An Advanced Face Glyph Design for Accessible Multivariate Cluster Analysis

XIAOFEI ZHOU, University of Rochester, United States
ABDUL MOID MUNAWAR, University of Rochester, United States
DANIEL MUNOZ HUERTA, University of Rochester, United States
ZHEN BAI, University of Rochester, United States

To facilitate more accessible and engaging cluster analysis for novice learners without computing background, we proposed three innovative face glyph design ideas, which aim to provide a playful interaction experience and tackle existing challenges of face glyph visualization. Preliminary results for one design idea echo with known challenges for face glyphs and indicate directions for improvements. Evaluation is needed for the rest two face glyph design ideas.

CCS Concepts: • **Human-centered computing** → **Visual analytics**.

Additional Key Words and Phrases: glyph visualization, cluster analysis, multidimensional data

ACM Reference Format:
Xiaofei Zhou, Abdul Moid Munawar, Daniel Munoz Huerta, and Zhen Bai. 2021. An Advanced Face Glyph Design for Accessible Multivariate Cluster Analysis. 1, 1 (March 2021), 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

It's challenging for young learners to compute the differences between two multidimensional data points, or interpret the shared or differentiating patterns of a subset, because the common core mathematics curriculum for K-12 covers neither patterns in multivariate data nor Euclidean distance on multidimensional feature space [3]. Therefore, we are exploring how to design **beginner-friendly and engaging** data visualization to facilitate novice learners to interpret multidimensional data for cluster analysis, which is a common branch of pattern recognition [6].

Face glyph [2] is one type of the glyph-based data visualization that uses a graphical representation (e.g. a star, a leaf, a bug [4, 18]) to represent individual multidimensional data points. By mapping each data attribute into a facial feature such as eye radius and mouth width, face glyph takes advantage of people's special processing capacity and natural sensitivity to human faces [10], and has been applied to facilitate visual reasoning across various domains [1, 5, 9, 15]. Although previous research proved face glyph's effectiveness as a learning method for students' science learning [12] and pattern interpretation [16], design challenges remain [8, 9, 17], including the potential additional cognitive load brought by translating glyph graphical features back to data attribute (**c#1**), the limited number of data

Authors' addresses: Xiaofei Zhou, University of Rochester, 500 Joseph C. Wilson Blvd., Rochester, United States, xzhou50@ur.rochester.edu; Abdul Moid Munawar, University of Rochester, 500 Joseph C. Wilson Blvd., Rochester, United States, amunawar@u.rochester.edu; Daniel Munoz Huerta, University of Rochester, 500 Joseph C. Wilson Blvd., Rochester, United States, dmunoz2@u.rochester.edu; Zhen Bai, University of Rochester, 500 Joseph C. Wilson Blvd., Rochester, United States, zhen.bai@rochester.edu.

attributes that a face glyph can afford (**c#2**), perceptual impact on interpretation (**c#3**). In this work, we proposed three innovative face glyph design ideas and reported preliminary evaluation results to tackle these challenges.

2 HOVERING INTERACTION FOR FACE GLYPH AND EVALUATION RESULTS

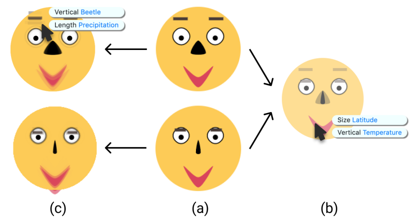

Fig. 1. Design idea #1 - Hovering interaction: (a) Face glyphs of field sites representing ecosystem clusters. (b)&(c) Overlay for pairwise & groupwise comparison.

We utilized the **face-overlay design** [16] for novices to interpret intra-cluster pattern (i.e. identify shared features through group-wise comparison in Fig. 1(c)) and inter-cluster pattern (i.e. identify differentiating features through pairwise comparison in Fig. 1(b)). To eliminate potential cognitive load added by translating between graphical features and data attributes (**c#1**), the **hovering interaction** is designed: users can hover their cursor on different parts of the face glyph to view the data attributes mapped to corresponding face areas (Fig. 1(b)&(c)). Preliminary evaluation with novices indicate manually hovering on face glyphs is still not sufficient for learners to generate scientific explanation with minimum cognitive load and more automated scaffolding is needed [13].

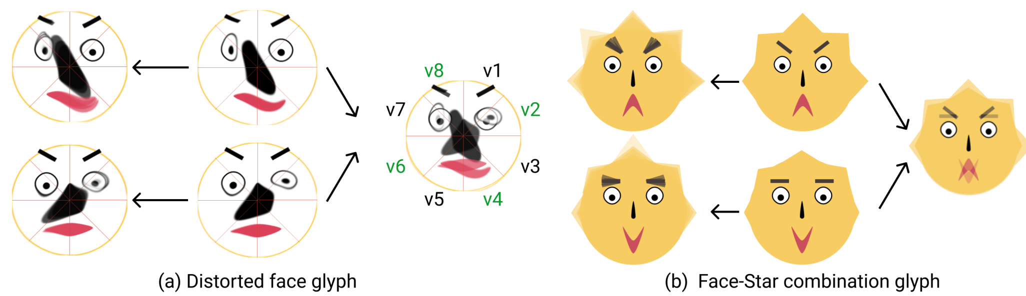

Fig. 2. Design idea #2 and #3: Two innovative face glyphs

3 DISTORTED FACE GLYPH AND FACE-STAR COMBINATION GLYPH

For challenges **c#2** and **c#3** originated from the particular visualization - face, we proposed to employ the mechanism of star glyphs [14, 19]. Compared to face glyphs, star glyphs have less perception-based bias and proximity-based bias by mapping individual data attribute values to the length of evenly spaced rays emanating from center [17], offer more flexibility in the number of data attributes. Existing work has proved its power in multivariate cluster analysis [7].

To address **c#2** that the number of data attributes a face glyph can afford is constrained by the number of facial features, we designed the **distorted face glyph** which translates individual data attribute values to corresponding force that distorts the face glyph from a certain direction (Fig. 2(a)). With this design idea, the distortion effect can be applied on the unlimited number of directions to generate glyphs for high-dimensional data. For **c#3** that face glyph with different data values may bring different perceptual impacts, we proposed a design idea: **Face-Star combination glyph** which can turn this limitation into an advantage. With a breast cancer dataset [11], the data attribute about whether a cell is benign or malignant can be mapped to mouth shape or brow slant so that benign cells produce happy faces and malignant cells produce sad faces. And other cellular morphological attributes can be mapped to the star-glyph features around the face which are more perceptually neutral to people. For **future work**, we will evaluate the effectiveness of two innovative face glyphs and interaction improvements in supporting young learners to interpret patterns for different data analysis tasks besides cluster analysis. The evaluation will keep emphasizing on the three challenges mentioned above (**c#1**, **c#2**, **c#3**).

REFERENCES

[1] Anthony Chan, Leyland F Pitt, and Deon Nel. 2014. Let's face it: Using Chernoff faces to Portray Social Media Brand Image. *Corporate Ownership & Control* 609 (2014).

[2] Herman Chernoff. 1973. The use of faces to represent points in k-dimensional space graphically. *Journal of the American statistical Association* 68, 342 (1973), 361–368.

[3] National Governors Association Center for Best Practices. 2010. C. o. C. S. S. O. Common Core Science Standards. National Governors Association Center for Best Practices, Council of Chief State School.

[4] Johannes Fuchs, Niklas Weiler, and Tobias Schreck. 2015. Leaf glyph visualizing multi-dimensional data with environmental cues. (2015).

[5] Wang Hongwei and Liu Hui. 2013. Quantitatively plotting the human face for multivariate data visualisation illustrated by health assessments using laboratory parameters. *Computational and mathematical methods in medicine* 2013 (2013).

[6] Leonard Kaufman and Peter J Rousseeuw. 2009. *Finding groups in data: an introduction to cluster analysis*. Vol. 344. John Wiley & Sons.

[7] Mandy Keck, Dietrich Kammer, Thomas Gründer, Thomas Thom, Martin Kleinsteuber, Alexander Maasch, and Rainer Groh. 2017. Towards glyph-based visualizations for big data clustering. In *Proceedings of the 10th international symposium on visual information communication and interaction*. 129–136.

[8] Young-seok Kim and Louise Cooke. 2017. Big data analysis of public library operations and services by using the Chernoff face method. *Journal of Documentation* (2017).

[9] Michael D Lee, Rachel E Reilly, and Marcus E Butavicius. 2003. An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data. In *Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24*. 1–10.

[10] Klaus Libertus, Rebecca J Landa, and Joshua L Haworth. 2017. Development of attention to faces during the first 3 years: Influences of stimulus type. *Frontiers in psychology* 8 (2017), 1976.

[11] Olvi L Mangasarian and William H Wolberg. 1990. *Cancer diagnosis via linear programming*. Technical Report. University of Wisconsin-Madison Department of Computer Sciences.

[12] José Jesús Reyes Nuñez, Anita Rohonczi, Cristina E Juliarena de Moretti, Ana María Garra, Carmen Alicia Rey, María V Alves de Castro, Anabella S Dibiase, Teresa A Saint Pierre, and Mariana A Campos. 2011. Updating research on Chernoff faces for school cartography. In *Advances in Cartography and GIScience. Volume 2*. Springer, 3–20.

[13] Chris Quintana, Brian J Reiser, Elizabeth A Davis, Joseph Krajcik, Eric Fretz, Ravit Golan Duncan, Eleni Kyza, Daniel Edelson, and Elliot Soloway. 2004. A scaffolding design framework for software to support science inquiry. *The journal of the learning sciences* 13, 3 (2004), 337–386.

[14] John H Siegel, SIEGEL JH, FARRELL EJ, and GOLDWYN RM. 1972. THE SURGICAL IMPLICATIONS OF PHYSIOLOGIE PATTERNS IN MYOCARDIAL INFARCTION SHOCK. (1972).

[15] Ruixia Song, Zhaoxia Zhao, and Xiaochun Wang. 2010. An application of the V-system to the clustering of Chernoff faces. *Computers & Graphics* 34, 5 (2010), 529–536.

[16] Xiaoyu Wan, Xiaofei Zhou, Zaiqiao Ye, Chase K Mortensen, and Zhen Bai. 2020. SmileyCluster: supporting accessible machine learning in K-12 scientific discovery. In *Proceedings of the Interaction Design and Children Conference*. 23–35.

[17] Matthew O Ward. 2008. Multivariate data glyphs: Principles and practice. In *Handbook of data visualization*. Springer, 179–198.

[18] Matthew O Ward, Georges Grinstein, and Daniel Keim. 2010. *Interactive data visualization: foundations, techniques, and applications*. CRC Press.

[19] Colin Ware. 2019. *Information visualization: perception for design*. Morgan Kaufmann.