PARALLEL PLOT VISUALIZATION FOR NOVICE USERS

Raja Jamilah Raja Yusof, Nazean Jomhari, Roziati Zainuddin, Mohd Sapiyan Baba
Faculty of Science Computer and Information Technology
Universitiy Malaya, 50603 Kuala Lumpur

Zulkifliy Mohd. Yusoff
Academic of Islamic Studies
Universitiy Malaya, 50603 Kuala Lumpur

Abstract— Parallel Plot for novice users is discussed. A relational model of parallel coordinate was used for analysis based on an interface called PaCQ. The advantage and disadvantages of parallel plot is discussed. Although through literatures it was found that parallel plot does not score well in graphical perceptibility among users, results shows that novice users can interpret parallel plots relatively well. The GOMS analysis of tasks related to parallel plot also shows that the time taken to complete tasks related to the parallel plot is relatively reasonable, well below 10s. The user study among novice with three different levels of education shows no significant difference in the scores of interpreting the parallel plot.

Index Terms— Information Visualization, Parallel plot, Multidimensional, GOMS analysis;.

1 INTRODUCTION

RESEARCH in visualization sometimes addresses the issue of expert and novice users. The term novice is sometimes used in an unclear manner. According to Fisher [1] novice users are those without experience in conducting certain tasks. Although given certain amount of period experience can be gained by novice user and expertise can be developed in the task involved. There is also the term naive users, those whom may be lacking analytical and critical skills. Particularly using a system without having to know how the system functions. Novice and naive users may coincide. Therefore in the context of this paper, the term novice and naïve will be refered as novice.

The usage of parallel coordinate visualization technique is usually associated with experts. However, very little research had been conducted on the perceptibility and performance of novice users using parallel coordinate. The perpendicular plots are much more popular to the generally especially to novice users since this type of plot had been introduced in high school level. In this paper, the advantages and disadvantages of the parallel plot will be discussed. A relational mathematical model of the parallel plot will be outlined based on an interface developed for novice user called PaCQ. Then the GOMS analysis is included based on a relational mathematical model. Finally, a user study on the performance of the novice users interpreting the parallel plot is discussed.

2 PARALLEL PLOT

2.1 Overview

A point can be represented in the perpendicular plot in a Cartesian system as$p(x_1, y_1)$, line l and distance d represented as the following equations:

$$l\text{: } y = mx + b \text{ – Equation 1}$$

$$d = | mx_1 - y_1 + \text{b} | / \sqrt{1 + m^2}\ . \qquad \text{- Equation 2}$$

A simple representation of this concept can be illustrated, the Cartesian system plane is represented as $\mathbb{R}^N$, where N=2 since it is 2-dimensional and the parallel plane $\mathbb{P}^2$ represents the $\mathbb{R}^2$ Cartesian system plane (Inselberg, 2009). A point in the in the $\mathbb{R}^2$ plane, p_r can be represented in the $\mathbb{P}^2$ plane as a line, ℓ_p. This is referred to as the principle of duality. Figure 1 (a) and (b) illustrate an example, point (3,0) in $\mathbb{R}^2$ plane is represented as a line connecting the corresponding points on the x and y axis drawn in parallel to each other. The position of the axis can be interchanged as shown in 1 (c).

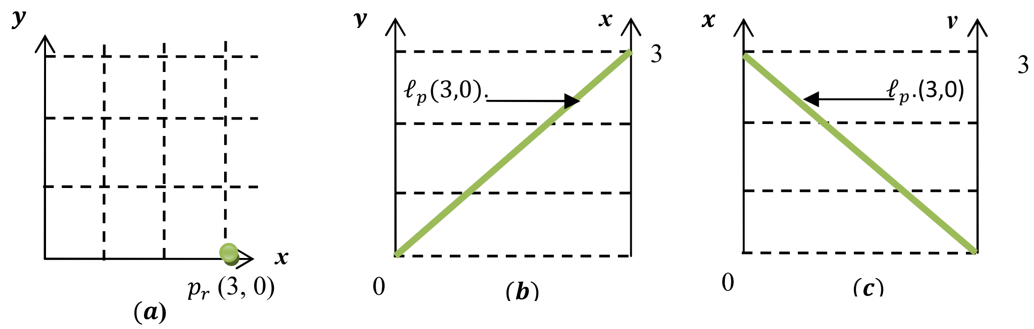

Figure 1: (a) a point (3,0) in the Cartesian $\mathbb{R}^2$ plane (b) point (3,0) on the line $\boldsymbol{\ell_r}$ in the parallel ($\mathbb{P}^2$) plane (c) interchangable axis position

Therefore, a line in the $\mathbb{R}^2$ plane, ℓ_r can be represented in the $\mathbb{P}^2$ plane as many lines, $\ell_{p(i-1),i}$ where i correspond to a point in the line,

$$\boldsymbol{\ell_r}: \begin{cases} \boldsymbol{\ell_{p\,1,2} : x_2 = m_2 x_1 + b_2} \\ \boldsymbol{\ell_{p\,2,3} : x_3 = m_3 x_2 + b_3} \\ \dots \\ \boldsymbol{\ell_{p\,N-1,N} : x_N = m_N x_{N-1} + b_N} \end{cases} \quad \textbf{Equation 0}$$

where m and b are values related to the slope and the point on the parallel lines. Figure 2 illustrate an example of the concept for a negative slope. Taking this further as a general point, indicated that the point of inter§ can be represented by

$$\bar{\boldsymbol{l}} : \left(\frac{\bar{\boldsymbol{d}}}{\boldsymbol{1-m}}, \frac{\boldsymbol{b}}{\boldsymbol{1-m}}\right), \boldsymbol{m \neq 1} \ \textbf{- Equation 3}$$

where $\bar{d}$ is the distance between the two adjacent parallel axis and m is the slope of the line in the $\mathbb{R}^2$ plane (Inselberg, 2009). Figure 3 illustrates example of a positive

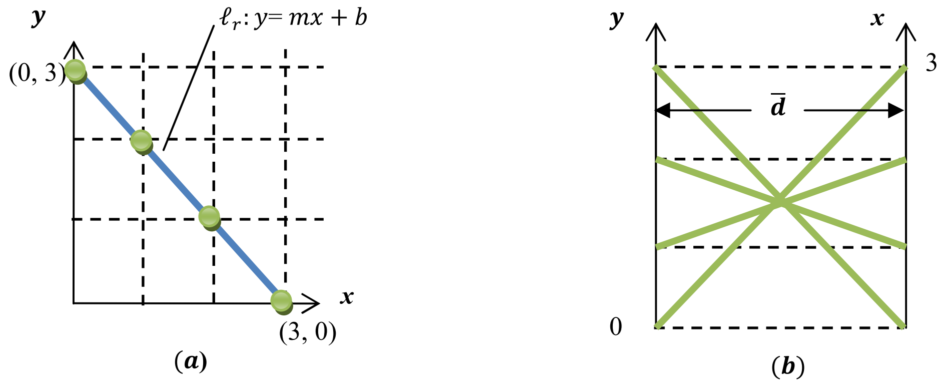

Figure 2: (a) negative slope line $\boldsymbol{\ell_r}$ in the Cartesian $\mathbb{R}^2$ plane (b) line $\boldsymbol{\ell_r}$ in the parallel ($\mathbb{P}^2$) plane

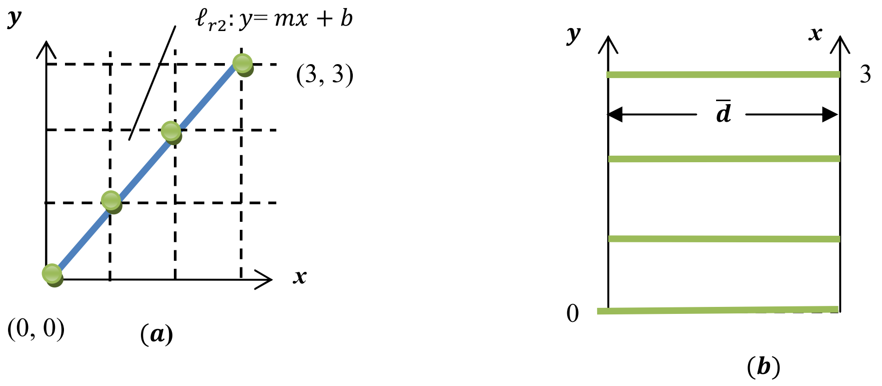

Figure 3: (a) positive slope line ℓ_r in the Cartesian $\mathbb{R}^2$ plane (b) line ℓ_{r2} in the parallel ($\mathbb{P}^2$) plane

When a third dimension is added in the $\mathbb{R}^2$ plane, a third line is added in the $\mathbb{P}^2$ plane as illustrated in Figure 4 (a) and (b).

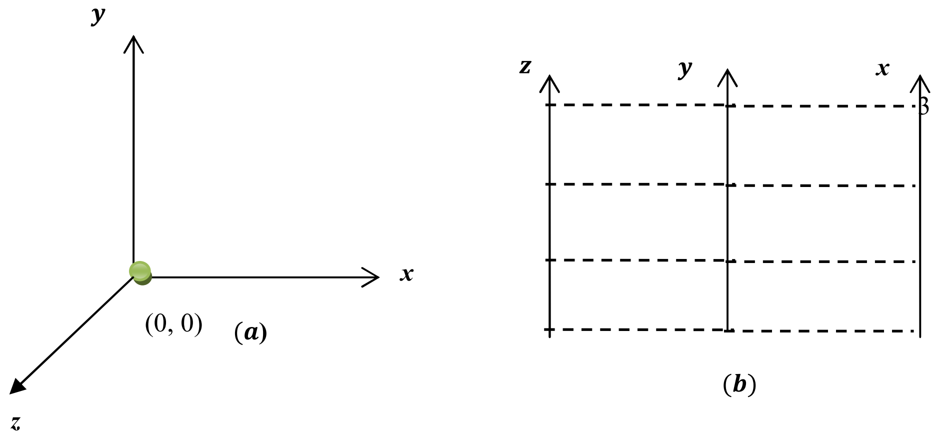

Figure 4: (a) 3 dimensional lines in the Cartesian $\mathbb{R}^2$ plane (b) 3 parallel lines represented in the parallel ($\mathbb{P}^2$) plane

Emphasis of the discussion on this § is therefore on the comparison between parallel coordinate (one type of geometry projection visualization technique) and pixel based plots under the family of scatter plot and line charts.

2.2 Advantages and Disadvantages

The main disadvantage of parallel plots is that the display is considered cluttered and hard to interpret if there are too many dimensions and set of attributes to be shown but this would depend on the size of the monitor displaying the data [5, 8, 11, 12]. This problem is related to the occlusion problem which had been mentioned earlier in perpendicular section. Ref. [8] and [13] pointed out that interpretation of parallel coordinates is harder and requires expert knowledge although there is also evidence showing that parallel plots are not difficult to manipulate and can result in a more accurate interpretation of data even by novice users [12-14]. Also, parallel coordinate is not suitable for visualizing large data sets which are more than the range of hundreds to thousands and the relationships between attributes can only be easily identified for two adjacent axes [10, 15]. Rearranging of the axes can overcome the mentioned problem.

The advantage of these plots is that it is considered as more effectively consuming the display space therefore increasing the amount of data a human can work with at the same time [16]. However, [10] found that 11 axes is the maximum threshold for a human to efficiently analyze the data. Also, N-dimensional properties can be visible from the screen [2].

2.3 Applications and tools

The operations related to parallel plots are very similar to the scatter plots i.e. zooming, panning, filtering and selection, detail-on-demand, focus+context, linking+brushing [6, 8, 17] and 3D manipulation. There is also operation such as rearranging the axes in order to permit analysis between two adjacent axes [8, 11]. Real application of parallel plots are also vast in number, in network attacks data [18], land satellite data [19], climate analysis [20], data envelopment analysis [21], fuel injection simulation [17], and others. The related tools for parallel plots usually include other types of visualization as well and these are such as GGobi [22], ComVis [17].

3 Relational model of parallel coordinate

A relational model can be associated with database systems. The parallel coordinate can be used to represent attributes of a table in a database system. This approach can be considered as a simpler approach in explaining the concept to a novice user. In the parallel plot however, the parallel lines will increase to *N-1* where *N* is the maximum number of variables. Let p_r be a point that is representing a binary relationship between two variables v_{i-1} and v_i in the Cartesian plane. If there are N variables, the set of binary relationship R that can be plotted are denoted by $R=\{R_{v1v2}, R_{v2v3}, R_{v1v3} \ldots, R_{(1-(N(N-1)/2)(N(N-1)/2)}\}$ where there contain N (N-1)/2 relationships. The corresponding plot in the parallel plane contains N axes line relation plotted in parallel denoted by $P=\{P_{v1}, P_{v2}, P_{v3} \ldots, P_N\}$ where the lines $l_{pv(i-1)v}$ are the points representations in the parallel plot. For each relation in R, there is a set of points with value v_{i-1} and v_i, $R_{vi(vi-1)}=\{(v_{i-1}, v_i), (v_i, v_{i+1}), (v_{i+1}, v_{i+2}), \ldots (v_{k-1}, v_k)\}$ where k is the maximum number of data set and v and i are arbitrary numbers.

3.1 The PaCQ interface and the relational model

The PaCQ interface was developed for supporting the Muslims of non-Arabic speakers to enhance their comprehension of an Arabic document called the Qur'an using the theory of word recognition. The problem faced by the non-Arabic speakers is not being able to comprehend what they read even though they are able to read the Arabic document. The reason is purely spiritual since the Qur'an is a source of guidance and is believed to be the word of Allah, God to the Muslims. PaCQ interface is targeted to support the Arabic word recognition process and therefore enhancing reading comprehension.

The PaCQ interface lets the user go through the Arabic document chapter by chapter which is called *sura*. Also, the user can browse the document sentence by sentence which is called *ayah*. The interface was built based on only 1/30 of the whole document and this is the 30th part or the last part of the document which is also named as *Juz Amma*. In browsing through the document, the user can add up their vocabulary to the list that will be stored in the system. The PaCQ interface lets the user see the overall percentage of their vocabulary in comparison to each *sura*. Also the PaCQ interface gives a comparison chart if all vocabulary is known in one *sura* and what percentage of these vocabulary words compared to other *suras*.

In this study, there are several variables sets:

- The Arabic word, A, where $A = \{A1 \cap A2\}$
  - $A1 \in$ *all Arabic words in Sura*1
  - $A2 \in$ *all Arabic words in Juz Amma*
- The Vocabulary words, V, where $V \subseteq A$
  - Note: the vocabulary for a user in the scope of the study is limited to only the Arabic words in *Juz Amma*
- The word count, C where $C = \{ c \in \mathbb{N}^+ \}$ and $\mathbb{N}^+$ is the set of positive natural numbers
- The corresponding *Sura* number, S where $S = \{ s \in \mathbb{N}^+ : 78 \leq s \leq 114\}$
  - Note: the *Sura* that contains in *Juz Amma* is from the *Sura* number 78 to 114
- The corresponding *Ayah* number, Y where $Y = \{ y \in \mathbb{N}^+ : y \leq 46\}$
  - Note: the maximum number of *Ayah* / verse in *Juz Amma* is 46
- The percentage, T where $T = \{ t \in \mathbb{N} : t \leq 100\}$ where $\mathbb{N}$ is the set if natural numbers

The main relations between these variables can be written as

- $R_1 = \{(c, a, s, y)$: There are c counts of Arabic word a in *Juz Amma* and can be found in *Sura* s and *Ayah* y } where $c \in C$, $a \in A$, $s \in S$ and $y \in Y$.
- $R_2 = \{(v, s, t)$: There is T percentage of the vocabulary words V in *Sura* $S\}$ where $v \in V$, $s \in S$, $t \in T$.
- $R_3 = \{ (c, s_1, s, t)$: There is T percentages of Arabic words in *Sura* S_1 compared to other *Suras* S in *Juz Amma* and C count of Arabic words in the *Sura* $S\}$ where $\{s_1$: P(s_1) } and P(s_1) is exactly one *Sura* in *Juz Amma* and $s \in S$, $t \in T$.

4 THE GOMS ANALYSIS

In this section GOMS [23] analysis will be used to predict the time taken for a user to interpret information for the parallel plots. GOMS is one of task analysis methods contain a set of techniques on how to model task in terms of Goals, Operators, Methods and Selection Rules while using a graphical user interface. As all task analysis methods starts with analyzing the goal of the task, GOMS also starts with a goal of a task and decompose these tasks to smaller tasks up to the key-stroke level (for example mouse click, pressing a key on the keyboard, finding a target object on the screen, pointing to an object and others). Furthermore, the technique gives an estimated time to accomplish these key-stroke level tasks known as operators. Table 1 shows the tasks involved in relation to the three relations identified for the PaCQ interface.

TABLE 1: Relations and the corresponding tasks.

Rel.	Scenario/ Goal	Tasks Involved
R_1	What is the word with the highest count and where can this word be found?	R1G: Find the word with the highest count and where can this word can be found
		S1:Find word
		S1.1:Locate/identify count axis/ symbol
		S1.2: Locate the highest count
		S1.3: Locate the corresponding word
		S2: Find where this word can be found
		S2.1: Locate the corresponding *surah no.*
		S2.1.1: Locate/identify *surah* axis/symbol
		S2.1.2: Read the corresponding *surah* no.
		S2.2: Locate the corresponding *ayat no.*
		S2.2.1: Locate/identify *ayat* axis/symbol
		S2.2.2: Read the corresponding *ayat* no.
R_2	What is the surah with the highest percentage based on the vocabu-lary words in the list?	R2G: Find the *surah* with the percentage of the vocabulary words in the list.
		S1: Find highest percentage
		S1.1: Locate/identify percentage axis/symbol
		S1.2: Locate/identify the highest percentage value
		S2: Find the corresponding *surah*
		S2.1: Locate the *surah no.*
R_3	What is the highest percentage of words in one *surah* compared to another and what are the *surahs*?	R3G: Find the highest percentage of words in one *surah* compare to another and the corresponding *surahs*.
		S1: Find the highest percentage of words
		S1.1: Locate/identify percentage axis/symbol
		S1.2: Locate the highest percentage value
		S2: Find the corresponding *surahs*
		S2.1: Identify the base *surah*
		S2.1.1: Locate/identify the base *surah* axis/symbol
		S2.1.2: Locate/identify the base *surah* no.
		S2.2: Identify the compared *surah*
		S2.2.1: Locate/identify the compared *surah* axis/symbol

4.1 GOMS analysis for relation R_1

In order to carry out goal for relation R_1, R1G, the lowest level tasks involve are identify count symbol, locate the highest count, locate the corresponding word, identify *surah* symbol, read the corresponding *surah* no., identify *ayat* symbol, read the corresponding *ayat* no. In a parallel plot these tasks may involve other subtasks:

- The user need to locate the count axis -1
- Locate the highest count; the user will also be confronting a cluttered display. However, there will not be much problem in comparing to the next highest point since the points are display vertically instead of horizontally in the perpendicular view. Therefore, the subtasks involved may be as the following:
  - Select/located the area with the highest point -2
  - Zoom into the area – 3
  - Read the highest count value -4

- Locate the corresponding word; at this point the words in the parallel plot will also be overlapping each other. Therefore, the subtask involve may be the following:
  - Click the highest count value -5
- This action can be connected to a filtering function. Once this value is clicked the system can automatically display the corresponding related attribute which is the word, the *surah* no. and the *ayat* no.. All other related values can be faded out or temporarily filtered off the display screen consequently simplifying all subtasks that comes after this.
  - View the display word on the word axis
    - Locate the word axis -6
    - Read the word – 7
- Locate *surah* / *ayat* axis - 8
- Read the corresponding *surah*/ *ayat* no.. At this point in the parallel plot there will also be many possible *surah* and *ayat* related to the highest word count but it is possible to read all the *surah* and *ayat* no.. -9

4.2 GOMS analysis for relation R_2

In order to carry out goal for relation R_2, R2G, the lowest level tasks involved are identify percentage axis/symbol, locate the highest percentage value, locate the *surah* no..

- Locate percentage symbol – 1
- Locate the highest percentage value- 2
- Locate the surah no. -3

4.3 GOMS analysis for relation R_3

In order to carry out goal for relation R_3, R3G, the lowest level tasks involve are identify percentage symbol, locate the highest percentage value, identify the base *surah* symbol, identify the base *surah* no., identify the compared *surah* symbol, identify the compared *surah* no.. In the parallel plot, the subtasks involve are as the following:

- Locate the percentage axis -1
- Locate the highest percentage value: in this case, since the scale of percentage from 0-100 is at a detectable value, the user can follow the sequence of task as in relation to locate the highest count, which will involve 3 subtask – 2-4
- At this point the user can use a filter function such as clicking on the highest percentage value that will display the corresponding attributes (the base *and* compared *surah*). This action will automatically reduce the subtasks that comes next after this one - 5
  - Locate the base *surah* axis -6
  - Locate the base *surah* no. -7
  - Locate the compared *surah* axis -8
  - Locate the compared *surah* no. – 9

4.4 GOMS analysis estimated time for tasks related to relation R_1, R_2 and R_3

TABLE 2: GOMS estimated operation duration

Operation	Duration(sec)
Click-mouse-button	0.20
Shift-click-mousebutton	0.48
Cursor movement	1.10
Mental Preparation	1.35
Determine Position	1.20
	0.28

Source [32]

Table 2 shows the GOMS estimated operation duration, from this table an estimated time for relation R_1, R_2 and R_3 is calculated and Table 3 shows this time.

It can be seen that the time taken to accomplish the tasks using parallel coordinate plots is 8.8s, 3.6s and 8.6s for R_1, R_2 and R_3 respectively. These timing can be considered reasonable.

5 THE USER STUDY

A user study was conducted to evaluate the performance of novice users interpreting the parallel coordinate graph in the PaCQ interface. This section describes the conducted study and the results found.

5.1 The Participants, Apparatus, Design and Procedure

Participants were chosen from around the Faculty of Science Computer and Information Technology of University of Malaya. Fifteen participants were involved, with 3 main categories according to their education level. All of them can be considered as novice users since they do not have experience in interpreting the parallel plot. Five participants in each category:

- High school academic qualification faculty staffs
- Bachelor degree students
- Master degree students

Participants were given a questionnaire with 7 questions based on the parallel coordinate graph in the PaCQ interface. The questions are related to the 3 relations described earlier, R_1 (word count and word position information), R_2 (vocabulary percentage information) and R_3 (comparing percentage of content of words between *surah*). The questions also incorporated instructions on how to interact with the parallel coordinate graph.

PaCQ was set at a 1280x800 pixel resolution and highest (32 bit) color quality. Participant interacted with the system using the mouse via a laptop Dell Vostro 1310 or 1210. The session was conducted in various conditions, sometimes simultaneous session with up to a maximum of 5 participants and sometimes individually. The questionnaires were given to the participant/s and they were told to answer the question as they go through them one by one. See appendix for the questionnaire.

5.2 The Objectives

The objectives of this study are as the following:

- To evaluate the graphical perceptibility of parallel coordinate before and after using PaCQ interface
- To investigate the correlation between the score and the time taken to complete the task.
- To investigate the performance of interpreting the parallel coordinate graph based on the questionnaire given:
    - The performance related to the percentage score of the questions
    - The performance related to the time taken to complete the task
    - Comparing the time estimated in GOMS analysis and the time to complete tasks in the questionnaire

5.3 Results

The graphical perceptibility: The participants were asked to rate what they thought of the parallel coordinate graph showed to them in the PaCQ interface. The rating was from 1 to 5 with 1 for “very hard to understand” and 5 for very easy to understand. Table 2 shows the detailed description. While Table 3 shows the mean and standard deviation. It can be observed that the mean rating at the start is 1.8 and at the end is 3.2 with standard deviation of 0.68 and 0.86 respectively.

TABLE 3: The estimated time for relation, R_1, R_2 and R_3

	Lowest Level Tasks	Time/s
R_1	S1.1:Locate/identify count axis/ symbol	1.2
	S1.2: Locate the highest count	1.2 +0.2+ 1.2
	S1.3: Locate the corresponding word	0.2
	S2.1.1: Locate/identify *surah* axis/symbol	1.2
	S2.1.2: Read the corresponding *surah* no.	1.2
	S2.2.1: Locate/identify *ayat* axis/symbol	1.2
	S2.2.2: Read the corresponding *ayat* no.	1.2
	TOTAL	8.8
R_2	S1.1: Locate/identify percentage axis/symbol	1.2
	S1.2: Locate/identify the highest percentage value	1.2
	S2.1: Locate the *surah no.*	1.2
	TOTAL	3.6
R_3	S1.1: Locate/identify percentage axis/symbol	1.2
	S1.2: Locate the highest percentage value	1.2 +0.2+ 1.2
	S2.1.1: Locate/identify the base *surah* axis/symbol	1.2
	S2.1.2: Locate/identify the base *surah* no.	1.2
	S2.2.1: Locate/identify the compared *surah* axis/symbol	1.2+1.2
	TOTAL	8.6

Table 2: Description of rating scale

Rating scale	Description
1	Very hard to understand
2	Hard to understand
3	Average/Neutral
4	Easy to understand
5	Very Easy to understand

Table 3: Graphical Perceptibility of Parallel Coordinate

	Perception of Parallel Coordinate at start	**Perception of Parallel Coordinate at end**
Mean	**1.80**	**3.20**
Std. Deviation	**0.676**	**0.862**

The Correlation: The Pearson correlation between percentage of the total score and the time taken to complete task were found to be non-significant, $r(13)= 0.457$; $p>0.05$. Table 3 shows this result. The scatter plot from Figure 4 also shows no linear relation between these two variables.

The performance related to the percentage score: It was found from the one way ANOVA test, that there is no significant difference in the percentage level scores between the three educational level of participants, $F(2,14) = 0.727$, $p > 0.05$. Refer to Tables 5 and 6. In Figure 2, we can see that 93% of participants scored above 70% in interpreting the parallel plot questionnaire.

Table 4: Person Correlations

		Percentage of the total score	**Time Taken to complete task**
Percentage of	Pearson Correlation	1	0.457

the total score	Sig. (2-tailed)		0.087
Time Taken to complete task	Pearson Correlation	0.457	1
	Sig. (2-tailed)	0.087	

The performance related to the time taken to complete the task: Table 7 shows the means and standard deviation of the time taken to complete task. High School, Bachelor Degree and Master degree has 18.2, 35.6 and 24.2 means to complete task respectively. It was found from the one way ANOVA test (see Table 8), that there is significant difference in the time taken to complete the tasks between the three educational level of participants, $F(2,14) = 12.64$, $p < 0.05$. In Table 9.14, the Scheffe's test shows that the bachelor degree participants take more time to complete tasks compared the other two groups. There is no significant difference in the time taken to complete between the Master degree and the High School group.

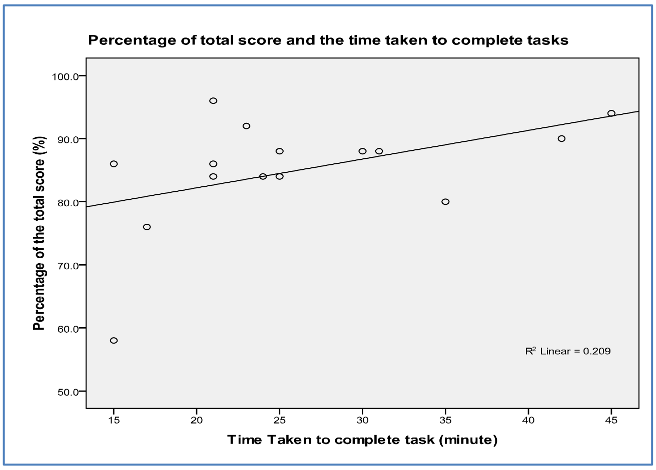

Figure 2: Percentage of total score and the time taken to complete task

Table 5: The means and standard deviation of 3 categories of participants for percentage level score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Highschool	5	1.80	.837	.374	.76	2.84
Bachelor Degree	5	2.20	.447	.200	1.64	2.76
Master Degree	5	2.20	.447	.200	1.64	2.76
Total	15	2.07	.594	.153	1.74	2.40

Table 6: One way ANOVA Percentage level score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.533	2	0.267	0.727	0.503
Within Groups	4.400	12	0.367		
Total	4.933	14			

Table 7: The means and standard deviation of 3 categories of participants for time taken to complete task

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
High School	5	18.20	3.633	1.625	13.69	22.71
Bachelor Degree	5	35.60	8.112	3.628	25.53	45.67
Master Degree	5	24.20	3.701	1.655	19.60	28.80
Total	15	26.00	9.071	2.342	20.98	31.02

Table 8: One way ANOVA for time taken to complete task

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	781.200	2	390.600	12.641	.001
Within Groups	370.800	12	30.900		
Total	1152.000	14			

Table 9: Comparison of time taken to complete task between participants in the 3 levels of education

	(I) Education Level	(J) Education Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	High School	Bachelor Degree	-17.400*	3.516	.001	-27.20	-7.60
		Master Degree	-6.000	3.516	.271	-15.80	3.80
	Bachelor Degree	High School	17.400*	3.516	.001	7.60	27.20
		Master Degree	11.400*	3.516	.023	1.60	21.20
	Master Degree	High School	6.000	3.516	.271	-3.80	15.80
		Bachelor Degree	-11.400*	3.516	.023	-21.20	-1.60

*. The mean difference is significant at the 0.05 level.

6 Discussion

The mean rating for the graphical perceptibility at the starting point when the participants were using the system was low (1.8), in other words the participant found the parallel coordinate graph hard to understand. This is also consistent with the findings of other researches on parallel coordinate [5, 8, 11, 12]. However, at the end of the session the average rating increases to 3.8, meaning that once the participants know how to interact with the graph, they found that it is much easier then what they have expected. The data shown to users can be interpreted accurately even by novice users, this was also found by Siirtola & Räihä [12] , Siirtola, Laivo, Heimonen, & Raiha [14] and Henley, Hagen, & Bergeron [13] as discussed in previous section.

It was investigated whether the time taken to complete the task has any correlation with the percentage score attained by the participants. Results shows no signifcant relation and the scatter plot between these two variables shows no linear relationship between them. However, the performance related to the mean time taken to complete the task between the Bachelor degree groups and the other two groups(the High School and the Master level) were significantly different. There is no significant difference in the mean time taken to complete the task between the the High School and the Master level groups. This indicates that the higher the education level does not nescessarily determine a faster speed of completion.

The performance related to the percentage of scores between the three levels of education groups was found to be not significant. Again, this indicates that higher education level does not nescessarily determine a better performance percentage score. Here, there is evidence showing that novice users can also interpret the parallel coordinate graph accurately contrasting the arguments of Yuan, Guo, Xiao, Zhou, & Qu [8] and Henley, Hagen, & Bergeron [13] stating that expert users are needed to interp-ret the graph.

7 Conclusion

A relational model of parallel coordinate was used for analysis based on an interface called PaCQ. Although through literatures it was found that parallel plot does not score well in graphical perceptibility among users, results shows that novice users can interpret parallel plots relatively well. The GOMS analysis of tasks related to parallel plot also shows that the time taken to complete tasks related to the parallel plot is relatively reasonable, well below 10s. The user study among novice with three different levels of education shows no significant difference in the scores of interpreting the parallel plot.

References

[1] Fisher, J., *Defining the novice user.* Behaviour & In-formation Technology, 1991. **10(5), 437-441**.

[2] Inselberg, A., Parallel Coordinates, visual multidimensional geometry and its application. Springer, London, 2009.

[3] Keim, D.A. and H. Kriegel, *Visualization Techniques for Mining Large Databases: A Comparison.* IEEE Trans. on Knowl. and Data Eng, 1996. **8, 6 (Dec. 1996), 923-938**.

[4] Chou, S.-Y., S.-W. Lin, and C.-S. Yeh, *Cluster identifi-cation with parallel coordinates.* Pattern Recognition Letters, 1999 **Volume 20, Issue 6, Pages 565-572**.

[5] Leban, G., et al., *VizRank: Data Visualization Guided by Machine Learning.* Data Min. Knowl. Discov., 2006. **13**(2): p. 119-136.

[6] Hauser, H., F. Ledermann, and H. Doleisch, *Angular brushing of extended parallel coordinates.* Proceedings of the IEEE Symposium on Information Visualization (InfoVis'02), IEEE Computer Society (2002) pp. 127-131, 2002.

[7] Rubik, G., *Extruded Parallel Coordinates.* Last update: 19 July 1997, Last Accessed: 26 March 2010, URL, 2010.

[8] Yuan, X., et al., *Scattering Points in Parallel Coordinates.* IEEE Transactions on Visualization and Computer Graphics, 2009. **pp. 1001-1008, November/December, 2009**.

[9] Heinrich, J. and D. Weiskopf, *Continuous Parallel Coor-dinates.* IEEE Transactions on Visualization and Computer Graphics, 2009. **Pages 1531-1538**.

[10] Johansson, J., et al., Perceiving patterns in parallel coordinates: determining thresh-olds for identification of relationships. Information Visualization 2008. **Vol 7, 2, 152-162**.

[11] Forsell, C. and J. Johansson, *Task-based evaluation of multi-relational 3D and standard 2D parallel coordinates.* IS&T/SPIE's International Symposium on Electronic Imaging, Conference on Visualization and Data Analysis 2007 (San Jose, USA, 2007), SPIE: Bellingham and IS&T: Springfield, 2007.

[12] Siirtola, H. and K.J. Räihä, *Interacting with parallel coordinates.* Interacting with Computers, 2006. **18(6):1278-1309**.

[13] Henley, M., M. Hagen, and R.D. Bergeron, *Evaluating Two Visualization Techniques for Genome Comparison.* Information Visualization IV '07. 11th International Conference , vol., no., pp.551-558, 4-6 July 2007, 2007.

[14] Siirtola, H., et al., *Visual Perception of Parallel Coordinate Visualizations.* Information Visualisation, 2009 13th International Conference , vol., no., pp.3-9, 15-17 July 2009, 2009.

[15] Novotny, M. and H. Hauser, *Outlier-Preserving Focus+Context Visualization in Parallel Coordinates.* Visualization and Computer Graphics, IEEE Transactions, 2006. **vol.12, no.5, pp.893-900**.

[16] Heer, J., N. Kong, and M. Agrawala, *Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations.* In Proceedings of the 27th interna-tional Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 1303-1312, 2009.

[17] Matkovic, K., et al., *Interactive visual analysis and exploration of injection systems simulations.* Visualization and Computer Graphics, IEEE Transactions, 2005. **VIS 05. IEEE , vol., no., pp. 391- 398**.

[18] Choi, H., H. Lee, and H. Kim, *Fast detection and visualization of network attacks on parallel coordinates.* Computers & Secu-rity, 2009. **Vol 28, Issue 5, July 2009, Pages 276-288**.

[19] Yong, G., et al., *Exploring uncertainty in remotely sensed data with parallel coordinate plots.* International Journal of Applied Earth Observation and Geoinformation, 2009. **Volume 11, Issue 6, pp 413-422**.

[20] Chad, S.A., et al., An interactive parallel coordinates technique applied to a tropical cyclone climate analysis. Computers & Geosciences, 2009. **Vol 35, Issue 7, pp 1529-1539**.

[21] Weber, C.A. and A. Desai, Determination of paths to vendor market efficiency using parallel coordinates representa-tion: A negotiation tool for buyers. European Journal of Opera-tional Research, 1996. **Volume 90, Issue 1, pp 142-155**.

[22] Swayne, D.F., et al., GGobi: evolving from XGobi into an extensible framework for interactive data visualization. Comput.

Stat. Data Anal. , 2003. **43, 4, 423-444**.

[23] John, B.E. and D.E. Kieras, *The GOMS family of user interface analysis techniques: comparison and contrast.* . ACM Trans. Comput.-Hum. Interact, 1996. **3**(4): p. 320-351.