically significant differences at the error level of 0.05.



(a) TS



(b) SS

Figure 7: A heat map generated from participants from TS and SS.

List of Tables

Table 1: Summary of results from Task 1.

Visualizations	Measures	Mean	SD	Min	Max
B	*Decision Quality*	0.84	0.21	0.00	1.00
	Time Spent (sec)	94.72	51.23	3.38	180.00
TS	*Decision Quality*	0.85	0.19	0.00	1.00
	Time Spent (sec)	97.88	47.73	7.29	180.00
SS	*Decision Quality*	0.94	0.10	0.39	1.00
	Time Spent (sec)	76.43	47.53	5.94	180.00
PT	*Decision Quality*	0.92	0.15	0.00	1.00
	Time Spent (sec)	68.35	41.08	11.06	180.00

Table 2: Tobit regressions by treatment.

Dependent Vars.	**B**	**TS**	**SS**	**PT**
Learning	-5.45	-10.99^{*}	-9.66^{**}	-4.07
1/Round	(5.64)	(5.26)	(3.36)	(4.20)
Time	0.05	-0.09^{*}	0.02	$-0.05^{\dagger}$
Seconds	(0.03)	(0.04)	(0.02)	(0.03)
Gender	-5.48	-6.34	-0.49	-5.31
1=male	(3.94)	(4.52)	(2.51)	6.25
Computer Use	5.19^{**}	4.15	-0.79	-0.84
1-7 comfort	(1.77)	(3.89)	(0.98)	(3.82)
New Software	-0.28	-4.63^{**}	-0.38	1.30
1-7 comfort	(1.45)	(3.89)	(0.58)	(2.51)
Sorting Use	-2.68	-0.95	0.15	2.61
1-7 comfort	(2.97)	(3.48)	(1.86)	(3.97)

Table 3: Summary of task performances using four visualizations.

Rank	**Task 1: Object Selection**		**Task 2: Low-level Analytic Tasks**				
			Characterizing distribution		**Correlating**	**Clustering**	**Filtering**
	Decision Quality	Time	Accuracy	Time	Accuracy	Time	Accuracy
1	SS	PT	SS	SS	SS	PT	TS
2	PT	SS	TS	TS	TS	B	SS
3	TS	B	B	PT	B		PT
4	B	TS	PT	B	PT	TS	B