Eurographics Conference on Visualization (EuroVis) 2015
H. Carr, K.-L. Ma, and G. Santucci
(Guest Editors)

Volume 34 (2015), Number 3

A Multi-task Comparative Study on Scatter Plots and Parallel Coordinates Plots

Rassadarie Kanjanabose[1], Alfie Abdul-Rahman[2], and Min Chen[2]

[1]Department of Computer Science and [2]Oxford e-Research Centre, University of Oxford

Abstract

Previous empirical studies for comparing parallel coordinates plots and scatter plots showed some uncertainty about their relative merits. Some of these studies focused on the task of value retrieval, where visualization usually has a limited advantage over reading data directly. In this paper, we report an empirical study that compares user performance, in terms of accuracy and response time, in the context of four different visualization tasks, namely value retrieval, clustering, outlier detection, and change detection. In order to evaluate the relative merits of the two types of plots with a common base line (i.e., reading data directly), we included three forms of stimuli, data tables, scatter plots, and parallel coordinate plots. Our results show that data tables are better suited for the value retrieval task, while parallel coordinates plots generally outperform the two other visual representations in three other tasks. Subjective feedbacks from the users are also consistent with the quantitative analyses. As visualization is commonly used for aiding multiple observational and analytical tasks, our results provided new evidence to support the prevailing enthusiasm for parallel coordinates plots in the field of visualization.

1. Introduction

Parallel coordinates plot (PCP) [Ins85] has been a favorable visualization technique among many, if not most, visualization researchers (e.g., [Weg90,tMR07,AR11,HW13]). However, there has also been some doubts about its effectiveness [HHB07, LMv08, HvW10], and it is often perceived to be difficult to learn and understand [SLHR09]. Recently, Kuang *et al.* reported an empirical study comparing PCP and scatter plot (SP) [KZZM12]. Their results indicated that PCPs were advantageous with datasets of low dimensionality and low density, whereas SPs performed better in handling the datasets with higher dimensionality and density. This study cast further doubts about the effectiveness of PCPs.

We noticed that the study by Kuang *et al.* [KZZM12] focused on a particular visualization task, that is, retrieving values of data from visual representations. Such a task has been considered less critical in data analysis (e.g., [Pla04]). We wondered whether the accuracy of performing value retrieval tasks using either SPs or PCPs will likely be poorer than reading data from a data table directly. If this is true, value retrieval tasks may not be the most representative tasks for comparing SP and PCP.

This supposition motivated us to conduct an empirical study to compare SP and PCP based on several tasks. We controlled our stimuli with three main variables:

Four tasks: *value retrieval*, *clustering*, *outlier detection*, and *change detection*;
Three representations: *data table* (DT), *scatter plot* (SP), and *parallel coordinates plot* (PCP);
Three levels of task difficulties: relatively *easy*, *medium*, and relatively *hard*.

As an individual study has a limited capacity for exploring multiple variables, we fixed other variables (e.g., the number of multivariate data points and the number of data dimensions) to minimize potential confounding effects. Prior to our study, we have the following hypotheses:

H1: For value retrieval, DT $\succ$ PCP $\succ$ SP.
H2: For clustering, PCP $\succ$ SP $\succ$ DT.
H3: For outlier detection, PCP $\succ$ SP $\succ$ DT.
H4: For change detection, PCP $\succ$ SP $\succ$ DT.

where $\succ$ denotes that the data representation on the left would result better performance than that on the right. If our study could not support above hypotheses, it would cast further doubts about the effectiveness of PCP. Considering that PCP is less intuitive than SP, while learning to use is more

difficult, it would suggest that the enthusiasm about PCP in the community might not be justifiable.

The results of our study have showed conclusively that hypotheses **H2**, **H3**, and **H4** are correct. The results have offered partial support to **H1**, but not a conclusive confirmation. The study has confirmed that PCP is advantageous over SP in performing several types of visualization tasks.

2. Related Work

The history of scatter plot (SP), or scatter diagram as it was called previously, can be traced back to the 18th century, and since then it has made remarkable contributions to the advancement of science [FD05]. It has been widely used as a form of statistics graphics, from which many variations (including its matrix form) have been derived [CM84, Cle85]. The uses of SP and its variations are ubiquitous across almost all disciplines (e.g., [TGS04, KD09]).

The history of parallel coordinates plot (PCP) can be traced back to the 19th century [GH83]. It became more widely used in the 1970s following the significant contribution of Inselberg [Ins85, Ins09]. A recent survey by Heinrich and Weiskopf [HW13] shows that over the past two decades the visualization community has displayed a high level of interest and enthusiasm in PCP and has invested a huge amount of effort in developing this technique.

PCP and the matrix form of SP are commonly used to depict multivariate data [Cle85, WB97, HG01, HW13]. Both techniques can support several visualization tasks such as observing clusters, outliers, and correlations. They can assist in analytical tasks such as classification, regression, and summarization. They can also be used to depict temporal data, aiding visualization tasks such as change detection, dependency reasoning, and identify temporal trends. There is a huge volume of works on both visualization techniques, and here it is only feasible for us to highlight a few examples focusing on works featuring both representations.

Wegman juxtaposed the patterns in PCPs with those of SPs [Weg90]. Yuan *et al.* proposed an integrated visual representation featuring SP and PCP [YGX*09]. Viau *et al.* designed another hybrid representation called "parallel scatterplot matrix" for aiding network visualization [VMCJ10]. Heinrich *et al.* used progressive splatting for rendering the continuous form of SP and PCP [JSD11]. Tam *et al.* added a 1D SP to each axis of a PCP, and used this integrated visualization to discover a decision tree as a classifier [TFA*11].

Previous works have evaluated PCP in various applications in terms of its usability [tMR07, JFLC08, AR11], and learnability [SR06, SLHR09] positively. However, comparisons between SP and PCP by several studies have been largely in favor of SP. Henley *et al.* compared the two techniques in the context of genome comparison, and they found that SPs were preferred in overall performance [HHB07]. Li *et al.* examined the two techniques for supporting correlation analysis of bivariate data, and concluded that SPs were more effective than PCPs [LMv08]. Holten and van Wijk compared variants of PCPs in cluster identification performance, and found that the variation with embedded SPs noticeably outperformed other variants [HvW10]. Note that this is not a direct comparison between SPs and PCPs.

Kuang *et al.* conducted a comparative study on multivariate data visualization using SP and PCP [KZZM12]. They examined user performance in value retrieval tasks between PCPs and three variants of SPs. The results showed that PCPs performed better under conditions where datasets were of lower attribute dimensions and data points were less dense. On the other hand, SPs performed better with datasets exhibiting higher dimensionality and dense data points.

As PCP is designed for supporting several different visualization tasks in multivariate data analysis, there are large gaps not covered by these studies. Our study was intended to fill in some of the gaps, in particular, to compare PCPs and SPs for supporting clustering, outlier detection, and change detection in multivariate data analysis.

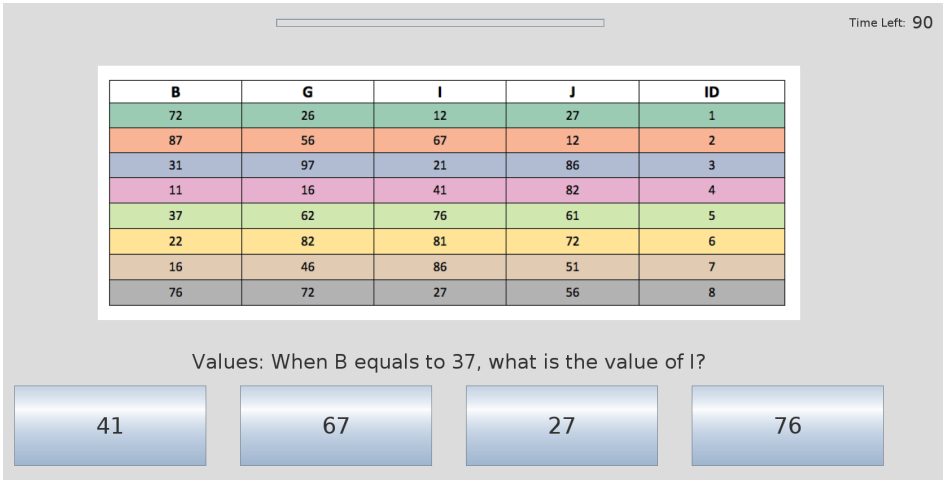
3. Experiment Overview

This empirical study was conducted in a laboratory environment. The main objective was to evaluate the four hypotheses listed in Section 1. The principal design criteria for the experiment were (i) featuring a reasonable number of visualization tasks in addition to the value retrieval task, (ii) featuring a reasonable number of data dimensions (> 2D), and (iii) using data table (DT) as a common reference representation for scatter plot (SP) and parallel coordinates plot (PCP).

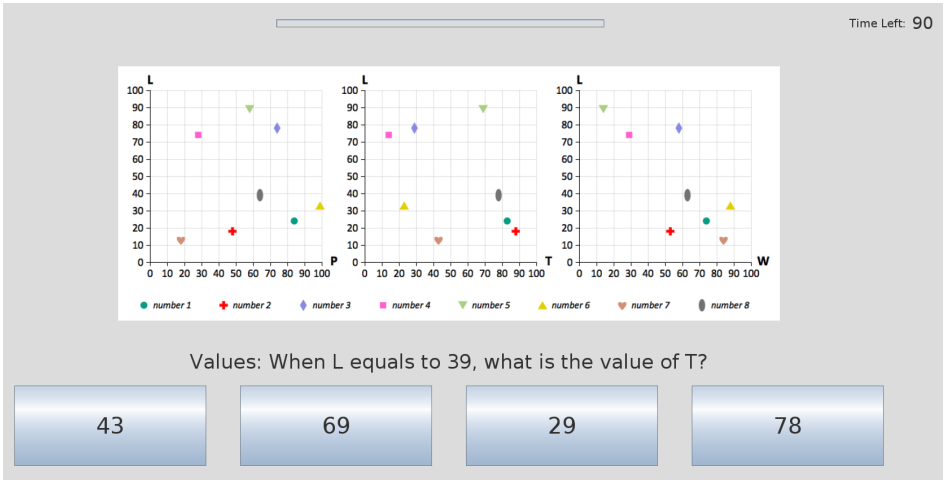
There are many variables that may potentially affect the performance of participants. Similar to most lab-based studies in visualization and cognitive sciences, we focus this study on a small number of variables while restricting others in order to minimize potential confounding effects.

Our study has three independent variables namely: *data representations*, *visualization tasks*, and *levels of task difficulty*. In the following subsections, we describe these three independent variables in detail.

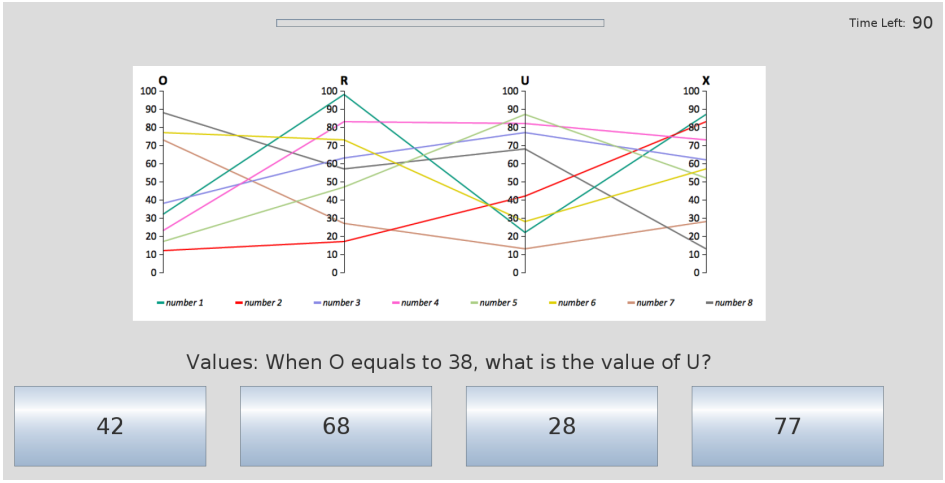
The study focused on two dependent variables, namely *accuracy* and *response time*, which are classical measurements for user performance. We used a combination of multiple choice questions and answers to obtain measurements for these two variables. Accuracy was measured as a ternary variable of *correct answer*, *incorrect answer*, and *no answer*. We anticipate that it may be very difficult to perform certain visualization tasks with a DT. We thus set a time limit for each task, and used a progress bar as a countdown timer. Any non-completion of a task is defined as a *no answer*. Response time was measured, in milliseconds, from the point when a given data representation was shown on the screen, to the time when the participant selected an answer. To avoid



(a) Data table (DT)



(b) Scatter plot (SP)



(c) Parallel coordinates plot (PCP)

Figure 1: *Screenshots of three example stimuli for a value retrieval task at the easy level in terms of task difficulty.*

interference from the time spent on reading a question and optional answers, we excluded such reading time by providing the textual information before showing the data representation part of the stimulus.

We controlled a number of variables to minimize the potential confounding effects. These include:

- *Number of data dimensions* — We set this number to 4 (excluding the ID of each data point) throughout the study. This dimensionality is more challenging than the bivariate data featured in some previous studies (e.g., [LMv08]), while it is still feasible for participants to derive an answer in most cases.
- *Number of data points* — We set this number to 8 for all stimuli throughout the study. Our pilot study suggested that this number offers a balanced search time for the 4 tasks and 3 levels of task difficulty featured in the study.
- *Data consistency* — User performance may likely vary with different datasets. In order to compare DT, SP, and PCP fairly, we designed stimuli in a *trio* consisting of three different data representations (i.e., DT, SP, and PCP). Within each trio, all three representations adopted the same baseline dataset (with 4 variables and 8 data points), but each data value is slightly altered to prevent learning effects. Further minor variations were also introduced to inhibit learning effects.
- *Positions of optional answers* — The positions of the correct answer and its distractors may lead to biases in choosing options. In order to remove the confounding effects due to such biases, all stimuli in the same trio had the same positional order of the four optional answers.
- *Levels of distractor difficulty* — It is unavoidable to have different distractors. Such variations present potential confounding effects. We alleviated this by designing each set of four optional answers with a consistent scheme, i.e., each containing 1 *correct answer* and 3 *distractors* (1 *easy*, 1 *medium*, and 1 *hard*). For example, for the value retrieval task, the three levels reflect the value-distance between each distractor and the correct answer (i.e., *easy*: [40,80], *medium*: [15,35], *hard*: [5,10]).

3.1. Data Representations

Figure 1 shows three representations of an example trio. In (a), the data table shows a set of 8 multivariate data points $\mathbf{p}_i (i = 1..8)$, each with 4 attribute values, $\mathbf{p}_i = (v_{i,1}, v_{i,2}, v_{i,3}, v_{i,4})$. The data is accompanied by a heading row for attribute labels and an extra column for the IDs of data points. The rows of data points are colored differently to ease visual search. We adapted a qualitative set of colors suggested by ColorBrewer [Bre].

In (b), three SPs are used to depict a similar dataset as (a). Similar to a type of stimuli used in [KZZM12], one attribute variable (y-axis) is used to connect other three attributes. As mentioned earlier, we applied randomized minor variation to each value $v_{i,j}$ in the baseline dataset when constructing a stimulus for DT, SP, or PCP. The standard range of variation is $\{+0, +1, -1\}$. As values ending with 0 and 5 are easier to recognize and should ideally be avoided, all baseline datasets thus contained only values ending with 2, 3, 7, and 8. For each attribute in a dataset, all data values are unique.

To add further minor variations for preventing learning effects, we used different letter sets for attribute labeling in different stimuli within the same trio, and mapped each data point to one of the eight colors randomly. As the three stimuli in the same trio were placed at least 5 trials apart, it was reasonably certain that participants could not transfer the answers between stimuli in the same trio.

In (c), a PCP is used to depict a similar dataset as (a) and (b). Similarly, minor variations were applied to data values, attribute labeling, and color mapping.

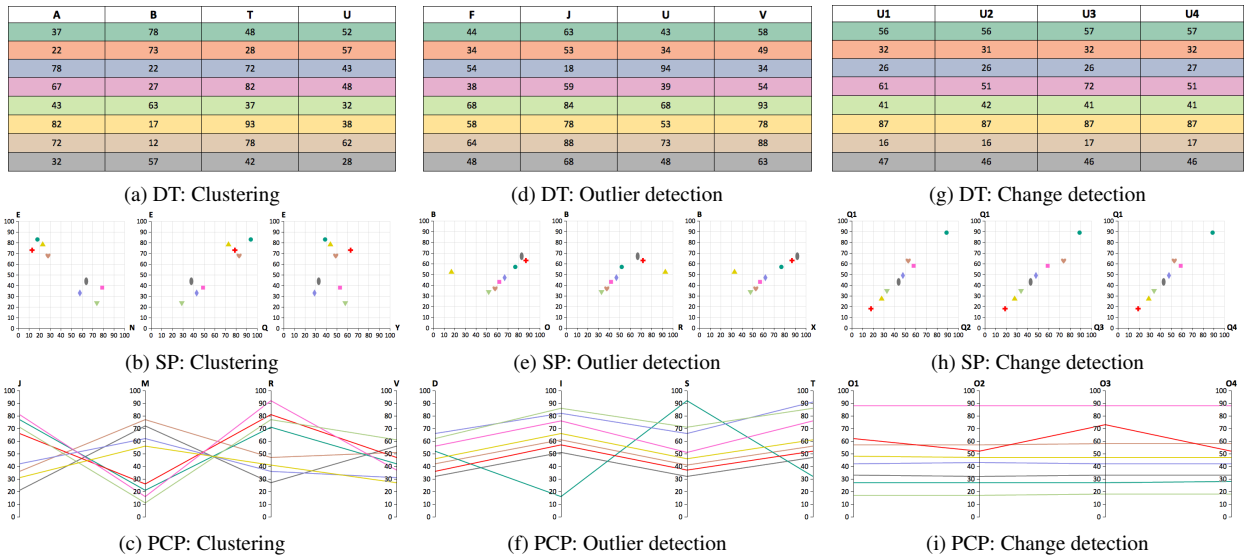
3.2. Visualization Tasks

We considered a variety of visualization tasks that may be performed with SPs and PCPs, and decided to choose four tasks that are detailed below. We restricted the number to four to avoid a lengthy experiment or any compromise of the minimal number of trials per task. We did not include the correlation task as it has been studied in detail in [LMv08], nor the classification task due to its similarity to clustering.

Value Retrieval: This task was the focus of [KZZM12], which involved only SP and PCP. We decided to reassess this task by comparing the performance of DT, SP, and PCP. This allowed us to correlate the results of our study with those of [KZZM12], with an additional comparison of SP and PCP with a non-visualization benchmark, i.e., DT.

For this task, as illustrated in Figure 1, participants are shown a set of 8 data points. Given one data value $v_{i,j}$ in a data point under a specific attribute j, participants were asked to find the corresponding data value $v_{i,k}$ for attribute k where $j \neq k$. For a DT, this involves a visual search for $v_{i,j}$ vertically along the column defined as attribute j, and then another visual search horizontally along row i to locate the cell i,k. For a row of SPs, this involves visual searches of the SPs featuring attributes j and k. If neither is mapped to the y-axis, a third attribute has to be used to aid the search. For a PCP, the search approach is expected to be similar to that with a DT, except that one has to trace a line from attribute j to attribute k rather than moving along a row of data values.

Clustering: Although [HvW10] studied clustering, they did not compare SP and PCP directly. Hence this was an important gap to be filled. For this task, participants were asked to choose an appropriate subset of data points to form a cluster. The participants were informed that the criterion of clustering was the similarity of data values. All participants had prior knowledge about observing clusters in a bivariate SP, so this criterion was easily understood. Nevertheless, apply-



(a) DT: Clustering (d) DT: Outlier detection (g) DT: Change detection

(b) SP: Clustering (e) SP: Outlier detection (h) SP: Change detection

(c) PCP: Clustering (f) PCP: Outlier detection (i) PCP: Change detection

Figure 2: *Example stimuli for three different visualization tasks: clustering (left), outlier detection (middle), and change detection (right). All stimuli are at the easy level in terms of task difficulty.*

ing this criterion to a multivariate dataset is more complex than observing a bivariate SP.

Three example stimuli for the clustering task are shown in Figure 2(a-c). Performing such a task with a DT is expected to be challenging. Performing the task with a row of three SPs involves a balanced judgment of different cluster patterns across three SPs. Meanwhile, with a PCP, the proximity of two polylines can be used to assess the likelihood whether the two corresponding data points belong to the same cluster. In addition, for this task, the optional answers showing both the IDs of the suggested data points and their corresponding visual objects, i.e., background row colors in a DT, colored shapes in a row of SPs, and colored lines in a PCP. The full screenshots of all stimuli can be found in [Kan14].

Outlier Detection: For this task, participants were asked to choose 1 or 2 data points as the outlier(s). Three example stimuli for the outlier detection task are shown in Figure 2(d-f). The strategy for performing this task is likely to be similar to that of clustering. As shown in Figure 2(e,f), an outlier can be considered as a data point that is numerically distant from other data points, or as a data point that does not correlate well with others. In some cases, as in the datasets in Figure 2(d-f), both factors co-existed, whereas in other cases, only one factor is apparent. Stimuli that feature different factors can be found in [Kan14].

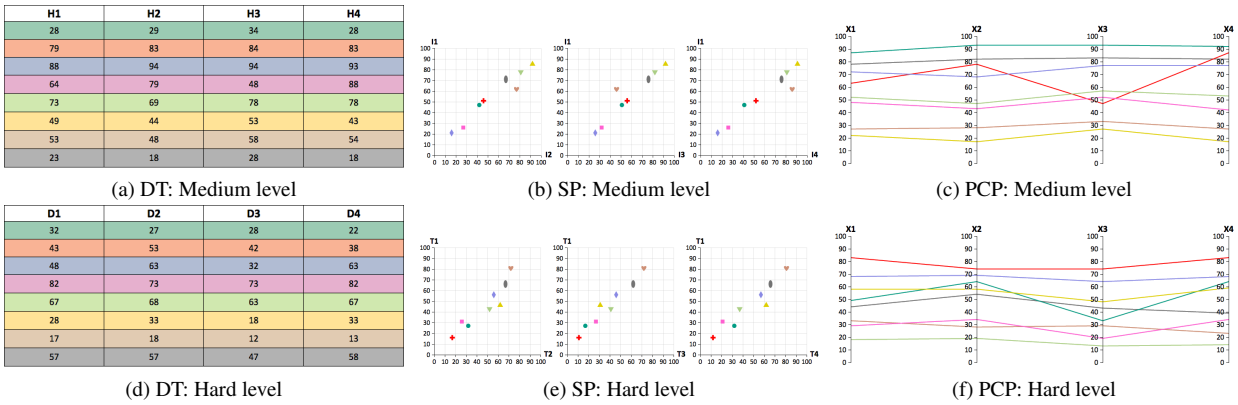
Change Detection: This task is commonly performed when different attributes variables exhibiting a temporal ordering. As shown in Figure 2(g-i), we explicitly labeled attributes using ordered labels, e.g., Q1, Q2, Q3, Q4. For this task, participants were asked to choose 1 or 2 data points that have the most changes. Note that an outlier is not necessarily a data point with the most changes. For example, one can observe that the top line in Figure 2(i) is an outlier, but does feature much variation from O1 to O4. Performing this task with a row of SPs may involve examining each data point across three SPs using its visual objects as the common identifiers. Performing this task with a PCP may involve examining each line to see how much changes occurred over the four axes.

3.3. Levels of Task Difficulty

All four visualization tasks feature three levels of task difficulty. It is not easy to control the levels of task difficulty in a precise manner. However, imposing some controls is necessary to reduce the confounding effects when the results of similar trials are grouped together in analysis. We thus defined three levels of difficulty for each task, denoting them as *easy*, *medium*, and *hard*.

For the **value retrieval** task, the three levels are defined by varying the selection of attribute variables and the numerical proximity of the distractors:

- *VR-easy* — The given attribute j is always assigned to the first column in a DT, the y-axis in a row of SPs, and the first axis in a PCP. Hence visual search is always from left to right in DT and PCP, and from y-axis to x-axis in SP.
- *VR-medium* — The target attribute k is always assigned to the first column in a DT, the y-axis in a row of SPs, and the first axis in a PCP. This requires backward searching from left to right in a DT or a PCP, and from one of the x-axes to the y-axis in a row of SPs.
- *VR-hard* — Neither attribute j or k involves the first column in a DT, the y-axis in a row of SPs, or the first axis in a PCP. In a row of SPs, this requires participants to re-



(a) DT: Medium level (b) SP: Medium level (c) PCP: Medium level

(d) DT: Hard level (e) SP: Hard level (f) PCP: Hard level

Figure 3: *Sample stimuli for the change detection task at the medium and hard levels of task difficulty.*

late the *x*-axis in one SP with that in another SP. Avoiding the "easily searchable attribute" may not have the same impact on DT or PCP as SP in term of task difficulty.

For the **clustering** task, the 8 data points are grouped into two subsets each with 4 data points. The correct answer for the clustering is determined using the *k*-means clustering algorithm. The distractors are obtained by swapping some data points in the two subsets, creating optional answers deviating from the result of the *k*-means clustering algorithm. The task difficulty level of distractors is based on the number of interchanged data points and their proximity. The three levels are defined as:

- *C-easy* — The two sets are separated by three attribute variables independently and obviously. It is thus relatively easy to observe the clustering pattern based on any of the three attributes and to reason about the uncertainty due to the 4th attribute.
- *C-medium* — The two sets are separated by two attribute variables independently and obviously. The other two attributes cast some uncertainty.
- *C-hard* — The two sets are separated by only one attribute variable independently and obviously. The other three attributes cast a fair amount of uncertainty.

For **outlier detection**, an outlier is defined by either (i) a different value range from other data points or (ii) a poor correlation with others. For example, the outlier in Figure 2(f) exhibits both factors. Each optional answer may contain 1 or 2 data points. The three levels are defined as:

- *OD-easy* — The outliers are based on both factors.
- *OD-medium* — The outliers exhibit only factor (i).
- *OD-hard* — The outliers exhibit only factor (ii).

For **change detection**, each data point may change its value from one attribute to the next following a temporal order. The contrast between different ranges of changes affects the perception of the most significant changes. Meanwhile the dynamics in changes creates distractive noise. The more dynamics, the harder to identify the most significant changes. Figure 3 shows example stimuli for change detection at the medium and hard levels of task difficulties (cf. Figure 2(h-i)). The three levels are defined as:

- *CD-easy* — The variation range of the data point(s) featured in the correct answer is 12, while that for the other data points is 2. This gives a high contrast and little noise.
- *CD-medium* — The variation range of the data point(s) featured in the correct answer is 16, while that for the other data points is 6. This gives a medium level of contrast and noise.
- *CD-hard* — The variation range of the data point(s) featured in the correct answer is 20, while that for the other data points is 10. This gives low contrast and high noise.

4. Empirical Study

The empirical study was conducted in four sessions, preceded by a pilot study. There were 2 participants (1 female, 1 male) in the pilot study, which allowed us to test various aspects of the experiment design, especially about timing. In this section, we will describe aspects of the main study, including its participants, apparatus, and procedure.

Participants. A total of 43 participants took part in the study in return for a £10 book voucher. Four sessions were organized with 8, 9, 14, and 12 participants respectively. Data for one participant was removed due to color-blindness. This leaves us with a total of 42 participants (20 females, 22 males). All participants were recruited from the University of Oxford. 27 participants were university students, and the other 15 were members of university staff. They were from a number of disciplines, including computer science, engineering, mathematics, statistics, zoology, medical sciences, business, and education. Among these participants, 26 were in the 20-29 age range, 8 in the 30-39 range, 5 in the 40-49 range, and 3 in the other three ranges respectively (19 or less, 50-59, and 60 or above). All 42 participants have normal or corrected to normal vision. Most participants were familiar with DT and SP technique, whereas about half of the participants were familiar with PCP.

Apparatus. The visual stimuli were generated using custom software written in Java and JavaScript. The stimuli were displayed to the participants through a custom-made software program, written in Java. The software was run in a full screen mode, using computers with 3.7 GB of RAM, 3.30 GHz quad-core Intel core i5-3550 processors, and running Fedora, a Linux based operating system, with GNOME version 3.4.2. Each computer had 24 inches Dell's LCD at 1920×1200 (16:10) resolution and with sRGB color display mode. We adjusted the monitors to the same levels of brightness and contrast. Each participant was required to interact with the software using a mouse at the desk.

Procedure. Prior to the experiment, the experimenter gave a 15-minute introduction, using self-paced presentation slides, to familiarize the participants with the study. The introduction included the explanation on the three visualization techniques, followed by the four visualization tasks. During the presentation, after each technique or each task had been presented, the participants were able to ask questions. This was designed to ensure that the participants understood each section of the presentation. The participants were also instructed to finish each trial as accurately and as quickly as possible, and were informed that each visualization image is independent to one another.

The experiment began immediately after the introduction. The time taken to complete an experiment was approximately 33 minutes ($min = 25$ and $max = 65$), excluding the introduction. The variation of time spent was due to the different amount of time used by the participants to read the instructions, perform each trial, and have a break.

The participants first completed a short form using the software program, providing their demographics information and familiarity rating about DT, SP, and PCP. The participants were then required to undertake a training session with 12 training trials. These were used to familiarize the participants with the four visualization tasks and the three data representations. To facilitate learning, the correct answer of each training trial was given as a feedback.

After the training session, the main body of the experiment started. The main body consisted of 72 trials, representing combinations of 3 data representations × 4 visualization tasks × 6 stimuli. Among the 6 stimuli, 2 are at the easy level, 2 medium, and 2 hard. The 72 trials were organized into 24 sections each with 3 trials using a pseudo-randomization mechanism abiding by the following rules:

- As mentioned in Section 3, the stimuli in the same trio feature similar datasets. So they cannot be placed in the same section, and they must be separated by at least 5 other trials. This prevents learning effects within each trio.
- In each section, there must be 1 DT, 1 SP, and 1 PCP trial, and they are placed in a random order. This ensures stimuli of each data representation are evenly distributed across all sections.
- In each section, there must be 1 easy, 1 medium, and 1 hard trial. This ensures stimuli at each level are evenly distributed across all sections. An "easy to hard" ordering eases the switch between sections for different tasks.
- In each section, all three trials must be for the same visualization tasks. Before each section, an instruction screen is shown to indicate the same task for the next three trials. This prevents potential confusion about different tasks. After each section, the participants were allowed to take a short break before continuing to the next one.
- The 24 sections are ordered in a random fashion.

Each trial has a time limit of 90 seconds. A progress bar serves as the timer. If a participant cannot select an answer in time, the software records *no answer* for that trial, and the participant is asked to move on to the next trial.

When all 72 trials had been completed, the participants were required to rate the effectiveness of each data representation in relation to each visualization task. The effectiveness rating used a five-level Likert scale: *not at all effective*, *slightly effective*, *moderately effective*, *very effective*, and *extremely effective*.

5. Results and Analysis

Figure 4 depicts the main descriptive statistics for the two objective measures, namely accuracy (Ac) and response time (RT), and one subjective measure of effective rating. There are three accuracy categories: *correct*, *incorrect*, and *no answer*. Following the convention in cognitive sciences, the response time is calculated based only on correct answers. From Figure 4(a,b), we can observe the following trends in relation to the four hypotheses in Section 1:

R1: (Ac) DT $\succsim$ PCP $\succsim$ SP; (RT) DT $\succ$ PCP $\succ$ SP
R2: (Ac) PCP $\succ$ SP $\succ$ DT; (RT) SP $\succsim$ PCP $\succ$ DT
R3: (Ac) PCP $\succ$ SP $\succ$ DT; (RT) PCP $\succsim$ SP $\succ$ DT
R4: (Ac) PCP $\succ$ SP $\succsim$ DT; (RT) PCP $\succ$ SP $\succ$ DT

where $\succ$ denotes that the data representation on the left would result better performance than that on the right, while $\succsim$ denotes the same order with some uncertainty.

The subjective measures of effective rating in Figure 4(c) are largely consistent with the objective measures. When the two objective measures deviate, the subjective measure is close to response time for clustering and outlier detection. We did not find any positive correlation between familiarity of SP vs. PCP and performance of SP vs. PCP.

We use ANOVA to derive inferential statistics as shown in Table 1. If *Mauchly's Test of Sphericity* shows the assumption of sphericity is violated, we use *Huynh-Feldt* or *Greenhouse-Geisser Corrections* according to [Atk11]. Figures 5–8 show the comparative results at each level of task difficulty for the four visualization tasks respectively. The detailed results of ANOVA analysis at each level of task difficulty can be found in [Kan14]. In summary, we can observe the followings:

Table 1: *The ANOVA results to accompany Figure 4(a,b).*

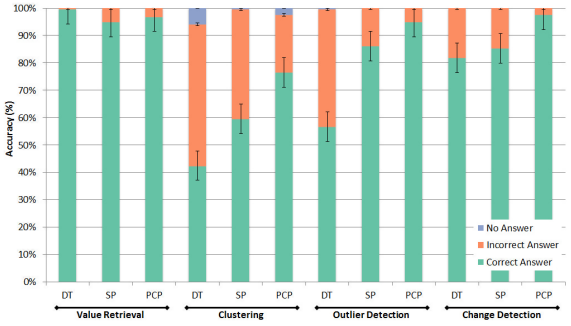
Vis. Task	Main Effect	DT v SP	DT v PCP	SP v PCP
Value Retr. (Ac)	=.037	=.037	=.153	=1
Value Retr. (RT)	<.001	<.001	<.001	<.001
Clustering (Ac)	<.001	<.001	<.001	<.001
Clustering (RT)	<.001	<.001	<.001	=1
Outlier Det. (Ac)	<.001	<.001	<.001	=.018
Outlier Det. (RT)	<.001	<.001	<.001	=.002
Change Det. (Ac)	<.001	=.423	<.001	<.001
Change Det. (RT)	<.001	<.001	<.001	<.001

- For **value retrieval**, DT resulted in the fastest response time, and SP was the slowest. The trend of accuracy is in favor of DT, though it is statistically insignificant. This suggests that the two visual representations, SP and PCP, do not exhibit much strength in supporting this task.
- For the **clustering** task, both SP and PCP are significantly better than DT. At the *easy* and *medium* levels, SP and PCP performed similarly. At the *hard* level, PCP performs better than SP in accuracy, but has no significant advantage in response time.
- For **outlier detection**, both SP and PCP are significantly better than DT. At the *easy* and *medium* levels, SP and PCP performed similarly. At the *hard* level, PCP performs better than SP in accuracy ($p = .035$) and response time ($p = .05$), but with some statistical uncertainty.
- For **change detection**, at the *easy* level, DT yields lower accuracy than SP ($p = .037$) and PCP ($p = .02$). PCP is faster than DT ($p < .001$) and SP ($p < .001$). At the *medium* and *hard* levels, PCP are better than DT and SP in both accuracy and response time (all $p < .001$ except in one case $p = .008$).

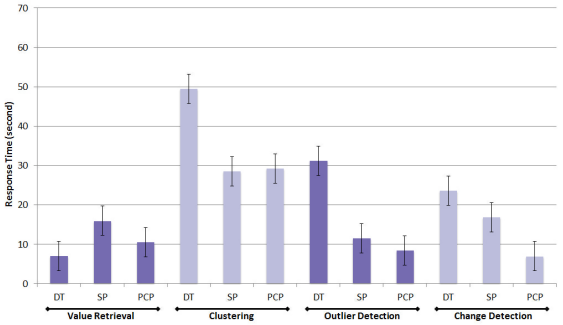
6. Conclusions

To the best of our knowledge, this is the first multi-task empirical study for comparing scatter plot (SP) and parallel coordinates plot (PCP) for visualizing multivariate data. Perhaps more significantly, it is also the first empirical study that has provided conclusive evidence in favor of PCP. While the findings of this study do not in any way dispute the findings of previous studies such as [HHB07, LMv08, HvW10, KZZM12], collectively they indicate that for some visualization tasks (i.e., clustering, outlier detection and change detection), PCP has the relative merits, and for some other tasks (i.e., correlation perception and similarity detection), SP is advantageous. Against the backdrop of the less favorable findings about PCP in the previous studies, it is also a great comfort to learn that the huge amount of enthusiasm about, and research effort made for, PCP in the field visualization has been totally justifiable.

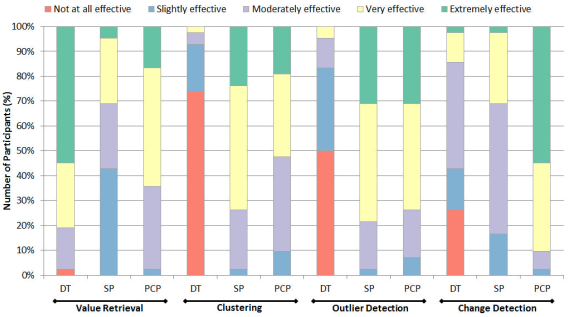
In addition, this study has shown that apart from the value retrieval task, both SP and PCT have outperformed data table (DT). This may suggest that the process of visualization is usually for supporting several tasks concurrently, and perhaps, the value retrieval task is not as important as some other tasks. This work also highlights the usefulness in considering textural data representations when comparing visual representations. The finding that PCP and SP are not advantageous over DT in the value retrieval task offers a new perspective to the results of [KZZM12]. Of course, this suggestion should be part of the continuing discourse in the field of visualization. It is desirable to study other variables, such as the numbers of data dimensions and data points (i.e., scalability) in future work. Since PCP has advantages over SP in several visualization tasks, we should enthusiastically encourage the use of PCP for multivariate data visualization, while devising more effective means for improving visualization literacy about PCP.



(a) Average accuracy for each visualization task

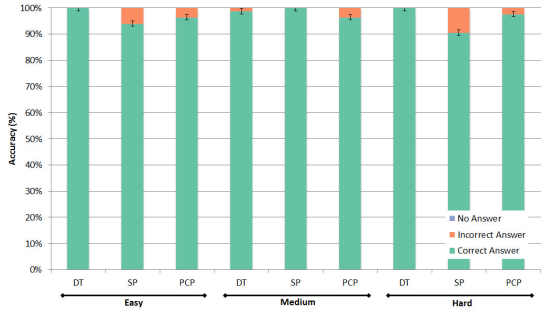


(b) Average response time for each visualization task

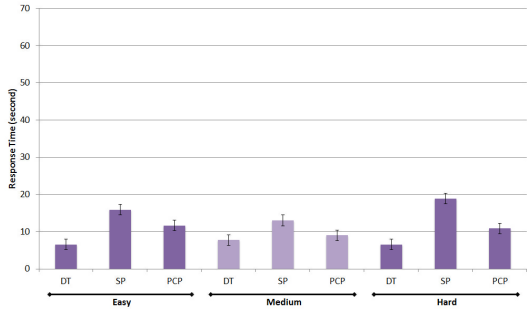


(c) Effectiveness rating for each visualization task

Figure 4: *Summary of the performance results of each data representation in conjunction with each visualization task.*

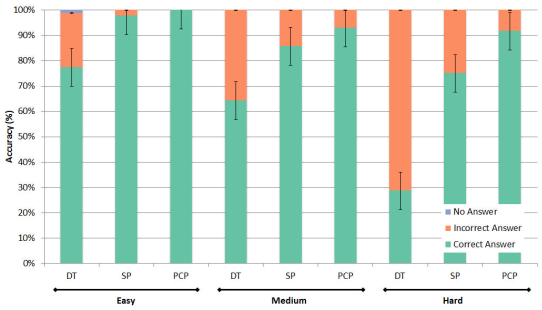


(a) Average accuracy

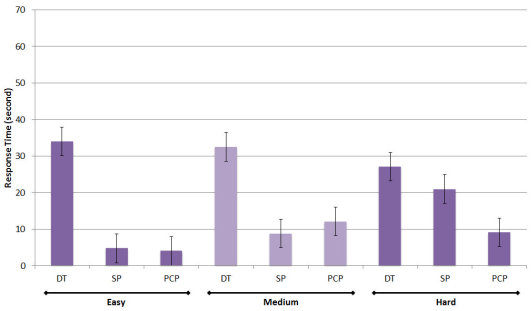


(b) Average response time

Figure 5: *Performance of the value retrieval task at different levels of task difficulty.*

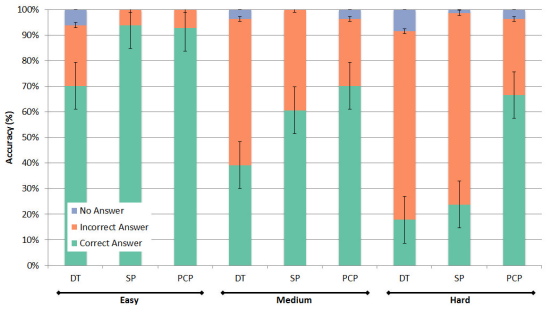


(a) Average accuracy

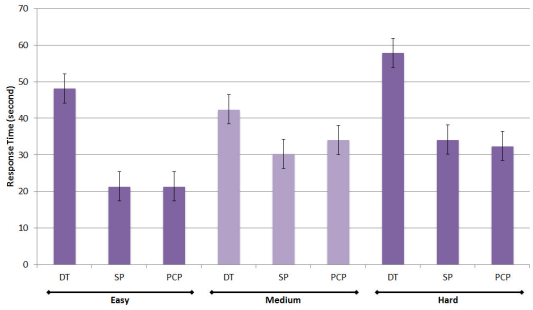


(b) Average response time

Figure 7: *Performance of the outlier detection task at different level of task difficulty.*

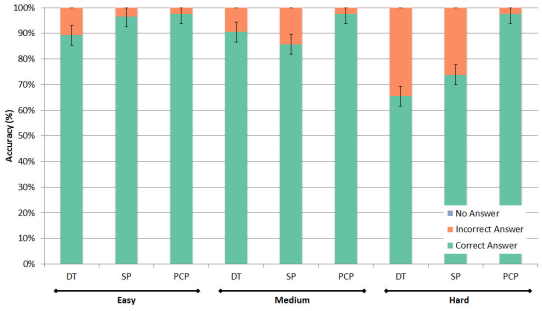


(a) Average accuracy

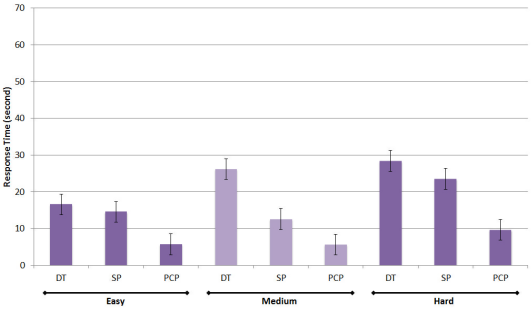


(b) Average response time

Figure 6: *Performance of the clustering task at different levels of task difficulty.*



(a) Average accuracy



(b) Average response time

Figure 8: *Performance of the change detection task at different levels of task difficulty.*

References

[AR11] AZHAR S. B., RISSANEN M. J.: Evaluation of parallel coordinates for interactive alarm filtering. In *Proc. Information Visualisation (IV)* (2011), pp. 102–109. 1, 2

[Atk11] ATKINSON G.: Analysis of repeated measurements in physical therapy research. *Physical Therapy in Sport 2*, 4 (2011), 194–208. 7

[Bre] BREWER C. A.: Colorbrewer 2.0: Color advice for cartography. http://colorbrewer2.org/. [Accessed on 30 May 2014]. 4

[Cle85] CLEVELAND W. S.: *The Elements of Graphing Data*. Wadsworth Advanced Books, 1985. 2

[CM84] CLEVELAND W. S., MCGILL R.: The many faces of a scatterplot. *Journal of the American Statistical Association 79* (1984), 807–822. 2

[FD05] FRIENDLY M., DENIS D.: The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences 41*, 2 (2005), 103–130. 2

[GH83] GANNETT H., HEWES F. W.: General summary, showing the rank of states by ratios, 1880. In *David Rumsey Historical Map Collection*. Charles Scribner's Sons, 1883. 2

[HG01] HOFFMAN P. E., GRINSTEIN G. G.: A survey of visualizations for high-dimensional data mining. In *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann Publishers, 2001, pp. 47–82. 2

[HHB07] HENLEY M., HAGEN M., BERGERON R. D.: Evaluating two visualization techniques for genome comparison. In *Proc. Information Visualization (IV)* (2007), pp. 551–558. 1, 2, 8

[HvW10] HOLTEN D., VAN WIJK J. J.: Evaluation of cluster identification performance for different PCP variants. *Computer Graphics Forum 29*, 3 (2010), 793–802. 1, 2, 4, 8

[HW13] HEINRICH J., WEISKOPF D.: State of the art of parallel coordinates. In *Proc. Eurographics State of the Art Reports* (2013), pp. 95–116. 1, 2

[Ins85] INSELBERG A.: The plane with parallel coordinates. *The Visual Computer 1*, 2 (1985), 69–91. 1, 2

[Ins09] INSELBERG A.: *Parallel Coordinates: Visual Multidimensional Geometry and its Applications*. Springer, 2009. 2

[JFLC08] JOHANSSON J., FORSELL C., LIND M., COOPER M.: Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships. *Information Visualization 7*, 2 (2008), 152–162. 2

[JSD11] J. H., S. B., D. W.: Progressive splatting of continuous scatterplots and parallel coordinates. *Computer Graphics Forum 30*, 3 (2011), 653–662. 2

[Kan14] KANJANBOSE R.: *An Empirical Study on Parallel Coordinates and Scatter Plots*. Master's thesis, Department of Computer Science, University of Oxford, September 2014. URL: http://www.ovii.org/sp-pcp/. 5, 7

[KD09] KINCAID R., DEJGAARD K.: Massvis: Visual analysis of protein complexes using mass spectrometry. In *Proc. IEEE Symposium on Visual Analytics Science and Technology* (2009), pp. 163–170. 2

[KZZM12] KUANG X., ZHANG H., ZHAO S., MCGUFFIN M. J.: Tracing tuples across dimensions: A comparison of scatterplots and parallel coordinate plots. *Computer Graphics Forum 31*, 3 (2012), 1365–1374. 1, 2, 4, 8

[LMv08] LI J., MARTENS J.-B., VAN WIJK J. J.: Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization 9*, 1 (2008), 1–18. 1, 2, 4, 8

[Pla04] PLAISANT C.: The challenge of information visualization evaluation. In *Proc. Working Conference on Advanced Visual Interfaces* (2004), ACM, pp. 109–116. 1

[SLHR09] SIIRTOLA H., LAIVO T., HEIMONEN T., RAIHA K. J.: Visual perception of parallel coordinate visualizations. In *Proc. Information Visualisation (IV)* (2009), pp. 3–9. 1, 2

[SR06] SIIRTOLA H., RAIHA K.: Interacting with parallel coordinates. *Interacting with Computers 18*, 16 (2006), 1278–1309. 2

[TFA*11] TAM G. K. L., FANG H., AUBREY A. J., GRANT P. W., ROSIN P. L., MARSHALL D., CHEN M.: Visualization of time-series data in parameter space for understanding facial dynamics. *Computer Graphics Forum 30*, 3 (2011), 901–910. 2

[TGS04] TYMAN J., GRUETZMACHER G., STASKO J.: InfoVis-Explorer. In *Proc. IEEE Information Visualization* (2004). 2

[tMR07] TEN CAAT M., MAURITS N. M., ROERDINK J. B.: Design and evaluation of tiled parallel coordinate visualization of multichannel EEG data. *IEEE Transactions on Visualization and Computer Graphics 13*, 1 (2007), 70–79. 1, 2

[VMCJ10] VIAU C., MCGUFFIN M. J., CHIRICOTA Y., JURISICA I.: The FlowVizMenu and parallel scatterplot matrix: hybrid multidimensional visualizations for network exploration. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1100–1108. 2

[WB97] WONG P. C., BERGERON R. D.: 30 years of multidimensional multivariate visualization. In *Scientific Visualization Overviews, Methodologies, and Techniques*. IEEE Computer Society Press, 1997, pp. 3–33. 2

[Weg90] WEGMAN E. J.: Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association 85*, 411 (1990), 664–675. 1, 2

[YGX*09] YUAN X., GUO P., XIAO H., ZHOU H., QU H.: Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1001–8. 2