Supplemental Materials for Kiviat Defense: An Empirical Evaluation of Visual Encoding Effectiveness in Multivariate Data Similarity Detection

M. Mantovani et al.

Reproducibility

This work uses a real dataset. The patient data, although anonymized and de-identified, is real, proprietary data collected at the MD Anderson Cancer Center (MDACC), University of Texas, and processed at the University of Illinois at Chicago (UIC). This dataset can be released through a Material Transfer Agreement between the requester and MDACC. All other user data (e.g., demographics) are protected by the UIC IRB.

Experimental Setup

Here we give more details about the interface used during the users tasks.

At the beginning of the study, each user was informed of the nature of the trial and that it would take approximately 30 minutes. All users were given consent forms to fill out inside the interface, followed by a demographics questionnaire (Figure 1 A-B). Afterwards, each user was given a tutorial that gave example tasks for each encoding type that was not timed, followed by the main trails.

In each task, the data items were shown on the screen; each item was identified by a numerical ID displayed next to that item. The tutorial consisted of 3 trials, one for each encoding, excluding color-cue variants. When performing tasks with glyph encodings, participants performed the similarity task on a 3 x 3 grid. Instructions were displayed at the top of the screen, in the following format: "This is a demo serving as introduction to the tasks. Click on the 3 most similar items to the item number 4" (Figure 2). The user selections were acknowledged by a brief highlighting of their selection. For the overlaid encoding section of the tutorial, the target item was shown with a thicker line, and hovering over another item highlighted that item, to better support visual identification of that item. This implementation replicated the brushing operation available in practice for this type of encodings. After each tutorial trial, we revealed the right answer by expanding the text message at the top of the screen (i.e.,"The correct answers are: 8, 2, 7"). During this stage, we answered any questions the participants had about the study or how similarity was measured.

For the main study, trial items were selected randomly from the database, and one item was randomly selected to be the 'target' item. For the glyph encodings, items were arranged in a grid of 4 by 4 (16 items), or 6 by 6 (36 items). Regardless of the encoding, the user was prompted to select the 3 items that were most similar to the target item by clicking on them with the mouse. When 3 items were selected, the user was allowed to process to the next trial by selecting the 'next' button (Figure 3). In the main experiment, participants performed 10 trials each, for the two test scales. During the test session, the ground truth similarity was not revealed to the user. A progress bar allowed participants to see their overall experiment progress.

Each study session was performed on a web browser on the same laptop (15.4-inch display, 2880 x 1800 resolution). The participants used only mouse interaction during each trial.

Additional Statistical Analysis Data

Please note that the results reported in the manuscript are based on the raw, untransformed data. Below we report skewness and results from normality testing, before and after applying a power transform. Tables 1 and 2 show error statistics before and after applying a Yeo-Johnson power transformation, respectively. Tables 3 show statistics for trial time across encodings with outliers removed before applying a data transformation, while Table 4 shows the same values obtained after performing a Yeo-Johnson power transformation.

Score-Time Analysis

We also generated scatterplots of trial time for each encoding type along with results from fitting spine fits to the data. Results are excluded as they did not give statistically meaningful results, although PCPs did register significantly lower scores despite reasonable attempt times (in the 30-90 second range).

Figure 1: Images of forms given to the user. (A) Consent form shown at the start of the study. (B) demographics form given before the beginning of the study. (C) Subjective questionnaire given at the end of the study.

Figure 2: Example image of one of the tasks given during the introductory tutorial given before the start of the study, showing the use of color-cue Kiviat encodings. Correct answers are shown at the top of the screen. Separate examples were given for each encoding type.

Figure 3: Screenshots of the interface during the similarity task. A progress bar and simple instructions are shown at the top of the screen. (A) Colored Kiviat diagrams for 16 items. The user is prompted with the id of the target item and selects the most similar items by clicking. (B) Parallel Coordinate plot with 16 items. The target item is shown as a thicker green line. Similar items can be brushed by mousing over them, and selected via a mouse click. Once 3 items are selected the Next button in the bottom left is enabled to allow the user to stop the timer on the current trial and show the next trial.

Table 1: Skewness and normality statistics for the raw error. The LSP (moderate), PCP (large) and Color-cue PCP values are normally distributed (Shapiro P > 0.05).

Encoding	Setting	Min	Mean	Max	Skew	Kurtosis	Shapiro W	Shapiro P
Overall	–	4.50	4.82	5.13	1.46	2.02	0.86	0.0000
Kiviat	Moderate	1.94	2.47	2.98	0.76	-0.52	0.90	0.0023
Kiviat	Large	3.62	5.04	6.37	1.22	0.46	0.83	0.0000
Color-cue Kiviat	Moderate	1.78	2.34	2.87	1.41	1.68	0.86	0.0001
Color-cue Large	Large	2.92	3.98	4.95	1.05	0.20	0.87	0.0002
LSP	Moderate	2.13	2.57	3.00	0.42	-0.78	0.96	**0.1156**
LSP	Large	4.84	6.21	7.49	0.96	0.28	0.91	0.0034
PCP	Moderate	3.40	3.97	4.49	1.17	1.14	0.90	0.0014
PCP	Large	9.51	10.81	12.08	0.53	0.61	0.98	**0.6989**
Color-cue PCP	Moderate	3.55	4.14	4.73	-0.07	-0.90	0.97	**0.3066**
Color-cue PCP	Large	7.14	8.50	9.86	0.25	-1.22	0.92	0.0089

Table 2: Statistics for the error after using a Yeo-Johnson power transformation. More distributions have normally distributed values (Shapiro P > 0.05) but the overall results are still not normally distributed.

Encoding	Setting	Min	Mean	Max	Skew	Kurtosis	Shapiro W	Shapiro P
Overall	–	1.53	1.59	1.64	-0.01	-0.59	0.98	0.0000
Kiviat	Moderate	0.94	1.11	1.29	-0.06	-0.94	0.95	**0.0840**
Kiviat	Large	1.25	1.51	1.77	0.09	-1.02	0.95	**0.0719**
Color-cue Kiviat	Moderate	0.91	1.08	1.25	0.21	-0.47	0.97	**0.3988**
Color-cue Large	Large	1.12	1.36	1.60	-0.07	-1.09	0.93	0.0176
LSP	Moderate	1.06	1.20	1.35	-0.36	-0.70	0.96	**0.2377**
LSP	Large	1.58	1.81	2.05	-0.32	-0.70	0.97	**0.2683**
PCP	Moderate	1.49	1.61	1.73	0.18	-0.13	0.97	**0.3790**
PCP	Large	2.44	2.58	2.73	-0.97	1.49	0.95	**0.0584**
Color-cue PCP	Moderate	1.44	1.60	1.76	-0.99	0.36	0.90	0.0023
Color-cue PCP	Large	2.04	2.24	2.43	-0.30	-1.33	0.91	0.0039

Table 3: Skewness and normality statistics for the raw time (in seconds), after removing three extreme outliers.

Encoding	Setting	Min	Mean	Max	Skew	Kurtosis	Shapiro W	Shapiro P
Overall	–	46.75	48.59	50.38	1.51	3.77	0.90	0.0000
Kiviat	Moderate	48.10	53.67	59.25	0.28	-0.82	0.97	**0.3139**
Kiviat	Large	43.67	48.09	52.61	0.22	-0.89	0.97	**0.4448**
Color-cue Kiviat	Moderate	34.70	39.93	44.70	1.69	3.84	0.85	0.0001
Color-cue Large	Large	32.28	37.36	41.81	1.94	6.38	0.85	0.0001
LSP	Moderate	38.48	46.66	54.27	1.29	0.65	0.83	0.0000
LSP	Large	42.65	49.59	56.15	0.90	0.47	0.94	0.0334
PCP	Moderate	60.12	70.57	80.21	1.36	2.02	0.89	0.0020
PCP	Large	49.15	57.65	65.30	1.34	1.84	0.89	0.0012
Color-cue PCP	Moderate	40.41	45.89	51.16	0.81	1.15	0.95	**0.1339**
Color-cue PCP	Large	40.75	46.29	51.67	0.60	-0.29	0.95	**0.1337**

Table 4: Statistics for the time after removing the three extreme outliers and using a Yeo-Johnson power transformation. The overall distribution has normally distributed values (Shapiro P > 0.05).

Encoding	Setting	Min	Mean	Max	Skew	Kurtosis	Shapiro W	Shapiro P
Overall	–	4.23	4.27	4.32	0.00	0.04	1.00	**0.9900**
Kiviat	Moderate	4.30	4.44	4.59	-0.40	-0.54	0.97	**0.3488**
Kiviat	Large	4.20	4.33	4.45	-0.34	-0.73	0.97	**0.4025**
Color-cue Kiviat	Moderate	3.92	4.06	4.21	0.31	0.12	0.96	**0.2271**
Color-cue Large	Large	3.83	3.98	4.14	0.02	0.61	0.98	**0.6892**
LSP	Moderate	3.99	4.17	4.36	0.49	-0.55	0.95	**0.0807**
LSP	Large	4.12	4.29	4.46	-0.10	-0.60	0.98	**0.8482**
PCP	Moderate	4.57	4.74	4.91	-0.01	0.21	0.99	**0.9937**
PCP	Large	4.30	4.48	4.66	-0.54	1.96	0.95	**0.0910**
Color-cue PCP	Moderate	4.08	4.23	4.39	-0.57	0.62	0.97	**0.3846**
Color-cue PCP	Large	4.08	4.24	4.40	-0.33	-0.20	0.98	**0.5947**

A

Case study: effectiveness of different visual encodings in item-item similarity identification

This is a web-based experiment aimed at analysing the effectiveness of different types of visual encodings for comparison tasks. You will have 12 total trials, initial instructions will be given at the beginning and are not part of the test. You'll also be asked some additional personal questions for the purpose of understanding if the perfomances are related to demographic features, all data will be anonymized. Participation takes between 15 and 25 minutes.

Please enter your nickname here:

User

CHECK ID

I AGREE

B

Please fill out the following demographics form.

Your age:

○ 18-24 ○ 25-29 ○ 30-39 ○ 40-49 ○ 50-59 ○ 60+ ● Unspecified

Your gender:

○ Male ○ Female ● Unspecified

Highest degree obtained:

○ High School ○ Bachelors ○ Masters ○ PhD ● Other

Familiarity with visualization and visualizing data in charts:

(**Not Familiar:** you have difficulties in reading and understanging basic charts, **Basic knowledge:** you know how to read basic charts as line or pie charts but you don't use or see them often, **Familiar:** you can interpret most of the common types of chart and you use or interpret them often, **Expert:** you have taken one or more courses in visualization and most likely would interpret correctly unseen types of chart)

○ Not familiar ○ Basic knowledge ○ Familiar ○ Expert ● Other

NEXT

C

k8yp0z5g

This study was designed to test the ability of people to find similar and dissimilar items of a multi-variate ordinal dataset using different visual encodings. It was hypothesized that the pre-attentive nature of some encodings would aid users in seeing similar and dissimilar examples. This study compares the performances using different type of encodings, both the state of the art charts for this type of tasks as well as new and unexplored types of visual encodings.

Which of the encodings did you think was most difficult for you to process during the task?

○ Kiviat Diagram (With Shapes) ○ Kiviat Diagram (Lines Only) ○ Stacked Bar Chart ○ Circular Slices Chart ○ Parallel Coordinate Plot ○ Unspecified

Why? (If there is no specific reason leave blank)

Which of the encodings did you think was easiest for you to process during the task?

○ Kiviat Diagram (With Shapes) ○ Kiviat Diagram (Lines Only) ○ Stacked Bar Chart ○ Circular Slices Chart ○ Parallel Coordinate Plot ○ Unspecified

Why? (If there is no specific reason leave blank)

Scalability: an encoding is scalable when increasing the number of items does not lead to excessive increase of efforts in completing the task

Which of the encodings do you think is most scalable?

○ Kiviat Diagram (With Shapes) ○ Kiviat Diagram (Lines Only) ○ Stacked Bar Chart ○ Circular Slices Chart ○ Parallel Coordinate Plot ○ Unspecified

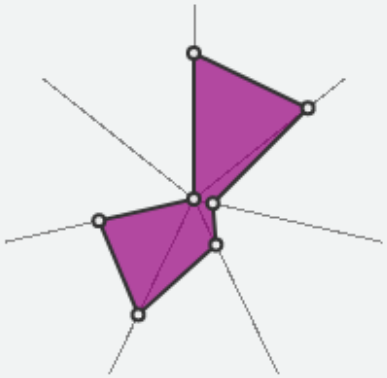
Which of the encodings do you think scales the worst?

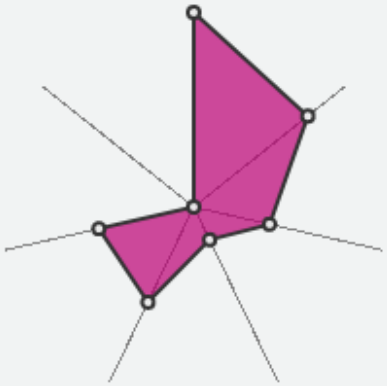
○ Kiviat Diagram (With Shapes) ○ Kiviat Diagram (Lines Only) ○ Stacked Bar Chart ○ Circular Slices Chart ○ Parallel Coordinate Plot ○ Unspecified

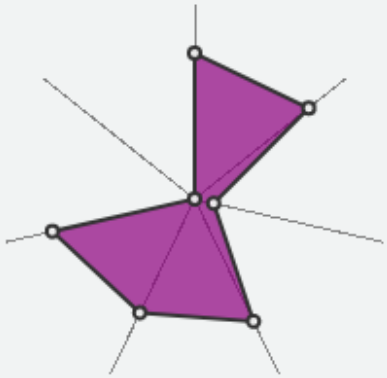
Thank you again for your participation. Feel free submit any additional comments below.

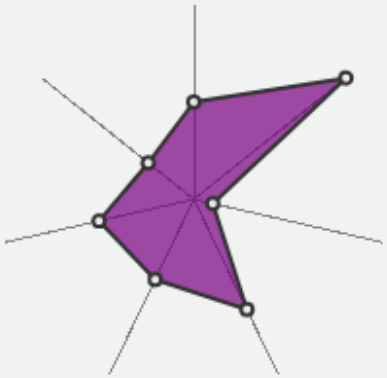
SUBMIT

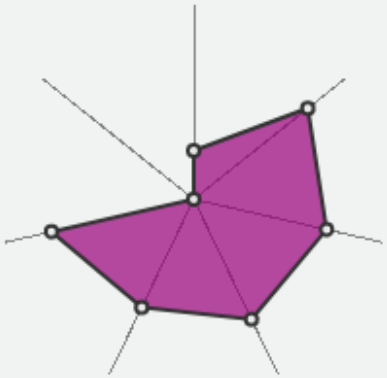
This is a demo serving as introduction to the tasks. Click on the 3 most similar items to the item number **6**. The correct answers are: **9, 4, 3**

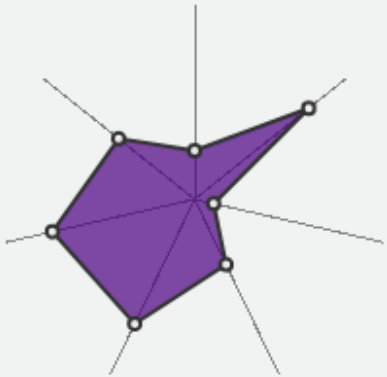

1

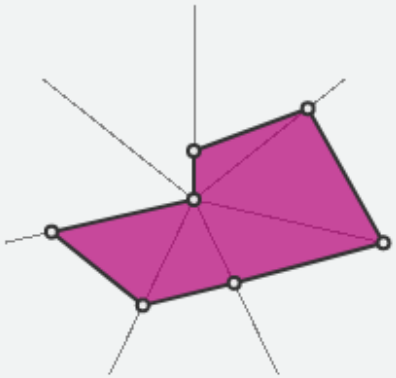

2

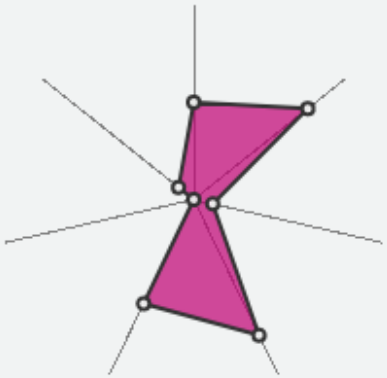

3

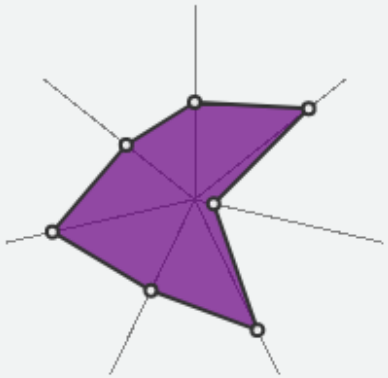

4

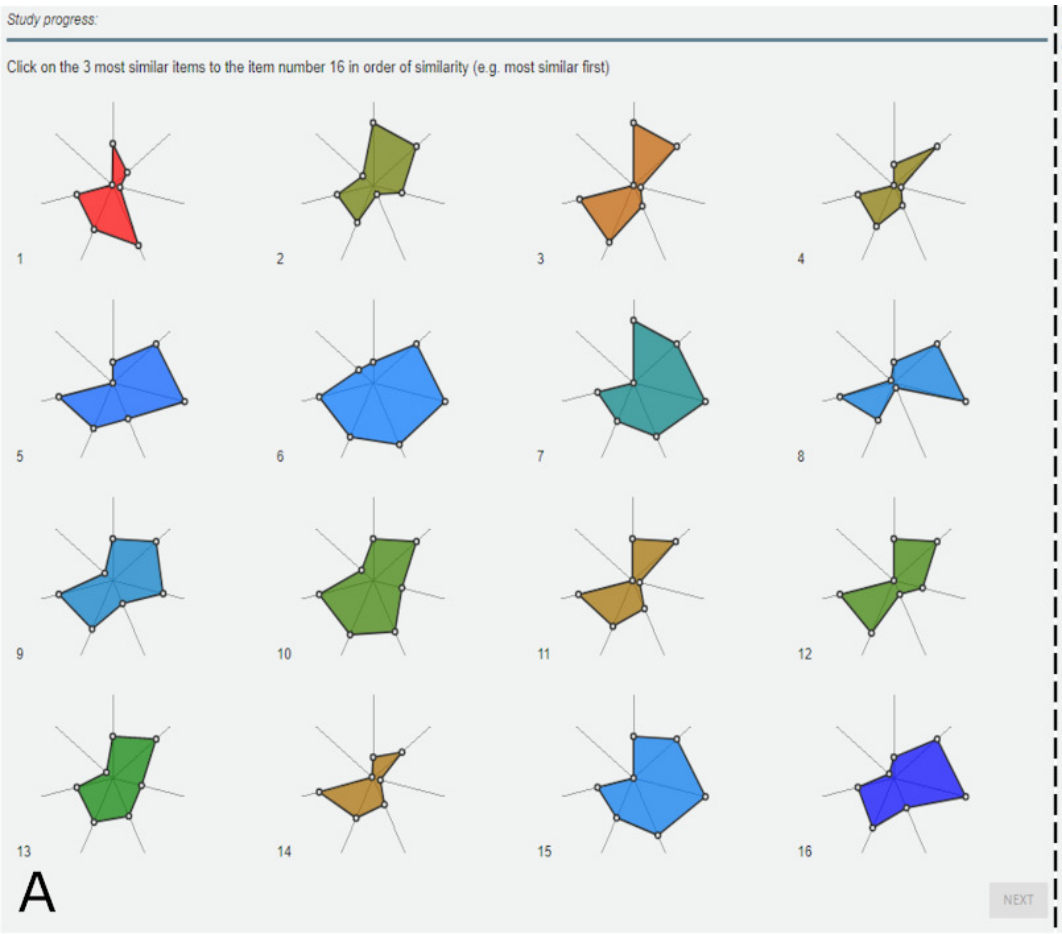

5


6

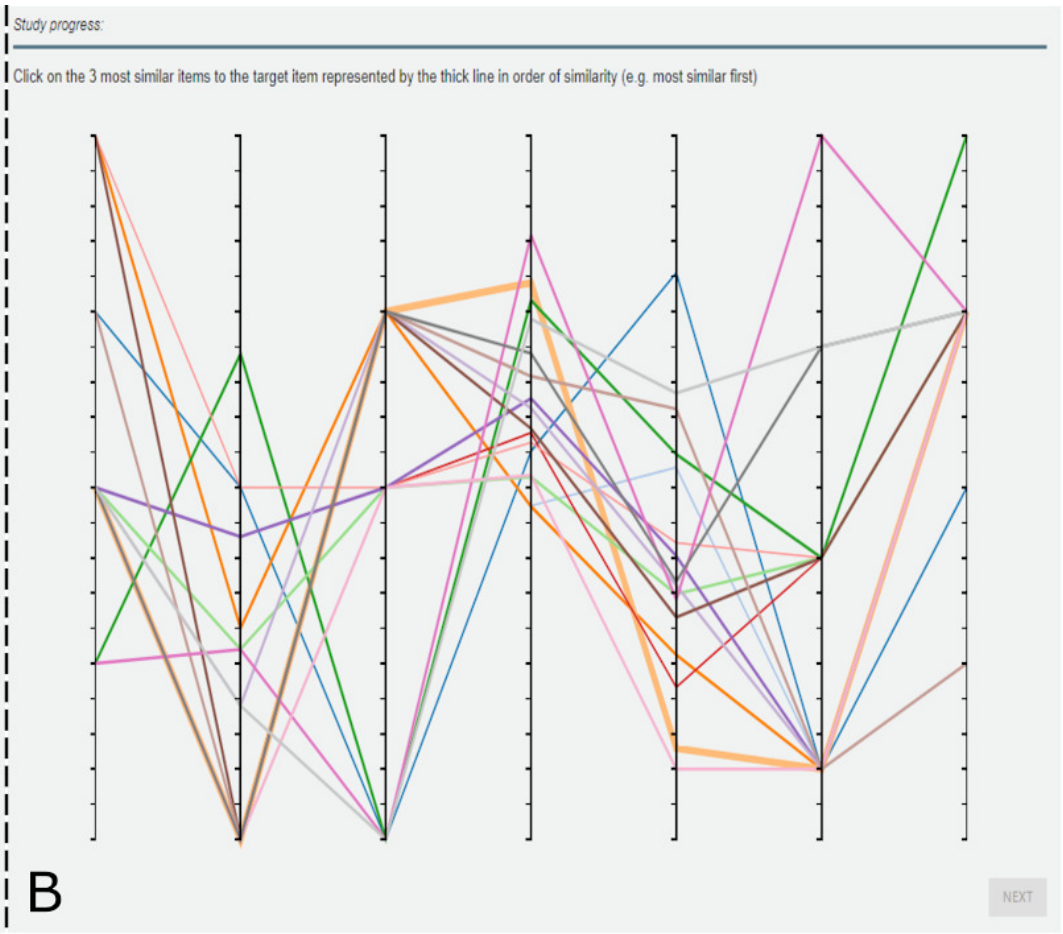

7


8


9


Study progress:
Click on the 3 most similar items to the item number 16 in order of similarity (e.g. most similar first)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
NEXT

A


Study progress:
Click on the 3 most similar items to the target item represented by the thick line in order of similarity (e.g. most similar first)
NEXT

B