



Square-glyphs: assessing the readability of multidimensional spatial data visualized as square-glyphs

Gianna Daniela Müller, Daria Hollenstein, Arzu Çöltekin & Susanne Bleisch

To cite this article: Gianna Daniela Müller, Daria Hollenstein, Arzu Çöltekin & Susanne Bleisch (2023): Square-glyphs: assessing the readability of multidimensional spatial data visualized as square-glyphs, International Journal of Cartography, DOI: [10.1080/23729333.2023.2235492](https://doi.org/10.1080/23729333.2023.2235492)

To link to this article: <https://doi.org/10.1080/23729333.2023.2235492>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 Jul 2023.



Submit your article to this journal [↗](#)



Article views: 91



View related articles [↗](#)



View Crossmark data [↗](#)

Square-glyphs: assessing the readability of multidimensional spatial data visualized as square-glyphs

Gianna Daniela Müller^a, Daria Hollenstein^a, Arzu Çöltekin^b and Susanne Bleisch^a

^aInstitute Geomatics, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttensz/Basel, Switzerland; ^bInstitute for Interactive Technologies, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Brugg-Windisch, Switzerland

ABSTRACT

Glyphs have long been used to approach the challenge of visualising multidimensional data with geospatial reference. Depending on the glyph design, data-dense visualizations of several concurrent data dimensions can be created. The square-glyph is a compound glyph to represent up to four data dimensions, e.g. walkability indices, with reference to a gridded geographic space (Bleisch and Hollenstein 2018 [*Exploring multivariate representations of indices along linear geographic features. Proceedings of the 2017 International Cartographic Conference, Washington D.C.* (pp. 1–5)]). In this paper, we present a user study to evaluate the readability and interpretability of the square-glyphs. We compare user performance with square-glyph plots containing two and four simultaneously mapped data dimensions under different value compositions. Our results show that the user performance with square-glyphs does not decrease as the number of data dimensions represented increases from two to four. The study results indicate no significant differences in efficiency and effectiveness between the four-dimensional square-glyphs and the two-dimensional square-glyphs. The average values of five adjacent glyphs can be estimated with a mean error of eight percentage points. The results suggest that equal value distances between the displayed dimensions are more accurately perceived in a lower-value composition than in higher-value arrangements.

RÉSUMÉ

Les glyphes ont longtemps été utilisés pour répondre au défi de représenter des données multidimensionnelles à référence spatiale. En fonction du design du glyphe, des représentations de données denses à plusieurs dimensions peuvent être créées. Le glyphe carré est un composant de glyphe pouvant représenter jusqu'à quatre dimensions, par exemple des indices de potentiel piétonnier référencé sur un espace quadrillé (Bleisch and Hollenstein 2018). Dans cet article nous présentons une étude utilisateur pour évaluer la lisibilité et l'interprétabilité des glyphes carrés. Nous comparons la performance utilisateur avec des

ARTICLE HISTORY

Received 14 April 2023

Accepted 8 July 2023

KEYWORDS

Glyph visualization; square-glyphs; multidimensional spatial data; evaluation; user study

structures de glyphes-carrés comportant deux ou quatre dimensions de données simultanément cartographiées sous différentes compositions de valeurs. Nos résultats montrent que la performance des utilisateurs ne baisse pas lorsque le nombre de dimension représenté passe de deux à quatre. Les résultats de l'étude ne révèlent aucune différence significative en efficacité ou en efficacité entre les glyphes-carrés de quatre dimensions et ceux de deux dimensions. Les valeurs moyennes de cinq glyphes adjacents peuvent être estimées avec une erreur moyenne de huit sur cent. Les résultats suggèrent que les distances de valeurs identiques entre les dimensions affichées sont perçues de façon plus précise pour les compositions de plus faibles valeurs que de plus grandes valeurs.

1. Introduction

Glyphs have long been used to represent multidimensional data in small space (e.g. Anderson, 1957). Square-glyphs (Figure 1) allow mapping up to four data dimensions or four data sets to differently colored centered squares in one representation (Bleisch & Hollenstein, 2018). Square-glyphs convey categorical data through color hue and quantitative values through area, following standard recommendations (e.g. Munzner, 2014). Placing the square-glyphs in a grid on a map adds the spatial dimension and avoids symbol overlap. Square-glyph visualizations can be employed for areal and network-related data representations (Figure 1).

The gridded square-glyphs on maps allow exploring up to four data dimensions or datasets simultaneously and in relation to space, to extract single values, to detect structures and patterns and to evaluate the interrelationships of multiple spatial attributes. The square-glyphs may be useful, e.g. in evaluating location quality in planning, social vulnerability (Strode et al., 2020), or walkability (Bleisch & Hollenstein, 2018) (Figure 1). It was previously shown that user performance with glyphs decreases with increasing data

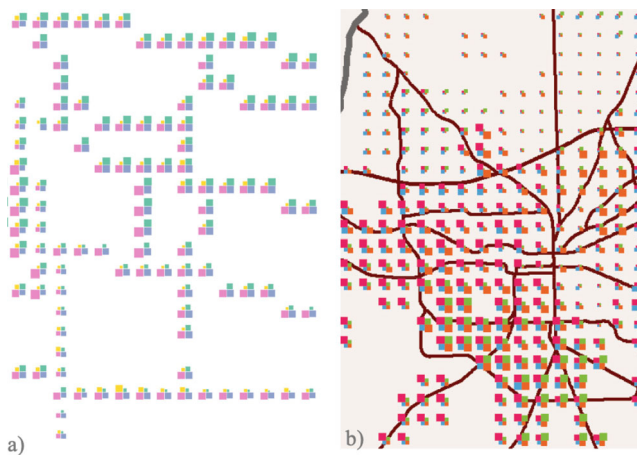


Figure 1. (a) Square-glyphs of four data categories along a gridded road network of a city (Bleisch & Hollenstein, 2018, p. 5) (b) Classed square-glyph visualization of four data categories across the gridded settlement area of Florida (Strode et al., 2020).

points and dimensions (Fuchs et al., 2017; Ward, 2008). However, it has not been evaluated (a) whether users can efficiently (i.e. fast) and effectively (i.e. accurately) read and interpret spatial attributes represented by square-glyphs; and (b) how useful square-glyphs are in analyzing the represented data to answer complex questions considering the multidimensionality and spatial relationships of the represented data.

In this paper, we explore how well can people read and interpret the differences and changes in values of different data dimensions for answering complex multidimensional questions. Specifically, we compare how effectively and efficiently the represented values can be recognized and interpreted when two (2D) or four (4D) data dimensions are present.

2. Related work

2.1 Glyph visualization

Visualization tools and processes offer powerful means for exploring large and complex data sets (Ward, 2002). Ward's mantra to visual information search, 'I know it when I see it', echoes Shneiderman's (1996) notion that we often do not know what we are looking for, but visualizations of the data can leverage the human perceptual and cognitive capabilities to extract meaning, focus attention, and uncover structures and patterns.

Glyph visualizations can help amplify perceptual processes and lead to quick and effective interpretation of multiple data dimensions. Compared to other symbols, displays of multiple glyphs allow for clear visualization of patterns that span multiple data dimensions, as long as they are within human visual perception capabilities (Ward, 2002). Data glyphs were used as early as the 1950s to represent multiple data dimensions: e.g. Metroglyphs (Anderson, 1957), which used line length to encode data. Over the years, many different glyph variations have been introduced, which fit specific data types or solve defined tasks more effectively (Monmonier, 1990). Glyphs can attract more attention than signs and thus stimulate more cognitive activity than other forms of visual representations (Borgo et al., 2013).

2.2 Square-glyph

The square-glyphs were proposed to enable visual analysis of walkability indices related to a linear road network in an urban environment (Bleisch & Hollenstein, 2018), but may be generalized to many other high-dimensional data visualization scenarios. Bleisch and Hollenstein's (2018) original proposition was linked to the need for a concurrent visualization of multidimensional data, as well as its relationship to the road network, which required a new approach, thus the gridded square-glyphs were developed (Figure 1a). The square-glyphs later found another application in a research project regarding socially vulnerable populations across a populated area (Strode et al., 2020). Strode et al.'s (2020) goal was to analyze different categories of data and their relationships to each other as efficiently as possible (Figure 1b). Before employing the square-glyphs, they created separate univariate area plots for each data set. They argue that since the relationships between the individual aspects of hazards are complex, the square-glyph visualization is able to better convey these relationships on a single map representation. Feedback from users in Strode et al.'s (2020) study indicated that it was possible for the participants

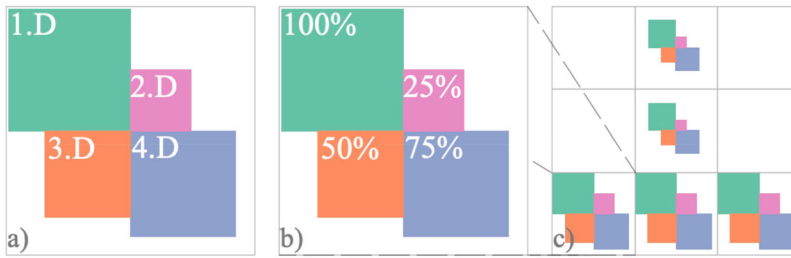


Figure 2. (a) Different color hues distinguish the data categories (b) the area of the squares represents the data values (e.g. normalised index values). In (b) green = largest value, pink = smallest value (c) the glyphs are placed in a grid and represent the values of up to four data categories in relation to each grid cell.

to successfully identify regions with clusters of values, as well as the dependence of the data dimensions represented in these regions.

2.2.1 Visual channels

Representing data values of different data dimensions, a glyph should be designed so that each value is mapped on a different display channel (Ware, 2019). For the square-glyphs, the square area is used to represent the value (Figure 2b), and the square color hue assigns the data dimension or category (Figure 2a). Thus, the variables can be read independently of each other. The recognition of quantities over sizes is possible, even if not too precise. Ware (2019) shows that the representation of quantities by area sizes is effective when relative quantities are to be assessed at a glance. Even if lengths can be judged more accurately than areas, the advantage of the area is that it can convey larger variations. However, visual separability of square size may vary along the value scale (Li et al., 2010).

2.2.2 Placement

The square-glyphs are placed within single grid cells of a defined spatial dimension. The squares of each category are arranged around the grid cell center and expand, according to the data values, to a maximum of $\frac{1}{4}$ of the grid cell. Thus, the square-glyph represents up to four data categories in relation to the cell location (Figure 2c). If less than four data values are represented concurrently, some of the grid quarters stay empty. Figure 3 shows

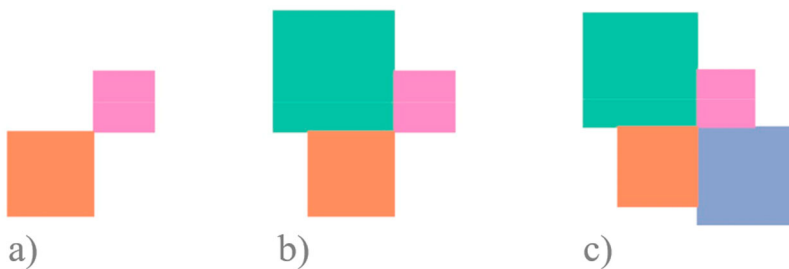


Figure 3. Representations of square-glyphs with two (a), three (b) and four (c) data categories (adapted from: Bleisch & Hollenstein, 2018).

square-glyphs with two (a), three (b) and four (c) data categories. The square sizes can represent continuous quantitative data. However, the size restriction within the grid cells may require the values of each category to be normalised or classed (i.e. representing quantile membership). Glyph placement is a major visual stimulus and can be used to convey information about the data. Ward (2002) gives an overview of placement strategies of multivariate glyphs.

2.2.3 Grid- and glyph-shape

The principle of arrangement and size variation is also possible with other shapes than squares. Bleisch and Hollenstein (2018) use triangles instead of squares for three data dimensions. Circles are not used because of the remaining white space between circles. Gröbe and Burghardt (2020) test different grid shapes (e.g. honeycomb grid) in addition to variations of glyph shapes. They conclude that squares and rectangular grid cells make the best use of space.

2.2.4 Context and density

A defined grid usually covers the whole geographic space of interest. However, not all grid cells may show a glyph. Figure 1a shows square-glyphs that reproduce the linear structure of an underlying road network to which they relate. Figure 1b uses the square-glyphs to represent data values related to a continuous spatial area. In area-based placement, glyphs tend to be close to each other. In network-based arrangements the number of empty cells depends on the density of the network. Strode et al. (2020) test display types in which they reduce the number of the area-based placement glyphs to achieve a better overview and focus on an area of interest. To do so they remove glyphs in less important regions (e.g. unpopulated areas), or do not display data outside a relevant analysis threshold.

2.2.5 Data set size

Depending on the task and data availability, the grid in which the square-glyphs are placed can be chosen at higher or lower cell size resolution. Large datasets may require small glyphs to represent them, making pattern recognition difficult (Fuchs et al., 2017; Ward, 2008). Borgo et al. (2013) describe that it becomes increasingly difficult to detect, classify, or measure features and relationships as the number of variables increases. But with more variables there is an opportunity to visually identify connections, trends, and relationships between different topics (Borgo et al., 2013). Visualization is about seeing patterns in the data rather than reading absolute quantity (Ware, 2019). Nevertheless, Ward (2008) notes that glyphs are primarily suited for qualitative analysis of datasets of modest size. Fuchs et al. (2017) come to the same conclusion when comparing the results of different user studies on glyphs. They show that mapping more dimensions has a negative impact on the performance of data glyphs.

2.3 Glyph user studies

Fuchs et al. (2017) provide a systematic overview of 64 papers investigating variations of glyph designs. Although different user studies have investigated different glyph designs and their variations, there are only a few glyph applications (Gröbe & Burghardt, 2017,

2020) that can be compared to the visual design features and application purpose of the square-glyph. Gröbe and Burghardt (2020) developed a glyph visualization method to distinguish categories and overall distributions of location-related point data sets of social media tweets. As with the square-glyph, their glyph diagrams are laid out in grid cells over a surface with a geographical background. Instead of squares, they use pie charts and area distribution (categorical value distribution of tweet language), as well as chart size (number of tweets), to reflect summary point data values for each grid cell. Through a user study, they showed that estimating values and percentages using their so-called micro-diagrams works better than using a dot map. However, the method only allows a rough determination of the data set values. Using the micro-diagrams, dominant categories could be better estimated than with dot maps.

Fuchs et al. (2017) indicate that most glyph studies did not test the number of different data dimensions represented. Only a few glyph studies, e.g. Fuchs et al. (2013), Fuchs et al. (2014) or Wilkinson (1982), have tested glyphs using different dimensions and used the number of illustrated data dimensions in the glyphs as a study factor. The results of these studies show that different designs are affected by the number of data dimensions represented to varying degrees. Overall, user performance with glyphs decreases as more data dimensions are represented in a glyph.

To add to the limited knowledge about the influence of the number of data dimensions represented through glyphs, and glyphs' overall usefulness in visually analysing multivariate spatial data, we designed a user study to evaluate square-glyphs. We compare the performance of glyph displays representing two (2D) or four (4D) data dimensions for value comparison and estimation tasks.

3. Experiment

3.1 Experiment design

We investigate whether the number of glyph components represented, i.e. the number of data dimensions (2D vs. 4D), influences efficiency and effectiveness of the participants in reading and comparing values from glyphs. Thus, our main independent variable is the number of data dimensions. We compare participants' performance in terms of speed (efficiency) and accuracy (effectiveness) as dependent variables. For control purposes and to obtain repeated observations under varying conditions, we observed participant performance with 2D and 4D glyphs as they executed two different tasks (T1 & T2), under three data conditions each. Data conditions were varied with respect to network layouts (resulting in different glyph placement), value spectra and between-component average value differences.

Based on the evidence presented in Section 2 on the effectiveness of the visual channels size and color (e.g. Borgo et al., 2013; Ware, 2019), we hypothesize that participants can distinguish between larger and smaller data ranges in the square-glyphs; and given a size legend, participants can estimate square component size of glyphs with a precision of around 10%. We assume that user performance decreases the closer the data values are to each other, and that with equal value differences, square-glyphs with low value ranges are more powerful than those with high value ranges. Furthermore, because it has been shown that user performance with glyphs decreases with increasing data points and

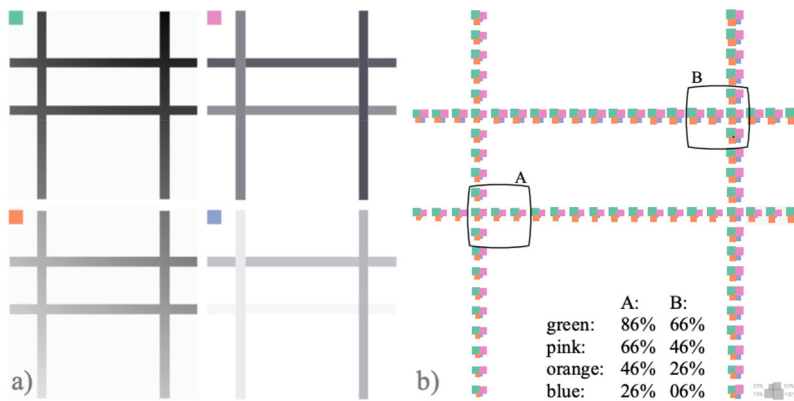


Figure 4. Example square-glyph generation from luminance value diagrams (pink and blue: continuous data, green and orange: discrete data). (a) four data sets along roads are represented as luminance values (b) the four data sets of (a) are gridded and represented as colored square-glyphs.

dimensions (e.g. Borgo et al., 2013; Fuchs et al., 2017; or Ward, 2008); we hypothesize that participants will read 2D square-glyphs more effectively and efficiently than 4D square-glyphs. 3D square-glyphs are not included in the study as their asymmetries (cf. Figure 3b) may introduce perceptual issues not accounted for.

3.2 Participants

The participants for the experiment were recruited through direct emailing and forwarding. A total of 24 expert participants (20 male, four female, 13 under and 11 over the age of 30) completed the study. 11 from the field of geomatics and 13 from the field of urban planning. These participants are assumed to have some knowledge and experience of visual analysis of spatial data due to their professional backgrounds. Participants were students and employees of private and public companies and organisations.

3.3 Materials: stimuli and data

We created a set of synthetic data to systematically vary parameters while creating glyphs with different data dimensions that reflect realistic and complex data combinations. In practice, the square-glyphs could represent data such as topography, noise, light or infrastructure availability, e.g. schools or shopping facilities. Each data set may relate to different spatial contexts, i.e. a network or an area. Depending on the nature of the data, values may be constant over several grid units, change progressively or change discretely at boundaries of the underlying spatial units. To represent different data types, we created different value distributions (e.g. continuously changing values along a road or discrete changes at intersections) using gray-scale luminance (Lerch & Bleisch, 2019). Figure 4a shows different gray scale value distributions along the lines of a network. The gray scale values are then aggregated into 2D or 4D square-glyphs (Figure 4b: 4D square-glyphs based on the values in Figure 4a) using a square grid structure. The grid

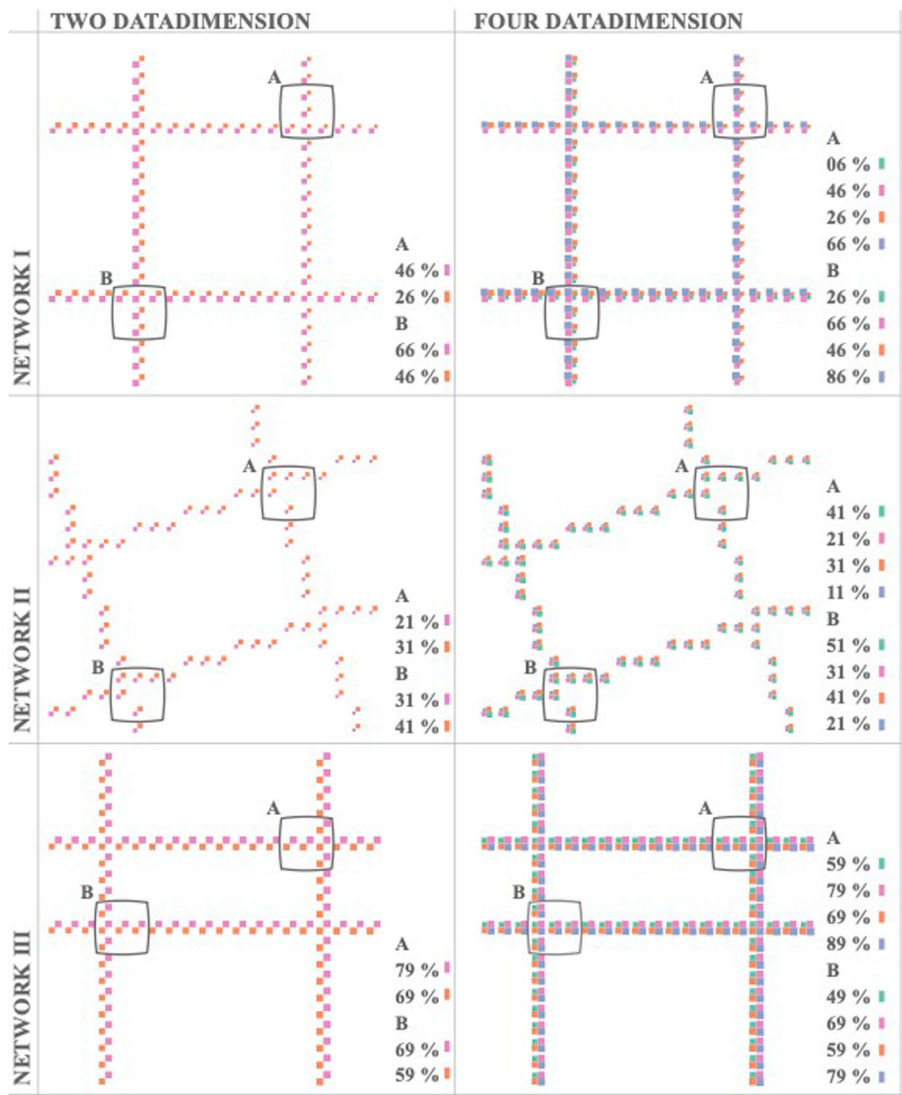


Figure 5. Example of data and layout conditions as used in the study. The three network variations (rows) with different data value, the two focus area frames A and B, and the representation as 2D and 4D square-glyphs (columns).

size defines the aggregation of the data. It was exploratively varied and finally set to allow for overviews as well as perception of individual square-glyphs.

The color choice for the four square-glyph components was made using the online tool Color-Brewer (Harrower & Brewer, 2003). The color hues vary (without order or conceptual ranking) but saturation and luminance are constant or nearly constant, in order to minimize effects of interaction between the perception of color and size (Ware, 2019) and to maximize for orthogonality of different attribute mappings (Borgo et al., 2013). The allocation of colors to the four components ensures that more similar hues are opposite each

other. Thus, the four data dimensions are easier to distinguish. Because of fixing saturation and luminance, the chosen color scheme neglects color vision deficiencies.

Using the described procedure, we generated three sets of synthetic data for each task type (six in total). The data sets vary with respect to network layout, value spectra and between-component value differences.

- (1) **Network layout variations** (Figure 5, rows): Firstly, we designed three variations of a simplified road network of four streets (Figure 5): two networks with horizontally and vertically running roads, and one network with diagonally running roads. Overlaying a square grid for data aggregation and glyph placement results in glyphs that are placed further apart for diagonally running roads. The placement of glyphs closer together or further apart could impact readability, as glyphs further apart lack a size reference through neighboring squares.
- (2) **Value spectra and between-component value differences:** One of the networks spans values from 3% to 88% (network I), the second network spans values from 9% to 52% (network II) and the third network spans values from 47% to 90% (network III) (Figure 5, rows). Two query frames (A & B) are placed on each network variation (Figure 5). For the tasks (cf. section 3.4) of type T1 (estimating average values for components within a frame), placing two query frames on three network variations yields six different tasks. For the tasks of type T2 (comparing A & B to a given set of values), this results in three different tasks.

Placing the query frames in different regions of the networks, results in the following per frame value spectra and between-component and -frame value differences: For the network with value spectrum 3–88%, average component sizes are 06, 26, 46, 66% units for frame A and 26, 46, 66, 86% for frame B. Thus, square-glyph component sizes differ on average 20 percentage points within a frame as well as between query frames. For the networks with value spectrum 9–52%, (average component sizes frame A: 11, 21, 31, 41%; average component sizes frame B: 21, 31, 41, 51%) and value spectrum 47–90% (average component sizes frame A: 59, 69, 79, 89%; average component sizes frame B: 49, 59, 69, 79%) square-glyph component sizes differ on average 10 percentage points within frames as well as between query frames on the same network.

Pairwise size difference between squares of the same component within a frame are between 0 and 5 or 0 and 3 percentage points. Thus, value variance within components per frame is less than half the variance between components and frames. Value variation per component inside a query frame takes two forms: For two components, values vary discretely, for the other two components values vary continuously. Placing query frames at intersections with lines of continuous and discrete value changes results in groups of three and two glyphs, each displaying a different value, depending on the axis they sample (Figure 4).

Finally, each of the three network variations per task was plotted twice, once displaying all four data dimensions and once displaying only two of the four data dimensions. This results in six pairs of T1 tasks and three pairs of T2 tasks with identical conditions except for the variation in data dimensions displayed. Specifically, in the 4D square-glyphs, the 2D glyphs are complemented with the components of lowest and highest values. For

all plots, this results in a configuration with the two middle values in diagonally opposite square quadrants and lowest and highest values in the remaining (opposite) quadrants. This is not ideal but ensures that the overall average value of glyphs in a query frame of a 2D plot is identical to the overall average value in the corresponding 4D representation. Also, a balance is kept between discretely and continuously varying components, by placing one of each on the diagonals of 2D and 4D glyphs.

3.4 Procedure

The study was approved by the MSE student research board and participants were informed about the purpose of the study, that participation is voluntary, and that their anonymity is guaranteed. The study was conducted online (i.e. not in a controlled laboratory environment), due to the restrictions imposed by the global Covid-19 pandemic. Participants completed the study unaccompanied, and on their own device. They were instructed to set aside 30 min to complete the tasks on a high-resolution computer screen. Starting the experiment, the participants were introduced to the principle of the square-glyph for potential application scenarios. This included showing them a display of square-glyphs on a background map to explain how the data dimensions of the square-glyphs may be related to geographic space even though the study tasks did not involve background maps. Then, the tasks and how to read the square-glyphs to answer the tasks were explained. Specifically, we designed two different task types (T1 and T2) to evaluate square-glyphs when either two or four components are present. The two task types are aimed at value ordering and value extraction in relation to the multidimensional nature of the square-glyph. In both task types, knowledge has to be extracted synoptically from groups of five glyphs at given locations within a display of a network comprising 76 glyphs. Initially, a third, more synoptic, task type was designed in which a group of glyphs with a given condition had to be found by exploring a display with 76 glyphs. Feedback from the pilot tests indicated that the task was very challenging and required concentration over a long period of time, that participants were not willing to invest. So, that task was excluded from the user study. Below we describe the two task types included in the study.

Task type 1 (T1): Average value estimation

For a defined group of five glyphs, an average value is to be estimated for each square-glyph component. To support exact value estimation a size legend is displayed at the bottom right of each test display. The legend consists of a single square-glyph with two or four dimensions labelled with their respective values. Continuous sliders are provided for the participants to set the estimated average value (0–100%) of components. Depending on the number of data dimensions, two or four sliders are present. The sliders are not subdivided, but the selected value is shown to the right. This task requires (a) a comparison of different glyph dimensions, and (b) the building of a rank order with respect to estimated average size which is expressed by the relative setting of slider handles. The estimation of average square size for each component is expressed by the absolute slider handle positions.

Task type 2 (T2): Location comparison

In the location comparison task, two groups of five adjacent glyphs each, are marked in a display and a set of sliders that define an average value for each data dimension are

given. The task is to estimate which of the two locations (A or B) is closer to the condition specified by the slider values. The slider values differ by 6 and 14 percentage points respectively from the average values of glyph components in the query frames for network type I, and by 3 and 7 percentage points for networks of type II or III. Again, a size legend is displayed at the bottom right of each test display. The legend consists of a single square-glyph with two or four dimensions labelled with their respective values. We assume, that this task involves checking if the relative order of any or both groups correspond with the slider values and then comparing each group based on their component values with the slider values to decide on the better match. To answer, the participants tick a box to indicate whether A or B is correct.

For both task types the target glyphs are located at network intersections and are delineated with a grey frame (Figure 4b, framed areas A and B). The frame is slightly convex to reduce its usefulness as a reference for value interpretation. In case of large neighboring data values, the convex frame may touch the squares outside the focus area.

Participants solved two test tasks (one of each task type T1 and T2). Then they completed the 18 experiment tasks (twelve T1 tasks and six T2 tasks) in randomized order. For each task, the completion time as well as the values from sliders and checkboxes were recorded.

The 18 data displays were prepared as described in section 3.3. In total, participants saw six different network representations (three for each task type). The placement of two query frames on each network yields six individual T1 tasks (presenting one frame at a time) and three T2 tasks (presenting both frames at a time). Duplicating these nine task plots and removing two of the four glyph components in the duplicates (reduction from 4D to 2D glyphs) results in six analogues 2D T1 tasks and three analogues 2D T2 tasks (18 in total).

The size of the square-glyph plots was set to 10 cm × 10 cm, so that the visualization is sufficiently large to read the square-glyphs and the task and sliders are visible on the same screen. The glyph display stays the same size in different browser window sizes but screen resolution was not controlled for. Participants were not tested for color vision deficiency. They could report qualitative feedback on the experiment at the end of the survey.

4. Results

We present the results per *task type* for each *dimension* (our main variable, i.e. 2D vs. 4D), taking *network type* into account. The results of statistical tests are detailed and further interesting exploratory post-hoc analyses reported. For interval-scale data, we check for normal distribution using a Shapiro–Wilk test, and accordingly use a parametric or non-parametric test for significance. Significant results ($p < .05$) are reported for test statistics.

4.1 T1 value estimation task

Our 24 participants solved 12 T1 tasks each, and set a total of 864 value sliders. Each participant solved six tasks in three 2D glyph plots setting 12 sliders; and six tasks in three 4D glyph representations, setting 24 sliders. The 12 tasks were solved independently of each other in random order.

4.1.1 Accuracy

Relative order: The set sliders were examined to see whether the size ranking of square-glyph components was correct with respect to the glyphs in the query frame. 253 out of 288 value estimation tasks (87.8%) were solved without error in size rank order (Figure 6).

To compare effectiveness of size ranking with respect to 2D and 4D glyphs, we looked at errors in the ranking of homologous glyph components in corresponding 2D and 4D figures. Using Wilcoxon signed rank tests, none of the three resulting 2D-4D comparisons yield a significant difference in task solving effectiveness for 2D vs. 4D glyph representations (lowest unadjusted $p = .041$ for the comparison of corresponding 2D and 4D representations of network III).

Absolute error: Pooling estimated values from all 864 sliders, an average absolute error of 8.3 percentage points with a standard deviation of 7.1 across participants is calculated. Looking at different configurations separately (2D vs. 4D under different network configurations), the mean absolute error varies only slightly (range), with the highest mean error for networks of type III (Figure 7).

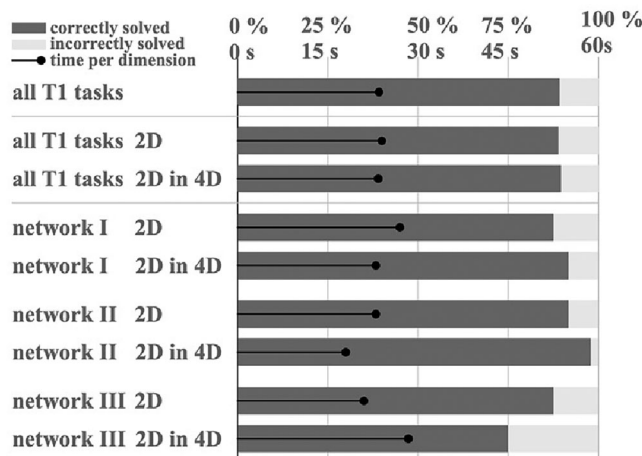


Figure 6. T1 results of ranking homologous glyph components in corresponding 2D and 4D conditions.

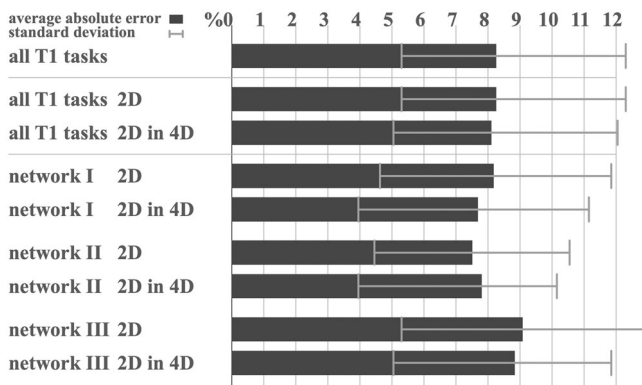


Figure 7. Results of T1 (average value estimation) mean and standard deviations of absolute errors.

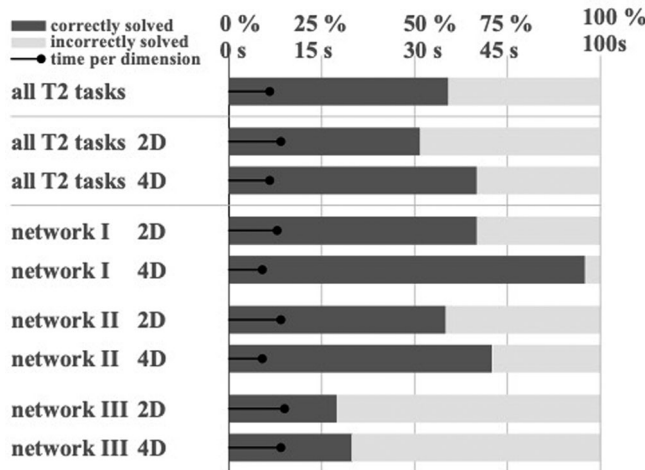


Figure 8. Results of T2 location comparison.

Using a Wilcoxon signed rank test with continuity correction or a paired t-test in case of normally distributed differences, we compared the sum of absolute errors for homologues glyph components in 2D and 4D plots per participant for each network type. None of the three comparisons yielded a significant difference in task solving effectiveness for 2D vs. 4D glyph representations (lowest unadjusted $p = .466$).

4.1.2 Task time

On average, the participants needed 47.3 s for setting the two sliders in 2D tasks and 93.8 s for setting the four sliders in 4D tasks (i.e. roughly twice the time for twice the number of sliders). Overall, it took participants 23.5 s to set a slider. On average, the participants took 23.7 s to set a slider in a 2D task, and 23.4 s to set a slider in a 4D task.

Using Wilcoxon signed rank tests with continuity correction to compare summed slider mean times per participant for 2D vs. 4D plots of each network type separately, no significant differences were found (lowest unadjusted $p = .493$).

4.2 T2 location comparison

Twenty-four participants conducted 144 (24×6) location comparison tasks. In each task, they compared two locations in one network to a set of target values.

4.2.1 Accuracy

Overall, 59% of type T2 tasks were solved correctly (Figure 8). We run a one-sample t-test on the proportions of correctly solved tasks per participant ($t = 2.1839$, $p = .039$) to check whether the binary answers of T2 differ significantly from a distribution with mean 50%, as would be expected by chance.

Across participants, 51.4% of the tasks in 2D glyph representations and 66.7% of the tasks in 4D glyph representations were solved correctly (Figure 8). Still, according to a

Wilcoxon signed rank test with continuity correction ($V = 27.5$, $p = .059$), that was calculated on the summarized scores for 2D and 4D tasks per participant, participants did not do significantly better, when solving the tasks in the 4D glyph representations.

Comparing the outcomes for 2D and 4D tasks for network types I, II, and III individually, the proportion of correctly solved tasks with 2D representations remains below the proportion of correctly solved tasks with 4D glyph representation consistently and diminishes from network I to network III with proportions markedly below 0.5 for network type III (Figure 8). Applying McNemar's tests, none of the three resulting 2D-4D comparisons reach a significant difference (lowest unadjusted $p = .023$ for the comparison of corresponding 2D and 4D representations of network I). Using a pairwise Wilcoxon signed rank test to compare all six configurations yielded significant differences for network I with 4D glyphs against network III with 2D glyphs (Bonferroni-adjusted $p = .001$) and network III with 4D glyphs (Bonferroni-adjusted $p = .002$) respectively.

4.2.2 Task time

Average task completion time was 28.6 s for 2D tasks and 42.1 s for 4D tasks. Normalized by the number of glyph dimensions, participants took 14.3s per dimension when solving the task in 2D plots and 10.5s per dimension in 4D plots. Task completion time slightly increases from network type I to III (Figure 8). However, task completion time for T2 tasks will likely not provide conclusive information with respect to the comparison of task solution efficiency for 2D versus 4D glyph plots as not all glyph components have to be evaluated to solve the task.

4.3 Open feedback

Qualitative feedback was reported by 19 of the 24 participants. 14 responses indicated that the quantitative assessment of the percentage values was perceived as difficult, demanding, strenuous or challenging. Eight participants perceived the assessment of the size ratios as easy. Eight feedbacks refer to the fact that value estimation is only possible with a size legend.

Two comments reported different task solving procedures. One mentioned that the value estimation was done via the area and the square length. The other described that the total squares were compared with the largest square of the group. Three comments mentioned that the square-glyphs with large value differences were found easier to interpret than those with small value differences. One feedback mentioned that the interpretation of the 2D glyphs was perceived as more complex than the interpretation of the 4D glyphs. Six comments referred to the color being difficult to discern in the 2D glyphs.

5. Discussion

We evaluated the influence of the number of represented data dimensions (2D vs. 4D) on average value estimation and comparison for square-glyph representations. Based on an average value estimation task (T1), solved by 24 participants, we find no significant difference in average value estimation accuracy for homologues components of five square-glyphs with two and four components. Also, time taken for average value estimation

per square-glyph component does not differ significantly for 2D and 4D glyph plots. This finding may point towards independence of the component-wise average value estimation and the number of glyph dimensions (for an ensemble of five glyphs and two vs. four components).

Results from task (T1) also indicate that generally, participants are successful in ordering glyph components with respect to size with no significant difference between equivalent displays of 2D and 4D glyphs.

When comparing two sets of five glyphs with respect to a set of target values (T2), participants solve the task of estimating the closer match consistently but not significantly better with 4D square-glyphs. However, overall success rate in solving this task is 59%, and the proportion of correctly solved tasks diminishes from network type I to network type III. In fact, results on network type III show a possible bias towards the wrong answer (correct answers remain markedly below 50%). Plots of network type III ask to estimate glyph component sizes in the upper value range (>47%). Task T2 was solved well only for 4D glyphs in network type I. Network I spans the largest average value range (06–86%) with a difference of 20 percentage points between glyph components (within and between glyph groups). Although other factors cannot be ruled out in this design, this outcome may be related to size range. The proportion of successfully solved T2 tasks with 4D glyph plots of network I (and also to a lesser extent with 4D plots of network II), may be associated to the fact, that the 4D glyph plot spans a wider value range (than the corresponding 2D plot), thus providing stronger extrema. With respect to the lower end of the size scale this may simplify solving the task (Li et al., 2010). Also, this could explain a potential bias towards the wrong answer (opting for the relatively larger squares) in task T2 on network III and would be inline with the known underestimation of larger area values (Stevens 1975 in Munzner, 2014). Additionally, for square-glyphs placed in a grid structure, users may consider the side length of the remaining white space to help estimation of square sizes close to 100%. Using a convex frame thus could amplify an underestimation of square sizes in the upper value range.

Finally, with 11 people from the field of geomatics and 13 people from the field of spatial planning, the results stem from a balance of participants with background knowledge from disciplines that may use square-glyphs. However, due to the gender imbalance (four female vs. 20 male participants) the results are dominated by the male perspective. The qualitative feedback of participants about the readability and interpretability of 2D and 4D square-glyphs, and different value ranges, supports the results and observations.

6. Conclusion

We report an experiment with 24 participants evaluating their performance of using square-glyphs with two and four dimensions and different value compositions. In particular, we quantitatively measured the relative and absolute accuracy and efficiency and collected the users' qualitative perception of analyzing the square-glyphs.

The results show that the value composition of five adjacent square-glyphs is readable in different data densities with mean absolute errors <10%. We find that the individual data dimensions of the square-glyphs are read without significant difference in effectiveness regardless of whether two or four dimensions are present. This indicates that the square-glyphs are suitable for representing value differences of up to four

dimensions. However, some participants report that exact value identification is challenging and demanding. And the overall rate of successful comparative value estimation across two groups of five glyphs is only 59%.

This experiment did not look into the limits of value levels and data densities that will make the identification of values or size classifications difficult or impossible. Future work could also test whether the placement or density of square-glyphs in a grid has an influence on the interpretation and readability or whether more complex tasks can be accomplished, such as evaluating groups of square-glyphs that fulfill a certain condition, trend or pattern within a larger structure of square-glyphs.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Gianna Daniela Müller completed her MSc in Engineering, specialization Geoinformation technology at FHNW and has a background in urban planning. She has worked as a research assistant at the FHNW Institute Geomatics, in the research group Geovisualization and Visual Analytics. Currently, she is employed by the city of Zurich as a 3D GIS specialist.

Daria Hollenstein is a research associate at the FHNW Institute Geomatics in the Geovisualization and Visual Analytics research group.

Arzu Çöltekin is the director of the Institute of Interactive Technologies at FHNW and a professor of Human-Computer Interaction and Extended Reality (HCI & XR). Arzu obtained her PhD from Helsinki University of Technology (now Aalto) and worked at the University of Zurich for more than a decade before taking on her current role. Besides HCI and XR, her research interests include visualization, visuospatial cognition, visual illusions as well as aging and technology.

Susanne Bleisch is a professor of Geovisualization and Visual Analytics at FHNW and is the program head of the MSc in Engineering masters at the School of Architecture, Civil Engineering, and Geomatics. She has a background in Geomatics Engineering and obtained her PhD in Geographic Information Science from City University London before doing a PostDoc at the University of Melbourne. Her research interests include the integration of suitable analytical visualizations in the data analysis process in different application areas such as urban planning or building energy loss.

References

- Anderson, E. (1957). A semigraphical method for the analysis of complex problems. *Proceedings of the National Academy of Sciences*, 43(10), 923–927. <https://doi.org/10.1073/pnas.43.10.923>
- Bleisch, S., & Hollenstein, D. (2018). *Exploring multivariate representations of indices along linear geographic features. Proceedings of the 2017 international cartographic conference, Washington D.C.* (pp. 1–5). <https://doi.org/10.5194/ica-proc-1-11-2018>
- Borgo, R., Kehrer, J., Chung, D. H. S., Maguire, E., Laramée, R. S., Hauser, H., Ward, M., & Chen, M. (2013). Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. *Eurographics 2013 – State of the Art Reports*, 25.
- Fuchs, J., Fischer, F., Mansmann, F., Bertini, E., & Isenberg, P. (2013). Evaluation of alternative glyph designs for time series data in a small multiple setting. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- Fuchs, J., Isenberg, P., Bezerianos, A., Fischer, F., & Bertini, E. (2014). The influence of contour on similarity perception of star glyphs. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 2251. <https://doi.org/10.1109/TVCG.2014.2346426>

- Fuchs, J., Isenberg, P., Bezerianos, A., & Keim, D. (2017). A systematic review of experimental studies on data glyphs. *IEEE Transactions on Visualization and Computer Graphics*, 23(7), 1863–1879. <https://doi.org/10.1109/TVCG.2016.2549018>
- Gröbe, M., & Burghardt, D. (2017). Micro diagrams: A multi-scale approach for mapping large categorised point datasets. *AGILE*.
- Gröbe, M., & Burghardt, D. (2020). Micro diagrams: Visualization of categorical point data from location-based social media. *Cartography and Geographic Information Science*, 47(4), 305–320. <https://doi.org/10.1080/15230406.2020.1733438>
- Harrower, M., & Brewer, C. A. (2003). Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1), 27–37. <https://doi.org/10.1179/000870403235002042>
- Lerch, T., & Bleisch, S. (2019). A visual approach to creating multivariate geovisualization test data. *Abstracts of the ICA*, <https://doi.org/10.5194/ica-abs-1-211-2019>
- Li, J., Martens, J. B., & van Wijk, J. J. (2010). A model of symbol size discrimination in scatterplots. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2553–2562. <https://doi.org/10.1145/1753326.1753714>
- Monmonier, M. (1990). Strategies for the visualization of geographic time-series data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 27(1), 30–45. <https://doi.org/10.3138/U558-H737-6577-8U31>
- Munzner, T. (2014). *Visualization analysis and design*. CRC Press.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–343. <https://doi.org/10.1109/VL.1996.545307>
- Strode, G., Mesev, V., Bleisch, S., Ziewitz, K., Reed, F., & Morgan, J. D. (2020). Exploratory bivariate and multivariate geovisualizations of a social vulnerability index. *Cartographic Perspectives*, 95, 5–25.
- Ward, M. O. (2002). A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4), 194–210. <https://doi.org/10.1057/PALGRAVE.IVS.9500025>
- Ward, M. O. (2008). Multivariate data glyphs: Principles and practice. In *Handbook of data visualization* (pp. 179–198). Springer.
- Ware, C. (2019). *Information visualization perception for design* (4th ed.). Elsevier, Imprint Morgan Kaufmann.
- Wilkinson, L. (1982). An experimental evaluation of multivariate graphical point representations. *Proceedings of the 1982 Conference on Human Factors in Computing Systems - CHI '82*, 202–209. <https://doi.org/10.1145/800049.801781>