A Systematic Review of Experimental Studies on Data Glyphs

Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, and Daniel Keim

Abstract—We systematically reviewed 64 user-study papers on data glyphs to help researchers and practitioners gain an informed understanding of tradeoffs in the glyph design space. The glyphs we consider are individual representations of multi-dimensional data points, often meant to be shown in small-multiple settings. Over the past 60 years many different glyph designs were proposed and many of these designs have been subjected to perceptual or comparative evaluations. Yet, a systematic overview of the types of glyphs and design variations tested, the tasks under which they were analyzed, or even the study goals and results does not yet exist. In this paper we provide such an overview by systematically sampling and tabulating the literature on data glyph studies, listing their designs, questions, data, and tasks. In addition we present a concise overview of the types of glyphs and their design characteristics analyzed by researchers in the past, and a synthesis of the study results. Based on our meta analysis of all results we further contribute a set of design implications and a discussion on open research directions.

Index Terms—Survey, glyphs, quantitative evaluation, glyph design

---

1 INTRODUCTION

REPRESENTING multi-dimensional data is a common task in data visualization and, thus, a multitude of techniques has been developed [1]. Data glyphs are one such technique, in which single data points are encoded individually by assigning their dimensions to one or more marks and their visual variables. Data glyphs have a long history, going back to the 1950s, with metroglyphs being one of the first designs using line length to encode data [2]. A somewhat infamous, and thus well researched, example of data glyphs are Chernoff faces [3] which encode data values in 2D facial features such as the length of the nose or the orientation of eyebrows. Star Glyphs are another type of glyph-based encoding that has received research attention [4] and has been used in various applications [5], [6], [7].

Over the years, many different glyph variations were introduced to better fit certain data types, or to solve specific tasks more effectively. Yet, while many designs have already been explored, the mapping possibilities of data dimensions to visual glyph encodings is nearly endless [1] and many more designs are certainly imaginable. This flexibility allows designers to come up with new and innovative glyph representations for specific data, tasks, or contexts. However, without any guidance, this freedom and large design space can become overwhelming. Knowledge of when and which types of designs work best or are preferred by viewers, could aid designers and practitioners in creating new designs or in selecting among existing ones. Yet, while many user studies in the literature have investigated different data glyph designs and their variations, a systematic overview of these studies and what they pertain to, is so far missing.

With this paper we contribute such a systematic overview of the user-study literature on data glyphs, focusing on the analysis of 64 papers with quantitative controlled studies. In contrast to their qualitative counterpart, controlled experiments are more easily comparable and summarizable, as they test concrete hypotheses regarding design choices and isolate factors in the glyph designs [8]. We focus on data glyphs that: use different visual channels to encode multiple attribute dimensions, that can be independently spatially arranged, and that can vary in size (see Section 3.1). We categorize the studies we found according to a number of criteria that are meant to help researchers and practitioners choose amongst the most relevant literature to read, and ultimately to make informed choices about glyph use, design, and potential future studies. These criteria include glyph types (see Fig. 1 for an overview), presentation settings, datasets, tasks, and study goals. We extend this characterization with a summary of study outcomes to help practitioners select the most appropriate data glyphs according to different criteria like visual design, data density, or task. By tabulating existing studies according to these criteria we provide a novel user-study-centric view on the design space and study outcomes, and contribute a discussion on open research areas that will help to further our understanding of data glyphs.

2 RELATED WORK

In this section we provide an overview of glyph-related survey- or meta-papers that, similar to ours, have attempted to structure existing glyphs according to various criteria. Since a large number of glyph designs have been proposed in the past, a tabulation of existing individual designs is beyond the scope of this section.

- *J. Fuchs and D. Keim are with the University of Konstanz, Konstanz 78464, Germany. E-mail: fuchs@dbvis.inf.uni-konstanz.de, keim@uni-konstanz.de.*
- *P. Isenberg is with Inria, Paris, France. E-mail: petra.isenberg@inria.fr.*
- *A. Bezerianos is with Univ Paris Sud, CNRS & Inria, Paris, France. E-mail: anastasiab@gmail.com.*
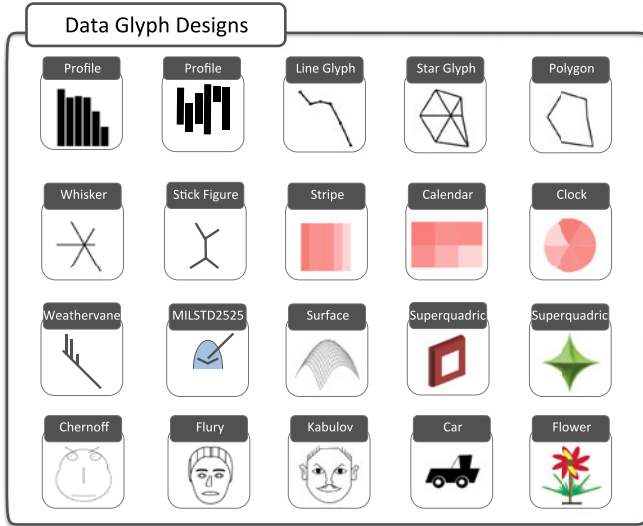
Fig. 1. *Data glyphs:* A selection of the different data glyph designs used in the quantitative experiments we analyzed.

Ward [9], [10] was one of the first to provide a structure of the glyph design space by classifying different layout possibilities into data-driven and structure-driven layouts. In a data-driven arrangement spatial position is determined by data: this can either be the raw data used as in a scatterplot, or a projection-based approach such as PCA. A structure-driven layout makes use of relations between the data points to calculate a layout. For example, hierarchical information can be used to lay out glyphs in a tree structure. Ward extended this work with a categorization of the visual characteristics of data glyphs. He structured data glyphs based on their mapping of data to visual attributes into three different classes: a *many-to-one mapping* where each data dimension is mapped to the same visual variable (e.g., profile glyph [11]); a *one-to-one mapping* showing each data attribute with a different visual variable (e.g., Chernoff faces [3]); and a *one-to-many mapping* representing the data dimensions redundantly with many different visual attributes (e.g., compound glyph [12]). In this paper, we use this categorization to structure our own categorization of data glyph user-studies.

In contract, Chung et al. [13] proposed a categorization based on the visual channels used to represent the data and the spatial dimensionality of the glyph (2D, 2.5D, and 3D). The authors also discussed critical design aspects and guidelines for glyph visualizations, such as the normalization of data input for each dimension, the use of redundant mappings, and the visual orthogonality of different glyph components to ensure best performance. Since some of these guidelines cannot be followed for a high number of dimensions, designers have to choose between few single complex glyph designs, or many simple designs. Additionally, they suggested using halos to limit the negative effect of overplotting. In our survey we extend this list of guidelines based on our review of experimental results and provide further open research questions.

An extensive survey on data glyphs was presented by Borgo et al. [14]. The authors cover different glyph representations and propose guidelines for designing data glyphs based on a collection of design principles in the literature. While Borgo et al. also include several empirical studies in their survey, their focus is on design study papers showing the applicability of data glyphs to different data sets and tasks. In contrast to this work, we provide an overview of performance assessments from quantitative user studies.

A more data-specific survey on glyphs in the medical domain was presented by Ropinski et al. [15]. The authors classified glyph-based visualizations for medical data into two groups: pre-attentively and attentively identifiable glyph designs. Based on this grouping the authors further derive design guidelines for developing glyphs for this domain, but provide no additional empirical results from user studies.

While there is no systematic assessment of glyph user-studies that we know of, some researchers have categorized subsets of the study design space. Nelson [16], for example, discusses the history of Chernoff faces [3] with its many variations such as the Flury-Rydwiel [17] or Kabulov faces [18]. She also discusses studies investigating performance changes for different data types or visual variations. We took this work as inspiration, but provide a much more comprehensive view on the study design space. Ware's [Ch. 5] [19] discussion on "Glyphs and Multivariate Discrete Data" is related to our work in that he categorizes two types of user study tasks for glyphs. He focuses on tasks designed to find out which display dimensions are perceived holistically (integral) or perceived separately (separable): restricted classification tasks and speeded classification tasks. Among others, we include both types of tasks in our discussion based on slightly different terminology [20]: similarity search tasks (related to restricted classification) and lookup tasks (similar to speeded classification).

In summary, while there are several overview papers on data glyphs, we know of no prior work, such as ours, that provides a systematic review of glyph user studies, study designs, and results.

3 METHODOLOGY

For our systematic review we focused on sampling user studies in which participants performed controlled, quantitatively measured tasks with data glyphs. These quantitative measurements could (but did not have to) be accompanied by a subjective assessment of the tested glyphs (e.g., according to aesthetics, confidence, etc.) Next, we categorized the found studies according to the criteria discussed in this section.

3.1 Data Glyphs: Sampling Characteristics

During our previous work on data glyphs we noticed that definitions of data glyphs vary in the literature. Yet, we needed to base our systematic literature search on a more formal, grounded understanding of existing data glyph definitions. We, thus, extracted definitions from survey- and meta-papers on glyphs, several books on visualization, many papers in the literature that used the word "glyph," and interviewed several visualization experts on their understanding of the term glyph.

We found that in early years the term "glyph" was often used as a synonym for the metroglyph [11], [21] but that with increasing numbers of different glyph designs published, more abstract definitions of a data glyph emerged. To compare how researchers have defined the term, we extracted descriptive keywords from published definitions, and

TABLE 1
Overview of Defining Glyph Characteristics
Mentioned in the Literature

References	Data glyph characteristics		
	small/ compact	encodes different attribute dimensions	uses different visual channels
[14]	X	X	X
[24]	X	X	-
[25]	X	X	-
[26]	-	X	X
[27]	-	X	X
[28]	-	X	X
[29]	-	X	X
[1]	-	X	X
[15]	-	X	X
[19]	-	X	X
[10]	-	X	-
[30]	-	-	X
[31]	-	-	X

summarized in which publications they appeared in Table 1. While Table 1 is certainly not exhaustive, it serves to show the wide variety of ways researchers think of, and define, data glyphs, an observation corroborated recently by Munzner [1].

After a careful assessment of these definitions, we approached the paper sampling as follows: first, we included all user studies that used the word "glyph". In a second step we excluded all papers which used the term glyph as a synonym for simple data marks (e.g., [22], [23]), such as points in a scatterplot. We chose this exclusion criteria because we wanted to focus our assessment on glyphs that encode multiple attribute dimensions at once. This notion of a data glyph as a multi-dimensional encoding also aligns with the majority of historical data glyph definitions. We, thus, chose studies in which glyphs fit the following general definition:

Data glyphs are data-driven visual entities, which make use of different visual channels to encode multiple attribute dimensions. They can be independently spatially arranged and can vary in size.

After this filtering we categorized the tested glyph designs according to the main descriptive criteria we found in the literature: visual channels, encoded dimensions, and presentation size (see Table 1).

3.2 Paper Sampling and Collection

To find relevant papers for our review we used a snowball sampling technique in which we first searched for the keyword "glyph" in the title, abstract, and keywords in the the ACM digital library (leading to 80 potential results), the IEEE Xplore digital library (leading to 255 potential results), the EG digital library (leading to 66 potential results), and the DBLP computer science bibliography (leading to 134 potential results).

In a next step we excluded papers that did not include at least one user study with quantitative measures or did not study glyphs that fit our definition. This filtering step removed 505 of the 535 candidate papers, leaving 30 relevant papers for our survey. From this initial set of papers we recursively scanned references for further user studies about data glyphs. Using this approach we collected 64 papers from the visualization literature as well as work from statistics and psychology.

3.3 Analyzed Study Characteristics

In the design of any quantitative user study several characteristics are important: the tasks to be performed, the collected measures, the presentation of the stimuli (glyphs), the size and type of data visualized, the general presentation setting, and the study goals (or main research questions) [8]. We categorized the 64 study papers using these characteristics as explained in more detail next.

3.3.1 Glyph Types and Data Encoding

We used Ward's data mapping taxonomy [10] to distinguish between glyphs using many-to-one and one-to-one mappings (see Section 2). The rows and columns of Table 3 give an overview of this categorization. We only found two occurrences of Ward's third group: one-to-many mappings. Thus, we do not highlight this group as a category in our result table. The two studies we found ([32] and [33]) are, however, discussed throughout the paper.

Since the many-to-one group encodes multiple data point dimensions using the same visual variable, we further split this group into categories based on the visual variables used: position/length, color saturation, and orientation/angle (see Table 3). We also distinguished whether or not a linear or circular layout was chosen to lay out the dimensions.

The category of one-to-one mapping was structured slightly differently as it includes a wide variety of design choices. As we mostly found facial glyph representations or three-dimensional designs, the result table includes these two categories: Faces and 3D Glyphs. A third category on car glyphs was added, since in one paper [34] faces were compared against unique car glyph representations. Car glyphs are abstract two dimensional representations of vehicles, which use unique characteristics (size of the trunk or hood) to encode data.

We additionally found 12 studies that tested unique glyph designs that were not compared to alternative representations: PlanningLines [35], weather vanes [36], [37], shapes [38], [39], roses [40], themes [41], arrows [42], Motifs [43], [44], flowers [45], and MILSTD2525 glyphs [46]. Rather they were either compared against textual information, tested on varying backgrounds (changes in the topological level of detail), or against different types of visualizations. Since they were not compared to other designs in the table, we positioned them slightly apart in the "One-to-One Mapping" category.

3.3.2 Glyph Presentation Setting

For the examined studies, we categorized how many glyphs were presented to a viewer on the screen: individual glyphs, multiple glyphs of fixed number, or multiple glyphs of varying numbers. In the category of multiple glyphs we further noted how the glyphs were arranged on the screen, as grids, scatterplots, node-link diagrams, on geographic maps, or other layouts.

3.3.3 Datasets

The glyphs used in the studies all encoded either multi-dimensional data of a general nature, or time-series data. Additionally, we noted how many dimensions a glyph encoded. The number of dimensions is related to the visual

complexity of a glyph. Independent of data type and density we further recorded whether the data was synthetically created, or if real data was used in the study.

3.3.4   Tasks and Measures

Important for understanding any study results is the nature of the task participants had to perform. We group tasks in broad categories, differentiating between tasks involving the glyph as a whole (synoptic tasks [20]) and tasks where participants had to focus on single specific characteristics of a glyph (elementary tasks [20]). An elementary task is typically a lookup task during which participants focus on single dimensions of a glyph and read individual values.

We further subdivided synoptic tasks into three categories: 1) *visual search* where participants had to find a glyph differing from others, or tell whether a specific glyph is present or not; 2) *similarity search* where participants had to compare the overall structure of glyphs and group similar representations; and 3) *trend detection tasks* where participants had to keep track of the development of data values across dimensions.

3.3.5   Study Goals

We found three different general study goals: 1) a comparison of *various glyph designs* according to their performance and a ranking of designs based on it; 2) a comparison of different *variations* of a single glyph, to detect visual features improving a specific glyph design; and 3) a comparison of *single glyphs versus data tables*, to motivate the use of these visual objects over textual representations.

3.3.6   Study Results

We summarized study outcomes on a high level, reporting findings on the impact of presentation settings, number of data points and dimensions on the tested glyphs. We further report overall ranking of different glyph types, offering explanations to seemingly contradictory results across studies. We do not enter into detail on findings regarding variations of a single glyph type. Our goal is to provide researchers and practitioners with a better grasp of the overall picture of the performance of different glyphs, and to point to individual papers for detailed study results.

4   RESULTS: STATE-OF-THE-ART IN GLYPH EVALUATION

In this section we discuss the findings from our systematic review based on the characteristics discussed in the previous section. A summary of the results is presented in individual tables and in highlighted paragraphs throughout the section. Many study descriptions did not include all information needed for our characterization and subsequently our counts do not always add up to 64–the total number of papers examined.

4.1   Study Goals

We found three higher-level study goals—all related to different types of comparisons: a) comparison of glyph designs, b) comparison of glyph variations, and c) comparison of glyphs with data tables or text. As can be seen in the
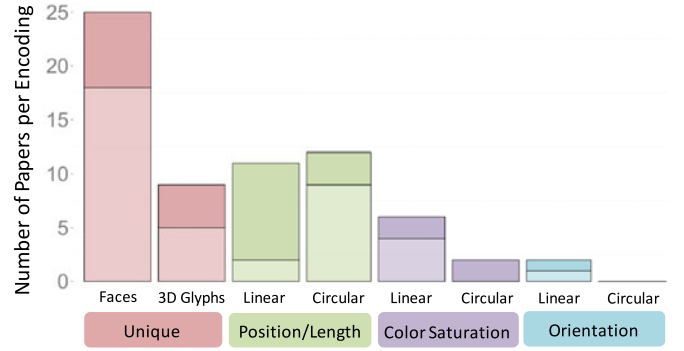

Fig. 2. *Statistics:* Ratio of papers evaluating different visual encodings (distinguished by color). Low saturation indicates experiments evaluating design variations of this encoding, and high saturation other experiments (e.g., comparisons to other encodings).

diagonal of Table 3,[1] the majority of studies (39/64, 60.94 percent) tested case b) or c). Design variations within a glyph category were more frequently tested against each other (32/39, 82.05 percent) than glyphs versus a common data table or text description (7/39, 17.95 percent). The latter group was most often used to motivate the use of visuals over text descriptions or data tables [21], [64], [84], [85], [86].

To measure participant performance all studies but one [44] recorded accuracy scores, additionally 65.63% measured completion time (42/64), and 29.69 percent collected qualitative feedback (19/64) as well. It is interesting to note that participants' preferences did not always match with their performance [56], [58], [60], [90]. Therefore, a preferred design was not always a guarantee for a good user performance.

Summary: We found similar study goals across many experiments, yet varied were factors like number of data points and dimensions, task, or glyph design. These variations make individual study outcomes hard to compare. Thus, we will discuss the individual factors in the following sections before discussing the study outcomes in Section 4.6.

4.2   Glyph Types and Data Encoding

Table 3 summarizes evaluated glyph types and their encodings based on Ward's data mapping taxonomy outlined in Section 3. The table is meant to be read like a matrix. The intersections of rows and columns show which glyph types and encodings a particular study compared against each other. The diagonal (top left to bottom right) of the table contains references to studies that tested design variations of the same glyph category, or an evaluation of one specific design against plain text or data tables (marked with a * in the table). Empty cells indicate new research possibilities.

Fig. 2 shows that face glyphs were evaluated most frequently (39.06 percent), followed by glyphs with position/length encodings (linear: 17.19 percent, circular: 18.75 percent), and 3D glyph designs (14.06 percent). We note that from the studies involving position/length encodings or 3D glyphs (27 in total), 8 were in fact compared to faces (Table 3). Color (linear: 9.38 percent, circular: 3.13 percent) and orientation encodings (linear: 3.13 percent, circular: 0 percent) have received little research attention.

---

1. The flower [45] and theme [41] glyphs also fit into this category, however, due to their visual encoding they are not represented on the diagonal.
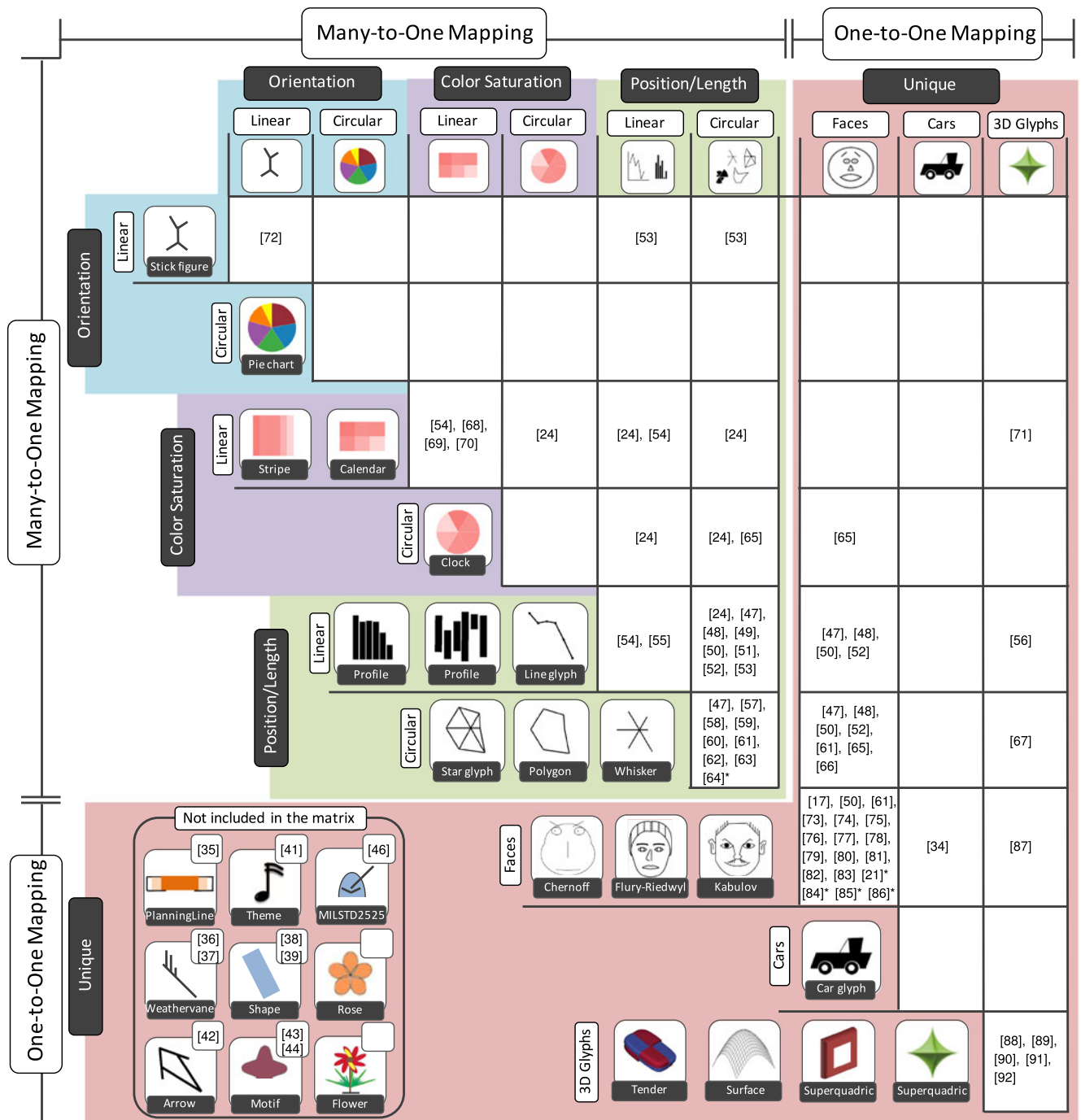
Fig. 3. *Glyph design table:* Columns represent the different categories of glyph encodings, replicated in rows with glyph examples for each category. Additionally, color is used to visually separate the different categories. References refer to articles in our study bibliography that compare glyph variations from the respective row and column. Studies placed in the diagonal evaluate either variations of the same glyph type, or comparisons of the glyph with data tables (starred *). Note that papers can fall in multiple cells. Since PlanningLines [35], weather vanes [36], [37], shapes [38], [39], roses [40], themes [41], arrows [42], Motifs [43], [44], flowers [45], and MILSTD2525 glyphs [46] use a unique encoding and are not compared to other glyphs, we positioned them slightly apart in the "One-to-One Mapping" category.

The high number–28.13 percent–of user studies on face variations, stands out compared to studies that only focus on other variations, e.g., circular position/length encodings (14.06 percent), 3D glyphs (7.81 percent), linear color (6.25 percent), or linear orientation encoding (1.56 percent). A possible reason for this imbalance are the many ways one can design faces and their data mappings (e.g., Chernoff faces [3], Rydwiel-Flury faces [17], Kabulov faces [18]).

We found only two studies [54], [55] that compared different linear position/length design variations. This is an interesting research gap given that profile glyphs that use this encoding are well established in practice (i.e., sparklines [93], profiles [11]). In general, we only found three main categories of visual variables used to encode data in glyphs with many-to-one mappings (Position/Length, Color, Orientation). Almost all glyph designs in these studies mapped

TABLE 2
Datasets: This Table Illustrates Detailed Information about the Real Datasets used in the Experiments

Dataset	Availability
Anthropometrical data about twins [17], [76]	[95]
Patients rated by psychiatrists [21]	Minnesota Multiphasic Personality Inventory
Medical data [32]	unknown
Medical images [56]	unknown
Cars dataset [34]	http://davis.wpi.edu/xmdv/datasets/cars.html
Project plans [35]	unknown
MM5 weather information [36]	unknown
Weather information [37]	NCEP forecast model
Weather information [42]	Operational Regional Atmospheric Prediction System
Modified U.S. census data [38]	unknown
Financial data [40]	Investment in education USA (2008)
Financial data [84]	Wall Street Journal Index (1974 and 1975)
Financial data [85]	Standard and Poor's firm list (1974 and 1975)
Classical music data [41]	unknown
k data [43]	Lostpedia wiki edits
Network data [44]	10 best ranked movies (IMDb)
Google search results [45]	http://www.google.de
Marathon runners [50]	unknown
Power plant statistics [53]	unknown
Audio information [62]	One laptop per child sound library
Biological data [70]	unknown
Economic variables [77]	U.S. Department of commerce & labor
Tensor data [92]	DTI dataset

TABLE 3
Presentation Setting: This Table Distinguishes between the Number of Data Points Shown to the Participants During the Studies and the used Layout

	Layout	References
Single		[48][49][50][54][64][91]
	Text	[55]
Multiple Glyphs	Grid	[17][21][24][33][39][45][47][51][52][53] [57][58][59][61][66][67][71][72][73][74] [75][76][77][78][80][84][85][86][87][89]
	Geo map	[36][37][38][40][60][65][68][69]
	Scatterplot	[34][88][90]
	Other	[32][35][56]
	Node-link	[41][70]
Varying	Grid	[46][62][79][82][83][92]
	Node-link	[43][44][63]
	Geo map	[42][81]

Color is used to better distinguish between the different categories.

For the 46 studies that tested a fixed number of multiple glyphs at a time, we found five types of layouts. The most frequent was a common small-multiples grid (65.22 percent), followed by geographic maps (17.39 percent), scatterplots (6.52 percent), node-link diagrams (4.35 percent), and other layouts (6.52 percent) like different 3D environments (see Fig. 4).

The goal of most of the studies with varying number of glyphs was to investigate changes in performance when increasing the number of visible data points in grid layouts [79], [82], [83], geographic maps [42], [81], and node-link diagrams [63]. The amount of glyphs visible to participants changed from 5–50 [79]; 5–15 [82], [83]; 6–18 [46]; 9–23 [81]; 4–300 [42]; and 30–48 [63]. In all seven studies participants were affected negatively by an increasing number of data points, as we discuss in Section 4.6.2. In the studies conducted by Aigner et al. [35], Dunne et al. [43], Cayli et al. [44], and Zhang et al. [92] the varying number of data points was not treated as a factor in the analysis.

In seven studies (10.94 percent) a single data point was shown to participants at a time [48], [49], [50], [54], [55], [64], [91]. These studies tried to control all parameters and avoid

quantitative information to visual variables. The only exception was Lee et al.'s work [66] which compares star glyphs, faces, and 2D projected data points using bivariate data. Here bivariate information, however, was still mapped either to the length of the whiskers (star glyph) or to different face characteristics.

Summary: Faces and circular profiles have been investigated in detail, in contrast to color value and orientation encodings on glyphs that only few studies investigated. Surprisingly, we found only two studies comparing different variations of linear profiles.

4.3   Glyph Presentation Settings

Presentation settings can be characterized by the number of glyphs presented to viewers, as well as by how the glyphs are layed out in space. We identified three types of studies when considering the number of glyphs presented (Table 3): those that presented only individual glyphs to the viewers (7/64, 10.94 percent), those that presented a fixed number of more than one glyph at a time (46/64, 71.88 percent), and those in which the number of presented glyphs varied but was always higher than one (11/64, 17.19 percent). Seven papers did not report the exact number of glyphs represented on the screen: [35], [37], [41], [43], [44], [45], [56].
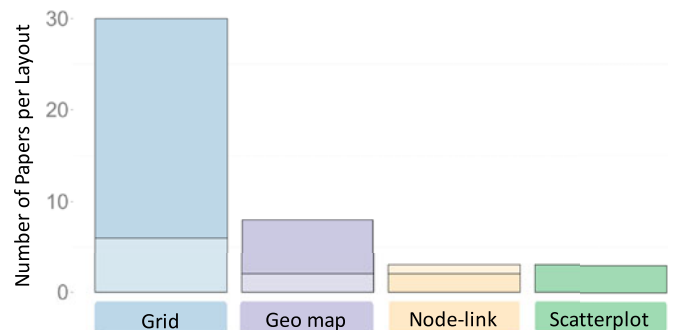

Fig. 4. *Statistics:* Ratio of papers evaluating different glyph layouts (distinguished by color). Low saturation indicates experiments evaluating a varying number of data glyphs, and high saturation a fixed number of data glyphs.

TABLE 4
Number of Dimensions: This Table Illustrates the Different Data Dimension Densities used in the Studies

Number of Dimensions	References
2 & 3 Dimensions	[32][35][36][37][38][39][40][44][48] [51][60][67][71][89][91][92][96]
4 & 5 Dimensions	[21][41][42][45][46][49][50][64][65] [68][69][72][80][81][82][83][88][90]
6 & 7 Dimensions	[77][85]
8 & 9 Dimensions	[33][34][53][59][61][73][79]
10 - 15 Dimensions	[54][66][70][75][84][87]
17 - 20 Dimensions	[17][52][57][74][76][78][86]
Varying	[24][47][58][63]

Color is used to better distinguish between the different categories.

possible confounding factors, so as to better reason about changes in performance when modifying specific aspects of the same design [91], or when comparing it with other representations [48], [49], [50], [54], [64], or layouts [55].

Regardless of whether a fixed or changing number of glyphs was tested, the vast majority of studies (56.25 percent) arranged glyphs in a grid layout, followed by geographic arrangements (15.63 percent), node-link diagrams (7.81 percent), and scatterplots (4.69 percent) (see Fig. 4). The choice of a grid layout for quantitative studies is understandable. Grids can help to avoid confounding factors in visual search, comparison, or classification tasks. For example, the information provided by a background, e.g., an underlying geographic map, may influence the perception of glyphs. The background color, for example may influence the perception color hues [94], while topology (e.g., rivers, mountains, land borders) may act as grouping enclosures or as reference structures for reading data values of glyphs. We only found a single study [36] that examined the influence of reading data glyphs with different geographic backgrounds; and one [39] that studied how the reading of a glyph is affected by the presence of other glyphs around it. We discuss their results in Section 4.6.1.

Summary: Only a small number of user studies varied the amount of data glyphs as a study factor. Most studies were conducted with a fixed number of glyphs arranged in a grid layout. Surprisingly, only four papers investigated the influence of different background information and layout on reading data glyphs [36], [39], [55], [62].

4.4 Datasets & Number of Dimensions

The number of data dimensions tested can help us compare results across studies, and inform us of the imagined use-case setting for data glyphs. Only four studies (6.25 percent) used the number of dimensions itself as a study factor and thus varied between glyphs with different dimension counts [24], [47], [58], [63]. The remaining 60 studies tested glyphs with various fixed numbers of dimensions. Of these, 44 tested less than 10 dimensions. An overview of different dimensionality settings is provided in Table 4. Three papers did not report about the number of dimensions encoded by the glyph designs [43], [55], [62].

In the vast majority of studies (54/64, 84.38 percent) glyphs encoded general multi-dimensional data, both real and synthetic. Eight studies tested glyphs encoding time-series data (8/64, 12.5 percent) and in two experiments [43], [44] glyphs were used to represent network topologies.

Only a small number of studies (24/64, 37.5 percent) used real data to investigate the performance of different glyph designs. The respective papers and real datasets can be found in Table 2. For the other experiments (41/64, 64.06 percent) the data was created synthetically.

Summary: Overall, most studies used synthetically created multi-dimensional data (41/64, 64.06 percent). The majority (44/64, 68.75 percent) of studies used glyphs encoding less than 10 dimensions.

4.5 Task Space

We used the Andrienko & Andrienko task taxonomy [20] to distinguish between two higher-level tasks as discussed in Section 3. Synoptic tasks (i.e., similarity search, visual search, trend detection) were the most common type of task used in the studies (44/64, 68.75 percent). This is perhaps not surprising as glyphs are often meant to provide quick overviews over a large number of varying multi-dimensional data points—and the use of synoptic tasks may reflect the authors' desire to test glyphs in a realistic use context.

TABLE 5
Data and Tasks: Most Studies were Conducted using a Lookup or Similarity Search Task with Multi-dimensional Data

Data Type	Task Description				
	Elementary Task		Synoptic Task		
	Lookup	3D Navigation (distance calculation)	Trend Detection	Similarity Search	Visual Search
Multi-dimensional	[21], [34], [36], [37], [38], [40], [42], [45], [46], [48], [49], [50], [51], [53], [60], [65], [72], [79], [88], [90], [91]	[32]	[38], [84], [85]	[17], [33], [45], [47], [52], [57], [58], [59], [61], [63], [64], [66], [73], [74], [75], [76], [77], [78], [80], [86], [87], [89], [92]	[37], [39], [40], [56], [65], [67], [71], [81], [82], [83]
Time-series data	[24], [35], [54], [55]		[24], [55], [68], [69]		[54], [62], [70]

As shown in Table 5, we found the following classes of synoptic tasks: similarity search (23/44, 52.27%), followed by visual search (14/44, 31.82 percent), and trend detection tasks (7/44, 15.91 percent). An example of a similarity search task can be found in two studies by Klippel et al.: Using a visualization tool showing 81 glyphs each representing one car, participants had to group these glyphs into different categories based on their attributes [33], [59].

In contrast to these synoptic tasks, 26/64 studies (40.63 percent) used elementary tasks, i.e., lookup (25/26, 96.15 percent) and 3D distance calculation (1/26, 3.85 percent). These studies focused on more perception-related questions such as the reading accuracy for visual variables used to encode a data value. In these studies, participants did not focus on reading the entire shape of the glyph, but on single glyph characteristics. For example in the user study conducted by MacGregor and Slovic [50] participants had to read the completion time of 48 marathon runners from bar chart glyphs, faces, and star glyphs. Faces performed best, followed by bar chart glyphs and star glyphs.

Summary: Most studies used a similarity search or a direct lookup task to measure the performance of glyph designs.

4.6 Study Outcomes

While we cannot discuss the study results individually for all 64 papers, we collected higher-level observations on study outcomes. Results on the study of factors such as number of dimensions and datapoints tested, is consistent across experiments. Nevertheless, when it comes to a general ranking, experimental results apply to a study's specific setting and should be generalized with caution. We discuss these results next.

4.6.1 Influence of Background Information and Layout

Understanding the influence of layout strategies or additional context information is crucial since data glyphs can be arranged in various different ways and settings. Four studies investigated the influence of positioning or background information on the performance of data glyphs [36], [39], [55], [62].

A common setting for data glyphs was the positioning in scatterplots, or projections from a high dimensional dataset to a two dimensional space. Frisson et al. used a visual search task to examine the benefits of a two dimensional projection compared to a grid layout used in small multiple settings [62]. Performance was lower for the two dimensional projection, since after projection, some data glyphs ended up overlapping each other, which caused a loss of information making it difficult to detect the stimulus. In a follow-up study, the authors added a proximity grid [97] as an additional layout to the study setting. Results indicated that participants performed best in a visual search task when using the proximity grid.

Glyphs were also used in textual documents to communicate statistical data not only with words but visually. Sparklines are a famous example of such small visual representations [93], which are usually positioned in the reading direction next to the statistics (e.g., on the right hand side). To backup this design decision, Goffin et al. conducted a

user study to compare different layout possibilities of glyphs within sentences [55]. Surprisingly, there was no significant effect on accuracy or reading performance for the different layouts. However, participants preferred the glyph being positioned above the words.

The influence of reading data glyphs with different geographic backgrounds was investigated in only one study conducted by Martin [36]. He measured the performance of participants working with weather vane glyphs while varying the underlying geographic map. Surprisingly, his results indicated the background had no influence on the performance of reading data-glyphs. However, the glyphs in his study were arranged in a grid on top of a map, and not according to their geographic position. Using different glyph designs or an irregular layout may, nevertheless, influence their performance.

Healey and Enns conducted an experiment to compare the interaction of different visual features in the surroundings of the glyph stimulus for a visual search task [39]. Results indicated that color variations due to the presence of other glyphs in the neighborhood of the stimulus glyph, caused a significant interference effect when participants had to judge heights of glyphs or density patterns. However, different densities in the surroundings of the stimulus or heights of neighboring glyphs had no effect on the detection of colored glyphs.

Summary: The influence of background and layout on reading data glyphs has so far received little research attention. The limited evidence from this work suggests that the background and neighborhood of a glyph did not affect glyph readability. Nevertheless more work is needed to determine the perceptual difficulties of reading glyphs depending on their background and layout.
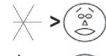
4.6.2 Influence of Number of Data Points

Seven studies varied the number of visible data points as a factor. The glyphs used in these experiments were either faces [79], [81], [82], [83], unique glyph designs (i.e., MILSTD2525 [46], arrow glyphs [42]), or star glyphs [63].

For the studies involving face glyphs, participants had to perform visual search tasks and find a certain stimulus in a growing set of data points. The researchers tested whether pre-attentive identification was possible, in which case search time would not have been seriously impacted by increasing the number of glyphs. Yet, in all studies the performance dropped with an increasing number of data points independent from the mapping of data to face characteristics. Based on this outcome Siva and co-authors concluded that participants performed a serial search and were not able to pre-attentively identify the stimulus [82], [83]. Therefore, the perception of abstract data glyph faces compared to human faces was shown to be different. This is an interesting finding, which lessens the basic motivation for using abstract faces. However, researchers could also show that a redundant visual mapping of data to face characteristics improved the performance [79].

Summary: Increasing the number of data points negatively affects search within a set of data glyphs, indicating that they—even face glyphs—cannot be read pre-attentively.

TABLE 6
Studies and their Result Rankings:
Faces versus Circular Profiles

Elementary Task	Synoptic Task
[profile] > [face] [48]	[face] > [profile] [47], [52], [61]
[star] > [face] [65]	[star] > [face] [65]
[face] > [polygon] [50]	[polygon] > [face] [66]

TABLE 7
Studies and their Result Rankings:
Faces versus Linear Profiles

Elementary Task	Synoptic Task
[face] > [linear profile] [50]	[face] > [linear profile] [47], [52]
[linear profile] > [face] [48]	

Conflicting results are marked with orange color.

4.6.3 Influence of Number of Dimensions

The results of studies varying the number of dimensions as a factor showed that different designs were impacted to different extents. In a study by Fuchs et al., for example, the performance of star glyphs dropped significantly in a lookup task when increasing the number of dimensions from 24 to 96, whereas the performance of line glyphs stayed stable [24].

Wilkinson also varied the number of dimensions to investigate changes in performance for different glyph representations. His results indicated that increasing the number of dimensions had no significant effect on the *ranking* of tested glyph designs [47], although there was a drop in performance overall.

However, it is interesting to note that even slight variations of a glyph design can be affected differently by the number of dimensions. Fuchs et al. tested the effect of increasing the number of dimensions on whisker glyphs (star glyphs without a contour line), traditional star glyphs and polygon variations. Although the performance dropped for all variations, whisker glyphs were affected the least [58].
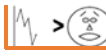
Summary: Increasing the number of dimensions negatively affects the performance of data glyphs [24], [47], [58], [63].

4.6.4 Influence of Tasks and Visual Encoding

The outcome of individual user studies often involved a ranking of data glyphs based on their performance in the study. These rankings were not always consistent for the same designs tested, and they changed, for example, based on tasks and details of the visual encoding. Table 6–Table 9 summarize the outcomes of the different experiments. The ">" symbol indicates that the glyph on the left outperforms the design on the right (either in terms of completion time or accuracy).

Seven studies compared faces against circular position/length encodings [47], [48], [50], [52], [61], [65], [66] (Table 6). In four, faces performed best [47], [50], [52], [61], while circular position/length encodings performed best in the remaining three [48], [65], [66]. These seemingly contradictory results are reconciled when we consider the tasks participants had to perform and how the glyphs were designed. In five of these studies the participants performed a synoptic task [47], [52], [61], [65], [66], in the other three a lookup task [48], [50], [65]. From the five synoptic task studies, in the three where faces performed best, the circular position/length encoding was a polygon (i.e., star glyph without whiskers, but only a contour) [47], [52], [61], while in the remaining two where faces performed worst the circular encoding was a star glyph with [66] and without contour line (i.e., whisker glyph) [65]. The remaining three
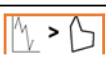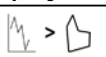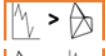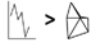
studies with lookup tasks also compared faces against polygons (with polygons performing best [48]), faces against star glyphs (with faces performing best [50]), and faces against whisker glyphs (with whisker glyphs performing best [65]). It seems that star glyphs compared to faces are more suitable for synoptic tasks. However, the whiskers glyph had the best performance independent from the underlying task. This finding has partially been confirmed for a similarity search [58] but not for lookup tasks.

Another example where glyph rankings change based on study characteristics can be found when comparing faces against linear profiles (Table 7). In three studies faces performed best [47], [50], [52], in the fourth study, profiles [48]. Again, the four studies used different tasks: lookup tasks [48], [50] and a similarity search task [47], [52]. When comparing the two lookup tasks the ranking of the two glyph designs is still different although they use a similar number of dimensions (4 [50] and 5 [48] dimensions), and just show one data point at a time. Yet, a major difference can be found when reading the task description more carefully. Although both tasks are a lookup task, participants had to either read a one-dimensional value [50] or detect when one dimension changes significantly compared to the other dimensions for a single data point [48].

When comparing linear and circular position/length encodings, we found glyph ranking differences in 8 studies [24], [47], [48], [49], [50], [51], [52], [53] (Table 8). In four, the linear design outperformed the radial [48], [50], [52], [53], while in two, circular designs were better [49], [51], and in the last one performance varied according to the underlying task [24]. However, only three out of these seven had a similar experimental setting with respect to design variations, presentation setting, number of dimensions and task [48], [49], [51]. These three all compare bar charts with polygons in a lookup task, using low dimensional data and presenting only one data point at a time. Surprisingly, the performance was still different: polygons ranked best in two of

TABLE 8
Studies and their Result Rankings:
Linear versus Circular Profiles

Elementary Task	Synoptic Task
[linear profile] > [profile] [48]	[linear profile] > [profile] [52]
[profile] > [linear profile] [49], [51]	
[linear profile] > [polygon] [24], [50], [53]	[linear profile] > [polygon] [24]
[polygon] > [linear profile] [24]	

Conflicting results are marked with orange color.

TABLE 9
Studies and their Result Rankings:
Color Saturati on versus Profiles

Conflicting results are marked with orange color.

them [49], [51] and bar charts performed best in the third [48]. Again, we have to look at the studies more carefully to come to a conclusion. In the two studies where polygons performed best, the bars in the bar charts were shown without a common baseline. This was not true for the third study were bar charts outperformed the polygons. We assume that a common baseline increases the performance of the linear profiles, a finding which is proposed as a design guideline from a study by Fuchs et al. [24]. However, a user study comparing linear profiles with and without a common baseline has, to the best of our knowledge, not yet been conducted.

Additionally, it is interesting to note that there were changes in performance depending on the kind of elementary task. For reading exact data values linear profiles outperformed star glyphs, however, when reading the position of an attribute dimension (e.g., a certain point in time for time-series data) star glyphs ranked first. [24]

Data glyph designs using color saturation to encode data values have not received much attention. We only found two papers, which report on results from quantitative experiments comparing these glyphs against alternative representations [24], [65] (Table 9).

For overview visualizations focusing on the overall appearance of a glyph, color value encodings were not found to be effective. In three out of four user studies participants performed better using a position encoding (i.e., linear and circular profiles) in synoptic tasks. In the fourth experiment the color encodings were visually enhanced to help participants solve certain tasks and outperformed linear profiles. Only faces performed worse. However, it is more difficult to draw conclusions for elementary tasks. When pursuing a direct lookup task (e.g., reading data values) radial color value encodings have outperformed star glyphs and faces [24], [65]. Whisker glyphs on the other hand have been shown to be as accurate as color value encodings but more efficient [65]. However, linear profiles were most accurate and, therefore, the best choice for direct lookup tasks [24].

Summary: Study results differed based on individual factors like number of dimensions, task, number of data points, or slight variations to the designs. Our summary tables can be considered as a performance overview pinpointing to relevant literature.

4.6.5 Influence of Metaphoric Glyph Design

One goal of information visualization is to present the underlying data in a way that can be easily understood by users. Thus, researchers have tried to improve intuitive understandability of visualizations, by using metaphors when mapping data to visual representations. One such example can be found for weather forecasts. In such a scenario weather status is communicated with small icons on top of a geographic map. These icons are metaphoric representations of the real environment to facilitate their understanding. Small cloud icons represent cloudy areas, rain drops encode rainy areas, and little suns illustrate sunshine in specific regions.

While data glyphs are different from icons [14], the general concept of representing the underlying data using metaphors can also be applied here. Since the visual representation of a glyph is data driven the idea is not to use a different glyph design for each individual data point (like in the weather forecast example), but to use certain glyph characteristics to display the data while being consistent with the metaphor.

However, it is not clear whether such metaphor-based representations are better than more abstract ones. Siirtola has attempted to provide an answer to this problem by introducing metaphoric glyph designs and comparing them with more abstract ones [34]. In his experiment he visualized car related data with abstract face representations, and with metaphoric car glyphs [34]. Car glyphs were created by mapping data to parts of the glyph with related meaning. For example the attribute horsepower was mapped to the size of the engine of the car, which is metaphorically reflected in a bigger hood. In his user study participants had to answer car related questions when working with either faces or car glyphs. The metaphor helped the participants in understanding the data. As a result, they performed better when working with car glyphs compared to faces.

Li et al. [40] provided another example where metaphors were used. In their quantitative experiment they compared RoseShape glyphs against abstract polygons to visualize multi-dimensional data about the education level in the US. The glyphs were positioned on top of a geographic map and participants had to either read data values or search for certain characteristics. Results suggest that participants were more accurate and more confident of their answers when working with the metaphoric designs.

In a study conducted by Flury and Riedwyl, data collected about monozygotic and dizygotic twins, such as their height or weight was mapped to two types of face glyphs [17]. Using abstract face representations (i.e., Chernoff faces) or more realistic faces (i.e., Flury Riedwyl faces) participants had to look at a glyph for each twin and rate whether or not the two glyphs showed data about monozygotic twins. The results indicated that participants were more accurate when working with the more realistic faces.

Jacob [21] gave another example where he tested the performance of a single metaphoric glyph design. He displayed data from patients having a certain psychological condition (e.g., depression, paranoia etc.) using faces. The abstract faces were created to show facial expressions resembling those of the human faces of the patients. Participants in his study had to judge which face corresponds to which behavior without being trained or knowing the patients. The results indicated

that people were able to name the correct psychological illness without knowing the mapping criteria of data to face representations.

Metaphors may help to explain the results obtained in a study conducted by Fuchs et al. [24]. The researchers ran a quantitative study using time-series data. Participants had to locate specific points in time using glyphs with either a linear dimension layout (e.g., sparklines) or a radial arrangement (e.g., star glyphs). Surprisingly, participants were more accurate when working with circular glyphs. This is interesting since the visual variable position (used in linear layouts) is considered more accurate compared to orientation (used in circular glyphs) [98]. However, participants argued that they were reminded of a clock when working with radial glyph designs, which facilitated locating certain points in time.

Summary: A small number of previous studies suggest that metaphors may help to better understand the underlying data.

4.6.6 Summary

While we found and reported on 64 papers, the vast design space of data glyphs and the possibility to test only a limited set of factors in a controlled user study makes it difficult to recommend a single best-of glyph design. Glyph performance depends on many different factors, such as the task used, the number of data points, or slight variations to the designs used across studies. Our analysis in Section 4.6.4 presents a summary of rankings from the articles we analyzed, and discusses how these factors can explain seemingly contradictory results.

We were able to draw general conclusions when it comes to number of dimensions and glyphs. Some study results indicate that increasing the number of data dimensions affects the performance of glyph designs negatively [24], [47], [58], [63] with position encodings (linear and circular profiles) being more robust compared to color encodings in high-density situations [24]. As with the number of dimensions, there is evidence that performance drops with increasing the amount of visible glyphs on the screen [42], [46], [63], [79], [81], [82], [83]. This seems like a logical conclusion due to the required additional effort in visual search involving a higher number of entities. In addition, a small number of past studies indicate that metaphoric glyph designs increase performance.

Finally, it has to be noted that our analysis was made difficult by a lack of standard for reporting study details on glyphs. For example specific information (e.g., stimuli size, viewing distance, number of visible data points, etc.), that could shed light on differences across experiments, were often missing.

5 IMPLICATIONS FOR DATA GLYPH DESIGN

In this section we abstract and summarize the study outcomes in the form of design considerations. We indicate the number of papers supporting each design consideration in brackets to indicate the extent of generalizability, and order them based on this support.

Do not expect glyphs to be perceived pre-attentively (7). All seven studies considering the scalability as a factor come to the conclusion that the performance drops significantly when increasing the number of data points. However,

these studies only included faces, arrows, star glyphs and MILSTD2525 glyphs.

Faces have possible value (5). Many experiments with different study designs have been conducted comparing faces against other data glyph design. Based on the underlying study setting the ranking of the tested designs was different. However, in at least five studies faces outperformed the alternatives [47], [50], [52], [61], [87].

Consider metaphoric designs (5). Five studies investigated the usefulness of metaphoric data glyph designs. The results indicate that metaphors helped participants in analyzing data. However, the type of metaphors and type of data are closely related, making the past studies somewhat limited. This a fruitful area for further research.

Linear data glyph designs profit from a similar baseline (3). In three studies comparing linear against circular profiles, linear designs were most effective when the dimensions where aligned to a common baseline. However, this finding has not been directly investigated by an experimental comparison of linear profiles with and without a common baseline.

In linear designs color encodings should be avoided, whereas they seem to be beneficial for circular layouts (3). Since only four studies have been conducted using color saturation to encode data values, this guideline has to be considered with caution [24], [65]. It seems that in circular layouts with position/length encodings (i.e., star glyps) the mental rotation required for comparisons affects participants more strongly than color comparisons that can be conducted without a common axis. For linear designs the ranking of visual variables from Cleveland and McGill seems to hold [98].

Pick line glyphs for lookup tasks with a large number of dimensions (2): Studies suggest that data glyph designs are influenced negatively by an increasing number of dimensions. However, the line glyph seems to be affected the least. For similarity search tasks one study found the whiskers glyph to be most effective in a high-dimensional setting.

6 DISCUSSION AND OPEN RESEARCH AREAS

In this section we identify and discuss directions for future research based on our analysis. The proposed research directions are ordered roughly according to their scope.

6.1 Types of User Studies

Even though we focused on user studies with quantitative components for this paper, we found only a few qualitative studies that considered how glyphs are used in practice within real applications. One such exception is the experiment conducted by Sreng et al. [99] where participants used a 3D automotive assembly tool and answered questions about the perceived usefulness of the embedded glyphs. Although this study provided qualitative observations in the form of questionnaires, we can envision more elaborate field experiments and observational studies on real use of glyphs. Observers could thus gather information on how people use glyph-based visualizations in real contexts, for which tasks, and with what kind of results. Such studies could inform our understanding of how glyph-based applications are adopted and used in practice and could, thus, provide new insights on which to base design choices.

Summary: Adding qualitative evaluations observing analysts working with different glyph designs, datasets, and

tasks, would help to better understand the glyphs design space. In particular, information about subjective preferences and the applicability of specific glyph designs in practice would be useful. It would be interesting to capture which design analysts choose to solve which analysis task.

6.2 Data to be Tested

There are several pros and cons for choosing real versus synthetic datasets for a study. On the one hand, real data has the advantage that it can demonstrate which visual representation performs best in realistic situations, providing valuable results for analysts of this data. However, real data often contains unique characteristics (e.g., size, structure, number of dimensions), that make the results noisy and hard to generalize.

On the other hand, one may argue that synthetic data does not always represent a real world scenario or problem well (ecological validity), making results again hard to generalize. However, artificial data can be easily controlled and focused on answering specific questions. Additionally, possible confounding factors due to the underlying data are excluded (e.g., visual search time according to the number of data points).

Given the above pros and cons, it seems an interesting open research question to see how glyphs behave when they undergo study using both synthetic and real data, similar to the approach taken by Caban [56].

Summary: Running quantitative experiments, using both datasets from synthetic to real world and vice versa will enhance our knowledge on the behavior of data glyphs in different situations.

6.3 Study Tasks and Measures

In the majority of studies participants had to perform synoptic tasks (Table 5). This is not surprising given that glyphs are often used to provide quick overviews over a large number of multi-dimensional data points. Nevertheless, there are glyph designs (e.g., some 3D glyphs) that have not or rarely been looked at for synoptic tasks, an interesting topic for further study.

Although results from specific tasks, such as these synoptic ones, are valuable, a common visualization task is free exploration, insight generation and hypothesis forming. Inspired by recent work on insight based evaluation [100], it would be worthwhile to investigate the performance of different glyph designs in such contexts.

Summary: Adding exploration tasks or extracting insights from an unknown dataset are realistic real-world analysis tasks. They should, therefore, be added to the repertoire of user study tasks in glyph evaluation to further reason about the practical applicability of data glyphs.

6.4 Glyph Presentation Setting

A large number of studies presented glyphs as small multiples using a grid layout. There were no studies on glyphs nested inside treemaps, or other types of representations apart from maps, scatterplots, node-link diagrams, and two 3D representations in the medical domain. This is interesting, as it is not clear that grid layouts present the most commonly assumed usage context for glyphs. For example, in the area of

scientific visualization, glyphs are often used on 3D volumetric surfaces or to represent 2D flow fields in order to indicate data at specific sampling points. These glyphs are approximately uniformly spaced apart, but this relative spacing changes depending on the view's magnification factor, making them appear more or less densely packed together. There is very little to no guidance from controlled user studies on how this apparent density affects their performance.

Moreover, we know little about the influence of the background information on the performance of glyphs. Only one study investigated performance changes for glyph designs when placed on top of different geographic maps [36], and one other their performance close to neighboring glyphs [39]. Many questions remain unanswered, for example, we do not know if glyphs are perceived differently when arranged in uniform grids compared to other arrangements, such as treemaps, that vary their relative distance.

It is also unclear what effect the glyphs have on the understanding of the underlying visualization itself: for example, it would be interesting to investigate if rectangular treemaps are more effective compared to circular treemaps when adding glyph designs; or if people are distracted by the additional context information in the form of glyphs in 3D environments. There is certainly much space for further research.

Summary: Since data glyphs cannot only be positioned in small multiple grids, evaluating different arrangements of more complex layouts (i.e., treemaps, etc.), would help to better understand the influence of specific data glyph designs on the context and vice versa.

6.5 Glyph Types and Data Encodings

Understand redundant encodings: Using Ward's glyph design categorization [10], we found only two studies that used glyphs with a one-to-many mapping (i.e., a redundant encoding). Ware [19], however, discusses interesting perceptual study approaches to learn how dimension encodings can be separable or integral. A better understanding of how redundant encodings work together, and could enforce data reading, would prove beneficial to glyph design.

Study missing mappings: In Table 3 we refer to 50 of the 64 studies examined, having left out the two one-to-many mappings [32], [33] and the twelve that were not compared to other designs [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46]. Looking at the table there is still clearly an imbalance in what kind of data encodings have been comparatively tested. Many cells remain empty, and there are several sparsely populated ones. One of the first things to notice is that there is no single study on circular orientation encodings, although they are used in visualization applications: representatives of this category are the compound glyph used in network graphs [12], pie chart glyphs for analyzing multi-dimensional data (e.g., global material composition [101], or biological binding properties [102]), or as provided in visualization toolkits (e.g., JIT[2]). Perhaps this type of encoding is a-priori deemed inferior based on Cleveland and McGill's [98] work that ranks orientation low for quantitative data representation. Given past use of these encodings however, it is certainly worthwhile to confirm that Cleveland and McGill's ranking does hold for circular-

2. JavaScript InfoVis Toolkit http://philogb.github.io/jit/

TABLE 10
Overall Summary: This Table is a Collection of All Design Considerations, Open Research Gaps,
and Findings Proposed throughout the Paper

	Section	Summary
Study Settings	4.1	We found similar study goals across many experiments, yet varied were factors like number of data points and dimensions, task, or glyph design. These variations make individual study outcomes hard to compare.
	4.2	Faces and circular profiles have been investigated in detail, in contrast to color value and orientation encodings on glyphs that only few studies investigated. Surprisingly, we found only two studies comparing different variations of linear profiles.
	4.3	Only a small number of user studies varied the amount of data glyphs as a study factor. Most studies were conducted with a fixed number of glyphs arranged in a grid layout. Surprisingly, only four papers investigated the influence of different background information and layout on reading data glyphs [36], [39], [55], [62].
	4.4	Overall, most studies used synthetically created multi-dimensional data (41/64, 64.06%). The majority (44/64, 68.75%) of studies used glyphs encoding less than 10 dimensions.
	4.5	Most studies used a similarity search or a direct lookup task to measure the performance of glyph designs.
Study Outcomes	4.6.1	The influence of background and layout on reading data glyphs has so far received little research attention. The limited evidence from this work suggests that the background and neighborhood of a glyph did not affect glyph readability. Nevertheless more work is needed to determine the perceptual difficulties of reading glyphs depending on their background and layout.
	4.6.2	Increasing the number of data points negatively affects search within a set of data glyphs, indicating that they— even face glyphs—cannot be read pre-attentively.
	4.6.3	Increasing the number of dimensions negatively affects the performance of data glyphs [24], [47], [58], [63].
	4.6.4	Study results differed based on individual factors like number of dimensions, task, number of data points, or slight variations to the designs. Our summary tables can be considered as a performance overview pinpointing to relevant literature.
	4.6.5	A small number of previous studies suggest that metaphors may help to better understand the underlying data.
Open Research Gaps	6	Adding qualitative evaluations observing analysts working with different glyph designs, datasets, and tasks, would help to better understand the glyphs design space. In particular, information about subjective preferences and the applicability of specific glyph designs in practice would be useful. It would be interesting to capture which design analysts choose to solve which analysis task.
	6	Running quantitative experiments, using both datasets from synthetic to real world and vice versa will enhance our knowledge on the behavior of data glyphs in different situations.
	6	Adding exploration tasks or extracting insights from an unknown dataset are realistic real-world analysis tasks. They should, therefore, be added to the repertoire of user study tasks in glyph evaluation to further reason about the practical applicability of data glyphs.
	6	Since data glyphs cannot only be positioned in small multiple grids, evaluating different arrangements of more complex layouts (i.e., treemaps, etc.), would help to better understand the influence of specific data glyph designs on the context and vice versa.
	6	Quantitative user studies should be conducted to compare data glyph designs which have not yet received much research attention (i.e., pie chart glyphs). The number of dimensions should be varied during the experiment and considered as a factor for analysis, to better understand glyph scalability.

orientation encodings in glyphs, in particular in the context of real multi-dimensional data. Similarly, several other cells of Table 3 are empty or populated by studies from a single paper. As discussed in Section 4.6, the ranking of glyph designs or their variations often depends on tasks and encodings, and as such more studies are needed to be able to provide reliable guidance for general glyph use and design. Especially glyph designs, which have not received much research attention but are used in practice (i.e., pie chart glyphs, or variations of linear profiles) should be prioritized in future studies.

Replicate studies on face glyphs: Many studies have been conducted investigating the performance of faces. Most of these studies were conducted in the 70s, and 80s when faces were newly introduced. In recent years face glyphs have been considered inferior but there are no recent studies or replications of earlier studies to confirm this. Given that some past studies showed good performance, it may be worthwhile to try and reproduce some earlier studies to confirm that they are indeed not as good as their current reputation in the community suggests.

Test larger number of dimensions: In addition to the data encoding, the number of glyph dimensions may highly influence performance. As we saw in Table 4 the vast majority of studies only examined glyphs under a fixed number of dimensions, often less than 10 data dimensions. Only four varied the number of dimensions systematically in their studies. To reliably understand how glyph performance scales, we need to further explore how glyph designs fare under different dimensions.

Summary: Quantitative user studies should be conducted to compare data glyph designs which have not yet received much research attention (i.e., pie chart glyphs). The number of dimensions should be varied during the experiment and considered as a factor for analysis, to better understand glyph scalability.

6.6 Summary

This section motivated promising open research directions for future experiments on data glyphs. In this summary, we revisit the most important gaps we identified and most promising research directions. Firstly, we need to give priority to experiments investigating glyph designs, which have not received much research attention, yet. For example, there is only little knowledge about the performance of radial layouts, such as pie chart glyphs. Having more evaluations about data glyph designs will help to better generalize the outcomes and argue about the performance of visual variables.

Additionally, different presentation settings need to be tested in more detail, since a big advantage of data glyphs is their flexible arrangement on the screen. In most experiments the data glyphs were positioned in a regular grid layout, however, data glyphs can also be arranged in more complex layouts like treemaps. Currently, there is only little guidance whether the performance of data glyphs will change according to context information or layout.

A wider variety of experimental factors should be considered such as: multiple datasets (i. e., synthetic data and real world data), different analysis tasks (e.g., exploration or insight generation), and different study types (i.e., qualitative and quantitative) to get a deeper understanding of the utility and performance of data glyphs.

7 CONCLUSION

In this paper we contribute a systematic review of research papers focusing on the evaluation of data glyphs in quantitative user studies. We organize this work using several criteria, such as glyph types, study presentation settings, datasets and tasks used. An overall summary of the most interesting findings can be found in Table 10. Our goal is to: first, help researchers and practitioners identify relevant previous studies that give insights into glyph design tradeoffs, and get inspired by previous study setups; second, provide a meta analysis of the study outcomes; and third, pinpoint open research directions for the study of data glyphs.

In summary, we found that faces and their variations were the most studied glyphs, followed by circular position encoding glyphs (e.g., star glyphs), which were often also compared to faces. Our analysis showed that at first glance performance rankings differ across studies. Yet, we discuss how some of these seemingly contradictory results can be explained by differences in the study criteria, such as the tasks, density and variations of glyphs tested. Our categorization provides readers with references to past studies that can not only serve as specific design considerations but also be an inspiration to design one's own user study.

Our work also aims to highlight gaps in the literature on data glyph evaluation. Few papers have evaluated variations of glyphs using color encoding, even though such glyphs are used in practice [103]. Moreover, only two works [54], [55] have compared design variations of glyphs using linear position or length encodings, that are well established in practice (i.e., sparklines [93] or profiles [11])—although some more have been compared to faces. We were also unable to find any study on circular orientation/angle encodings as already used in applications (e.g., [12]). At the present time we caution against making overly general recommendation for using one type of glyph over another given in particular the many criteria we needed to use to distinguish and categorize past studies (e.g., datasets, tasks, encodings). There are still several years of research possible to understand how humans perceive and use glyphs. We outlined many opportunities for design and evaluation and hope this work encourages researchers to pursue them.

REFERENCES

[1]   T. Munzner, *Visualization Analysis & Design*. Boca Raton, FL, USA: CRC Press, 2015.

[2]   E. Anderson, "A Semigraphical method for the analysis of complex problems," *Proc. Nat. Academy Sci. United States Amer.*, vol. 43, no. 10, pp. 923–927, 1957.

[3]   H. Chernoff, "The use of faces to represent points in K-dimensional space graphically," *J. Amer. Statist. Assoc.*, vol. 68, no. 342, pp. 361–368, 1973.

[4]   J. Siegel, E. Farrell, R. Goldwyn, and H. Friedman, "The surgical implications of physiologic patterns in myocardial infarction shock," *Surgery*, vol. 72, no. 1, pp. 126–141, 1972.

[5]   W. Peng, M. Ward, and E. Rundensteiner, "Clutter reduction in multi-dimensional data visualization using dimension reordering," in *Proc. IEEE Symp. Inform. Vis.*, 2004, pp. 89–96.

[6]   J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner, "Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets," in *Proc. IEEE Symp. Inform. Vis.*, 2003, pp. 105–112. [Online]. Available: http://dl.acm.org/citation.cfm?id=1947368.1947390

[7]   Y.-H. Chan, C. Correa, and K.-L. Ma, "The generalized sensitivity scatterplot," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 10, pp. 1768–1781, Oct. 2013.

[8]   D. W. Cunningham and C. Wallraven, *Experimental Design: From User Studies to Psychophysics*. Boca Raton, FL, USA: CRC Press, 2012.

[9]   M. O. Ward, "A taxonomy of glyph placement strategies for multidimensional data visualization," *Inform. Vis.*, vol. 1, no. 3/4, pp. 194–210, Dec. 2002. [Online]. Available: http://dx.doi.org/10.1057/palgrave.ivs.9500025

[10] M. Ward, "Multivariate data glyphs: Principles and practice," in *Handbook of Data Visualization*, Berlin, Germany: Springer, 2008, pp. 179–198. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-33037-0_8

[11] S. H. Du Toit, A. G. W. Steyn, and R. H. Stumpf, *Graphical Exploratory Data Analysis*. Berlin, Germany: Springer, 1986.

[12] J. Pearlman and P. Rheingans, "Visualizing network security events using compound glyphs from a service-oriented perspective," in *Proc. Workshop Vis. Comput. Security*, 2007, pp. 131–146.

[13] D. H. Chung, R. S. Laramee, J. Kehrer, and H. Hauser, "Glyph-based multi-field visualization," in *Proc. Sci. Vis.*, 2014, pp. 129–137.

[14] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, "Glyph-based visualization: Foundations, design guidelines, techniques and applications," in *Proc. Int. Conf. Eurographics State Art Rep.*, 2013, pp. 39–63.

[15] T. Ropinski, S. Oeltze, and B. Preim, "Survey of glyph-based visualization techniques for spatial multivariate medical data," *Comput. Graph.*, vol. 35, no. 2, pp. 392–401, 2011.

[16] E. S. Nelson, "The face symbol: Research issues and cartographic potential," *Cartographica*, vol. 42, no. 1, pp. 53–64, 2007.

[17] B. Flury and H. Riedwyl, "Graphical representation of multivariate data by means of asymmetrical faces," *J. Amer. Stat. Assoc.*, vol. 76, no. 376, pp. 757–765, 1981. [Online]. Available: http://www.jstor.org/stable/2287565

[18] B. Kabulov, "A method for constructing chernoff faces oriented toward interval estimates of the parameters," *Soviet J. Comput. Syst. Sci.*, vol. 30, no. 3, pp. 94–97, 1992.

[19] C. Ware, *Information Visualization: Perception for Design*. San Mateo, CA, USA: Morgan Kaufmann, 2004.

[20] N. Andrienko and G. Andrienko, *Exploratory Analysis of Spatial and Temporal data*. Berlin, Germany: Springer, 2006.

[21] R. Jacob, "Facial representation of multivariate data," in *Proc. Graph. Representation Multivariate Data*, 1978, pp. 143–168.

[22] J. Miller, C. Quammen, and M. Fleenor, "Interactive visualization of intercluster galaxy structures in the horologium-reticulum supercluster," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 5, pp. 1149–1156, Sep. 2006. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2006.155

[23] L. Micallef, P. Dragicevic, and J. Fekete, "Assessing the effect of visualizations on Bayesian reasoning through crowdsourcing," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2536–2545, Dec 2012.

[24] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg, "Evaluation of alternative glyph designs for time series data in a small multiple setting," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 3237–3246. [Online]. Available: http://doi.acm.org/10.1145/2470654.2466443

[25] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen, "Taxonomy-based glyph design with a case study on visualizing workflows of biological experiments," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2603–2612, Dec. 2012.

[26] U. Brandes, B. Nick, B. Rockstroh, and A. Steffen, "Gestaltlines," in *Proc. EuroVis*, Eurographics, 2013, pp. 171–180.

[27] D. H. S. Chung, P. A. Legg, M. L. Parry, R. Bown, I. W. Griffiths, R. S. Laramee, and M. Chen, "Glyph sorting: Interactive visualization for multi-dimensional data," *Inform. Vis.*, vol. 14, pp. 76–90, 2015.

[28] H. Levkowitz, "Color icons—Merging color and texture perception for integrated visualization of multiple parameters," in *Proc. IEEE VIS*, Oct. 1991, pp. 164–170.

[29] A. E. Lie, J. Kehrer, and H. Hauser, "Critical design and realization aspects of glyph-based 3D data visualization," in *Proc. 25th Spring Conf. Comput. Graphics*, 2009, pp. 19–26. [Online]. Available: http://doi.acm.org/10.1145/1980462.1980470

[30] M. Chuah and S. Eick, "Information rich glyphs for software management data," *IEEE Comput. Graph. Appl.*, vol. 18, no. 4, pp. 24–29, Jul. 1998.

[31] W. Ribarsky, E. Ayers, J. Eble, and S. Mukherjea, "Glyphmaker: Creating customized visualizations of complex data," *Computer*, vol. 27, no. 7, pp. 57–64, Jul. 1994.

[32] C. Dick, R. Burgkart, and R. Westermann, "Distance visualization for Interactive 3D implant planning," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2173–2182, Dec. 2011.

[33] A. Klippel, F. Hardisty, R. Li, and C. Weaver, "Colour-enhanced star plot glyphs: Can salient shape characteristics be overcome?" *Cartographica*, vol. 44, no. 3, pp. 217–231, 2009.

[34] H. Siirtola, "The Effect of data-relatedness in interactive glyphs," in *Proc. 9th Int. Conf. Inform. Vis.*, Jul. 2005, pp. 869–876.

[35] W. Aigner, S. Miksch, B. Thurnher, and S. Biffl, "Planninglines: Novel glyphs for representing temporal uncertainties and their evaluation," in *Proc. 9th Int. Conf. Inform. Vis.*, Jul. 2005, pp. 457–463.

[36] J. P. Martin, J. E. Swan, R. J. Moorhead, Z. Liu, and S. Cai, "Results of a user study on 2D hurricane visualization," *Comput. Graph. Forum*, vol. 27, no. 3, pp. 991–998, 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2008.01234.x

[37] D. H. Pilar and C. Ware, "Representing flow patterns by using streamlines with glyphs," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 8, pp. 1331–1341, Aug. 2013.

[38] B. Yost and C. North, "Single complex glyphs versus multiple simple glyphs," in *Proc. Extended Abstracts Human Factors Comput. Syst.*, 2005, pp. 1889–1892. [Online]. Available: http://doi.acm.org/10.1145/1056808.1057048

[39] C. G. Healey and J. T. Enns, "Large datasets at a glance: Combining textures and colors in scientific visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 5, no. 2, pp. 145–167, Apr./Jun. 1999.

[40] Y.-N. Li, D-J. Li, and K. Zhang, "Metaphoric transfer effect in information visualization using glyphs," in *Proc. 8th Int. Symp. Vis. Inform. Commun. Interaction*, 2015, pp. 121–130. [Online]. Available: http://doi.acm.org/10.1145/2801040.2801062

[41] W.-Y. Chan, H. Qu, and W.-H. Mak, "Visualizing the semantic structure in classical music works," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 1, pp. 161–173, Jan./Feb. 2010.

[42] C. M. Wittenbrink, A. T. Pang, and S. K. Lodha, "Glyphs for visualizing uncertainty in vector fields," *IEEE Trans. Vis. Comput. Graph.*, vol. 2, no. 3, pp. 266–279, Sep. 1996.

[43] C. Dunne and B. Shneiderman, "Motif simplification: Improving network visualization readability with fan, connector, and clique glyphs," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 3247–3256. [Online]. Available: http://doi.acm.org/10.1145/2470654.2466444

[44] M. Cayli, M. C. Cobanoglu, and S. Balcisoy, "Glyphlink: An interactive visualization approach for semantic graphs," *J. Vis. Languages Comput.*, vol. 24, no. 6, pp. 463–471, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1045926X13000542

[45] M. Chau, "Visualizing web search results using glyphs: Design and evaluation of a flower metaphor," *ACM Trans. Manag. Inform. Syst.*, vol. 2, no. 1, pp. 1–27, Mar. 2011. [Online]. Available: http://doi.acm.org/10.1145/1929916.1929918

[46] N. Siva, A. Chaparro, and E. Palmer, "Evaluation of MILSTD 2525 glyph features in a visual search paradigm," *Proc. Human Factors Ergonomics Soc. Annu. Meet.*, 2014, pp. 1189–1193. [Online]. Available: http://pro.sagepub.com/content/58/1/1189.abstract

[47] L. Wilkinson, "An experimental evaluation of multivariate graphical point representations," in *Proc. Conf. Human Factors Comput. Syst.*, 1982, pp. 202–209. [Online]. Available: http://doi.acm.org/10.1145/800049.801781

[48] E. J. Casey, "Visual display representation of multidimensional systems: The effect of system structure: And display integrality," *Proc. Human Factors Ergonom. Soc. Annu. Meet.*, 1987, pp. 112–115. [Online]. Available: http://pro.sagepub.com/content/31/1/112.abstract

[49] P. M. Jones, C. D. Wickens, and S. J. Deutsch, "The display of multivariate information: An experimental study of an information integration task," *Human Perform.*, vol. 3, no. 1, pp. 1–17, 1990. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/s15327043hup0301_1

[50] D. MacGregor and P. Slovic, "Graphic representation of judgmental information," *Human Comput. Interaction*, vol. 2, no. 3, pp. 179–200, 1986. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/s15327051hci0203_1

[51] M. C. Carswell and C. D. Wickens, "Information integration and the object display: An interaction of task demands and display superiority," *Ergonomics*, vol. 30, no. 3, pp. 511–527, 1987. [Online]. Available: http://dx.doi.org/10.1080/00140138708969741

[52] J. E. Mezzich and D. R. L. Worthington, "A comparison of graphical representations of multidimensional psychiatric diagnostic data," *Graph. Representation Multivariate Data*, pp. 123–141, 1978.

[53] R. J. Petersen, W. W. Banks, and D. I. Gertman, "Performance-based evaluation of graphic displays for nuclear power plant control rooms," in *Proc. Conf. Human Factors Comput. Syst.*, 1982, pp. 182–189. [Online]. Available: http://doi.acm.org/10.1145/800049.801778

[54] D. Albers, M. Correll, and M. Gleicher, "Task-driven evaluation of aggregation in time series visualization," in *Proc. Conf. Human Factors Comput. Syst.*, 2014, pp. 551–560. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557200

[55] P. Goffin, W. Willett, A. Bezerianos, and P. Isenberg, "Exploring the effect of word-scale visualizations on reading behavior," in *Proc. Conf. Human Factors Comput. Syst. Work-In-Progress*, 2015, pp. 1827–1832.

[56] J. Caban, P. Rheingans, and T. Yoo, "An evaluation of visualization techniques to illustrate statistical deformation models," *Comput. Graph. Forum*, vol. 30, no. 3, pp. 821–830, 2011.

[57] I. Borg and T. Staufenbiel, "Performance of snow flakes, suns, and factorial suns in the graphical representation of multivariate data," *Multivariate Behavioral Res.*, vol. 27, no. 1, pp. 43–55, 1992. [Online]. Available: http://dx.doi.org/10.1207/s15327906mbr2701_4

[58] J. Fuchs, P. Isenberg, A. Bezerianos, F. Fischer, and E. Bertini, "The influence of contour on similarity perception of star glyphs," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2251–2260, Dec. 2014.

[59] A. Klippel, F. Hardisty, and C. Weaver, "Star plots: How shape characteristics influence classification tasks," *Cartography Geograph. Inform. Sci.*, vol. 36, no. 2, pp. 149–163, 2009. [Online]. Available: http://dx.doi.org/10.1559/152304009788188808

[60] C. Weigle and R. Taylor, "Visualizing intersecting surfaces with nested-surface techniques," in *Proc. IEEE Vis.*, Oct. 2005, pp. 503–510.

[61] R. Jacob and E. Howard, "The face as a data display," *Human Factors*, vol. 18, no. 2, pp. 189–200, 1976.

[62] C. Frisson, S. Dupont, W. Yvart, N. Riche, X. Siebert, and T. Dutoit, "Audiometro: Directing search for sound designers through content-based cues," in *Proc. 9th Audio Mostly: A Conf. Interaction Sound*, 2014, pp. 1:1–1:8. [Online]. Available: http://doi.acm.org/10.1145/2636879.2636880

[63] Y. Wu and M. Takatsuka, "Visualizing multivariate networks: A hybrid approach," in *Proc. IEEE Pacific Vis. Symp.*, Mar. 2008, pp. 223–230.

[64] M. D. Boulette, B. G. Coury, and N. A. Bezar, "Classification of multidimensional data under time constraints: Evaluating digital and configural display representations," *Proc. Human Factors Ergonomics Soc. Annu. Meet.*, vol. 31, no. 1, pp. 116–120, 1987. [Online]. Available: http://pro.sagepub.com/content/31/1/116.abstract

[65] E. Nelson and P. Gilmartin, "An evaluation of multivariate, quantitative point symbols for maps," in C. Wood and P. Keller (eds). *Cartographic Design: Theoretical Practical Perspectives*, Wiley & Sons, London, pp. 191–203, 1996.

[66] M. D. Lee, R. E. Reilly, and M. E. Butavicius, "An Empirical evaluation of chernoff faces, star glyphs, and spatial visualizations for binary data," in *Proc. Asia-Pacific Symp. Inform. Vis.*, 2003, pp. 1–10. [Online]. Available: http://dl.acm.org/citation.cfm?id=857080.857081

[67] C. Forsell, S. Seipel, and M. Lind, "Surface glyphs for efficient visualization of spatial multivariate data," *Inform. Vis.*, vol. 5, no. 2, pp. 112–124, 2006. [Online]. Available: http://ivi.sagepub.com/content/5/2/112.abstract

[68] M. Livingston and J. Decker, "Evaluation of trend localization with multi-variate visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2053–2062, Dec 2011.

[69] M. A. Livingston and J. W. Decker, "Evaluation of multivariate visualizations: A case study of refinements and user experience," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 8294, pp. 82 940G–82 940G–12, 2012. [Online]. Available: http://dx.doi.org/10.1117/12.912192

[70] P. Saraiya, C. North, and K. Duca, "Comparing benchmark task and insight evaluation methods on timeseries graph visualizations," in *Proc. 3rd BELIV Workshop: BEyond Time Errors: Novel Eval. Methods Inform. Vis.*, 2010, pp. 55–62. [Online]. Available: http://doi.acm.org/10.1145/2110192.2110201

[71] C. Forsell, S. Seipel, and M. Lind, "Simple 3D glyphs for spatial multivariate data," in *Proc. IEEE Symp. Inform. Vis.*, Oct. 2005, pp. 119–124.

[72] M. A. Livingston, J. Decker, and Z. Ai, "An evaluation of methods for encoding multiple 2D spatial data," *Proc. SPIE*, vol. 7868, pp. 78 680C–78 680C–12, 2011. [Online]. Available: http://dx.doi.org/10.1117/12.872576

[73] H. Chernoff and M. H. Rizvi, "Effect on classification error of random permutations of features in representing multivariate data by faces," *J. Amer. Stat. Assoc.*, vol. 70, no. 351, pp. 548–554, 1975. [Online]. Available: http://www.jstor.org/stable/2285931

[74] G. De Soete and W. Do Corte, "On the perceptual salience of features of chernoff faces for representing multivariate data," *Appl. Psychological Meas.*, vol. 9, no. 3, pp. 275–280, 1985. [Online]. Available: http://apm.sagepub.com/content/9/3/275.abstract

[75] G. De Soete, "A Perceptual study of the Flury-Riedwyl faces for graphically displaying multivariate data," *Int. J. Man-Mach. Stud.*, vol. 25, no. 5, pp. 549–555, 1986.

[76] C. G. Hamner, D. W. Turner, and D. M. Young, "Comparisons of several graphical methods for representing multivariate data," *Comput. Math. Appl.*, vol. 13, no. 7, pp. 647–655, 1987.

[77] D. L. Huff and W. Black, "A multivariate graphic display for regional analysis," in *Graphical Representation of Multivariate Data*, Orlando, FL, USA: Academic, 1978, pp. 199–218.

[78] R. Jacob, "Investigating the Space of Chernoff faces," *Recent Adv. Statist.*, Academic Press, New York, pp. 449–468, 1983.

[79] C. J. Morris, D. S. Ebert, and P. L. Rheingans, "Experimental analysis of the effectiveness of features in chernoff faces," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 3905, pp. 12–17, 2000. [Online]. Available: http://dx.doi.org/10.1117/12.384865

[80] M. Naveh-Benjamin and R. G. Pachella, "The effect of complexity on interpreting "Chernoff" faces," *Human Factors*, vol. 24, no. 1, pp. 11–18, 1982. [Online]. Available: http://hfs.sagepub.com/content/24/1/11.abstract

[81] E. S. Nelson, D. Dow, C. Lukinbeal, and R. Farley, "Visual search processes and the multivariate point symbol," *Cartographica*, vol. 34, no. 4, pp. 19–33, 1997.

[82] N. Siva, A. Chaparro, D. Nguyen, and E. Palmer, "Visual search efficiency for features in chernoff faces," *J. Vision*, vol. 13, no. 9, pp. 689–, 2013. [Online]. Available: http://www.journalofvision.org/content/13/9/689.abstract

[83] N. Sivagnanasundaram, A. Chaparro, and E. Palmer, "Evaluation of the presence of a face search advantage in chernoff faces," *Proc. Human Factors Ergonomics Soc. Annu. Meet.*, vol. 57, no. 1, pp. 1610–1614, 2013. [Online]. Available: http://pro.sagepub.com/content/57/1/1610.abstract

[84] S. Moriarity, "Communicating financial information through multidimensional graphics," *J. Accounting Res.*, vol. 17, no. 1, pp. 205–224, 1979. [Online]. Available: http://www.jstor.org/stable/2490314

[85] D. Stock and C. J. Watson, "Human judgment accuracy, multidimensional graphics, and humans versus models," *J. Account. Res.*, vol. 22, no. 1, pp. 192–206, 1984. [Online]. Available: http://www.jstor.org/stable/2490708

[86] P. C. Saxena and K. Navaneetham, "Comparison of Chernoff-type face and non-graphical methods for clustering multivariate observations," *Comput. Stat. Data Anal.*, vol. 15, no. 1, pp. 63–79, 1993.

[87] R. L. Brown, "Methods for the graphic representation of systems simulated data," *Ergonomics*, vol. 28, no. 10, pp. 1439–1454, 1985. [Online]. Available: http://dx.doi.org/10.1080/00140138508963269

[88] E. Chlan and P. Rheingans, "Multivariate glyphs for multi-object clusters," in *Proc. IEEE Symp. Inform. Vis.*, Oct. 2005, pp. 141–148.

[89] C. D. Shaw, J. A. Hall, C. Blahut, D. S. Ebert, and D. A. Roberts, "Using shape to visualize multivariate data," in *Proc. Workshop New Paradigms Inform. Vis. Manipulation Conjunction 8th ACM Int. Conf. Inform. Knowl. Manag.*, 1999, pp. 17–20. [Online]. Available: http://doi.acm.org/10.1145/331770.331777

[90] D. Feng, Y. Lee, L. Kwock, and R. M. Taylor, II, "Evaluation of glyph-based multivariate scalar volume visualization techniques," in *Proc. 6th Symp. Appl. Perception Graphics Vis.*, 2009, pp. 61–68. [Online]. Available: http://doi.acm.org/10.1145/1620993.1621006

[91] T. Jankun-Kelly, Y. Lanka, and J. Swan, "An evaluation of glyph perception for real symmetric traceless tensor properties," *Comput. Graph. Forum*, vol. 29, no. 3, pp. 1133–1142, 2010. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2009.01711.x

[92] C. Zhang, T. Schultz, K. Lawonn, E. Eisemann, and A. Vilanova, "Glyph-based comparative visualization for diffusion tensor fields," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 797–806, Jan. 2016.

[93] E. Tufte, *Beautiful Evidence*, Graphics Press, New York, 2006.

[94] M. Stone, *A field guide to digital color*. Boca Raton, FL, USA: CRC Press, 2013.

[95] F. Vogel and G. Wendt, "Zwillingsuntersuchung über die erblichkeit einiger anthropologischer maße und konstitutionsindices," *Z. Menschl. Vererb. Konstitutionslehre*, vol. 33, pp. 425–446, 1956.

[96] D. S. Ebert, R. M. Rohrer, C. D. Shaw, P. Panda, J. M. Kukla, and D. A. Roberts, "Procedural shape generation for multi-dimensional data visualization," *Comput. Graph.*, vol. 24, no. 3, pp. 375–384, 2000.

[97] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood, "Does organisation by similarity assist image browsing?" in *Proc. Conf. Human Factors Comput. Syst.*, 2001, pp. 190–197. [Online]. Available: http://doi.acm.org/10.1145/365024.365097

[98] W. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *J. Am. Statist. Assoc.*, vol. 79, no. 387, pp. 531–554, 1984.

[99] J. Sreng, A. Lecuyer, C. Megard, and C. Andriot, "Using visual cues of contact to improve interactive manipulation of virtual objects in industrial assembly/maintenance simulations," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 1013–1020, Sep. 2006.

[100] P. Saraiya, C. North, and K. Duca, "An Insight-based methodology for evaluating bioinformatics visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 11, no. 4, pp. 443–456, Jul. 2005.

[101] A. Amirkhanov, B. Fröhler, J. Kastner, E. Gröller, and C. Heinzl, "Inspectr: Multi-modal exploration, visualization, and analysis of spectral data," *Comput. Graph. Forum*, vol. 33, no. 3, pp. 91–100, 2014.

[102] S. Oeltze, W. Freiler, R. Hillert, H. Doleisch, B. Preim, and W. Schubert, "Interactive, graph-based visual analysis of high-dimensional, multi-parameter fluorescence microscopy data in toponomics," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 1882–1891, Dec. 2011.

[103] J. Yang, A. Patro, H. Shiping, N. Mehta, M. Ward, and E. Rundensteiner, "Value and relation display for interactive exploration of high dimensional datasets," in *Proc. IEEE Symp. Inform. Vis.*, 2004, pp. 73–80.

Anastasia Bezerianos received the PhD degree in computer science from the University of Toronto, Canada. She is currebtly an assistant professor at the Univ Paris-Sud, France, and a member of the HCC (LRI) and ILDA (INRIA) teams. Her research interests include interaction and visualization designs for large displays, visual perception and user evaluation.

Daniel A. Keim is a full professor and the head to the Information Visualization and Data Analysis Research Group at the University of Konstanz. Keim received a habilitation in computer science from the University of Munich and has been program cochair of the IEEE InfoVis, the IEEE VAST, and the ACM SIGKDD Conference.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

Johannes Fuchs is a research scientist and lecturer at the University of Konstanz. He is currently working toward the PhD degree at the University of Konstanz. His main research interests include the area of visual analytics and information visualization with a special focus on data glyph design and evaluation.

Petra Isenberg received the PhD degree in computer science from the University of Calgary. He is a research scientist in Inria's Aviz Research Team. Her research interests include information visualization and visual analytics with a focus on novel contexts for visualization application, evaluation in visualization, and interaction.