REGULAR PAPER

Xiaonan Luo · Yuan Yuan · Kaiyuan Zhang · Jiazhi Xia[ID] **· Zhiguang Zhou · Liang Chang · Tianlong Gu**

Enhancing statistical charts: toward better data visualization and analysis

Abstract Conventional statistical charts are widely used in visual analysis. With the development of digital techniques, statistical charts are confronted with problems when data grow in scale and complexity. Accordingly, a huge amount of effort has been paid on the enhancement of standard charts, making the design space dramatically increased. It is cumbersome for naive users to choose appropriate design in a specific analysis scenario. In this paper, we survey the enhancement techniques for a compact set of statistical charts, and identify the types and usage scenarios. Motivated by the new problems, such as data volume and complexity, we present a challenge-and-task-driven framework to guide the understanding of the design space and the decision-making process.

Keywords Statistic charts · Enhancements · Design space

X. Luo · Y. Yuan
Institute of Artificial Intelligence, Guilin University of Electronic Technology, Guilin, Guangxi, China
E-mail: luoxn@guet.edu.cn

Y. Yuan
E-mail: yuanyuan168888@Outlook.com

K. Zhang
State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China
E-mail: zhangkaiyuan20@gmail.com

J. Xia
School of Computer Science and Engineering, Central South University, Changsha, Hunan, China
E-mail: xiajiazhi@csu.edu.cn

Z. Zhou
Zhejiang University of Finance and Economics, Hangzhou, Zhejiang, China
E-mail: zhgzhou1983@163.com

L. Chang (✉)
School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin, Guangxi, China
E-mail: changl@guet.edu.cn

T. Gu
Guangxi Key Lab of Trusted Software, Guilin University of Electronic Technology, Guilin, Guangxi, China
E-mail: cctlgu@guet.edu.cn

1 Introduction

Statistical charts, also known as statistical graphics, play an important role in statistical analysis (Friendly 2008) and exploratory data analysis. They present data in graphics form to provide insights into the underlying structure of data, such as distribution, trend, correlation, and outliers. A variety of statistical charts are developed to present different categories of data. For instance, scatterplots, line charts, bar charts, and parallel coordinates are proposed to present data of various dimensions.

In the big data era, conventional statistical charts are confronted with new challenges when data grow in scale and complexity (Sarikaya and Gleicher 2018). Therefore, the design space of statistical charts needs to be enhanced to address scalability, complex data characteristics, and various analysis tasks. Meanwhile, the usage scenarios of each chart have been expanded in terms of data characteristics and tasks. Generally, the enhanced charts are referred to as statistical charts that vary the original encoding (e.g., the continuous scatterplots Bachthaler and Weiskopf 2008) or add new visual channels to tackle the new challenges.

Being confronted with the large design space of enhanced charts, it is quite cumbersome for naive users to identify appropriate charts and fulfill the visual mapping and interactions (Li et al. 2018). In this work, we aim to help users select statistical charts and associated enhanced designs for specified usage scenarios. Specifically, we review use cases of line charts, bar charts, scatterplots and parallel coordinates. There are several considerations with respect to the reviews. Firstly, they cover the most representative graphic elements, including dots, lines, bars, and areas (a variation of line charts). Secondly, we focused on enhanced statistical charts only. For example, although the pie chart is also widely used, it is not included in this review because it has few enhancing instances. Thirdly, we limit our discussion on tabular data visualization. Following the framework of Sarikaya and Gleicher (2018) and Munzner (2014), we identify the data characteristics, tasks, and design space of each chart, and reason the design space and use cases by tasks and data characteristics. Besides the tasks, we discuss the challenges of each type of charts, such as data scalability and dimension scalability, and consider them as the targets of enhancements with respect to specific tasks. Although Sarikaya and Gleicher (2018) have given a thorough review of scatterplots, our survey remains to cover scatterplots to provide a systematic framework that guides designers to select charts and explore design spaces in terms of data characteristics and tasks as well as challenges. Our framework also provides a base for designers to develop new forms of statistical charts design.

Because statistical charts are widely used in visualization community, covering all papers using enhanced chart is infeasible. Instead, we focus on covering typical enhancement approaches and sample corresponding instances. We limit the scope of our survey in major conferences and journals on visualization, including IEEE VIS, EuroVis, PacificVis, IEEE TVCG, CGF, JOV, and JVLC. Papers in these venues in the last ten years are surveyed and sampled. In addition, a few papers from other areas, such as statistics, are included for their distinct enhancements.

The rest of the paper is organized as follows. In Sect. 2, we review the taxonomies of statistical charts. Section 3 reviews the application cases and taxonomies of each chart in detail. Subsequently, we discuss and conclude our survey in Sects. 4 and 5.

2 Background

In this paper, we illustrate the enhancements of statistical charts in a task-and-challenge-driven manner. According to Munzner's paradigm of visualization design, "What-Why-How" (Munzner 2014), the design spaces are elicited by challenges from data characteristics and tasks. In this chapter, we state the taxonomies of statistical charts in the aspect of data, tasks, and designs.

2.1 Data types and characteristics

Usually, data types and characteristics are the first consideration in information visualization design (Munzner 2014). Several taxonomies have been proposed to guide the selection of charts and designs. Shneiderman (1996) summarized seven data types. In his taxonomy, multi-dimensional data are close to our definition of tabular data that has multiple attributes. Munzner (2014) categorized the attribute types into three types, including categorical, ordinal, and quantitative. Specifically, Sarikaya and Gleicher (2018) summarized the characteristics of tabular data for scatterplots, including class label, number of points,

number of dimensions, spatial nature, and data distribution. They concluded that some characteristics, such as number of data items, number of dimensions, yield challenges to conventional charts.

The lexicon of the above-mentioned taxonomies touches several aspects, including data types (such as multi-dimensional, tree and network), data characteristics (such as the number of dimensions), and tasks (such as presenting the distribution of data). Among these taxonomies, we would like to choose a standard one for tabular data to put our discussion into a unified framework. Following Munzner (2014), we categorize the major data attributes into categorical, ordinal, and quantitative. Some attributes, such as class label, spatial nature, and temporal, can be classified into these three major types. We also identify a set of characteristics, such as number of data items and number of dimensions, as derived characteristics. These derived characteristics yield challenges to visualizations when the scales of them grow. The remaining characteristics, such as distribution, are considered as the aim of tasks for the clarity of discussion.

2.2 Tasks

Tasks are an important consideration in choosing visualizations (Schulz et al. 2013; Zhang et al. 2017; Mei et al. 2018). A thorough discussion on tasks is out of the focus of our review. We list primary tasks of each plot following Munzner (2014), and point out the reason of enhancing techniques in the aspect of tasks. For instance, line charts are designed for showing trend of a variable. Its variation, area chart, is proposed when users want to present the accumulation of the variable. In this sense, we introduce the tasks as the driven force of enhancements, as well as challenges from the data characteristics.

2.3 Designs

The design space of information visualization is built by visual marks and channels. Recently, Sarikaya and Gleicher (2018) collected the design decisions in scatterplots. They clustered the design choices as four major types: point encoding, point grouping, point position, and graph amenities. We extend their clustering to general statistical charts as encoding, grouping, position, and graph amenities, respectively.

Visual encoding, which contains visual marks and channels, is the first consideration of visual design. In the scope of this paper, the original visual marks include points (scatterplots), lines (line charts and parallel coordinates), bars (bar charts and histograms), and areas (area charts). As summarized by Sarikaya and Gleicher (2018), the visual channels contain color, size, symbols, outline, opacity, texture, depth of filed, and blurriness. Munzner (2014) summarized the theory of visual marks and channels and the ranking of channel effectiveness. Generally, the ranking of channels is task dependent.

3 Enhanced charts

3.1 Line charts

The line chart was firstly proposed by Playfair et al. (2005) in 1786. It has been widely used in visualization applications (Ma et al. 2016; Wu et al. 2018, 2019). The connections across data points (as shown in Fig. 1a) present the trends of a series. The temporal trends and other patterns of interest can be easily perceived according to the up and down slopes of data changes. However, when handling data with multiple series, large-scale, or extended tasks, standard line charts are infeasible to present needed information. A large amount of enhancements have been developed to eliminate these challenges.

3.1.1 Handling multiple series

As a commonly used enhancement, the colors of lines (Chen et al. 2015b, 2019) and the shapes of nodes (Liu et al. 2015) are employed to help viewers quickly identify and compare the trends of the different dimensions. Another example (Pagot et al. 2011) is shown in Fig. 1b. Alternatively, multiple series can be presented in parallel with the small multiples technique. For example, Chang et al. (2007) proposed a parallel line charts system to present the multiple dimensions of the represented data, with each row of line charts showing a specific series (Fig. 1c).

3.1.2 Handling large-scale data

Along with the growing number of series, it is difficult to identify lines and points in the line charts, because they overlap each other and bring much visual clutter. In order to explore the potential features of interest, a large amount of improvements have been conducted to enhance the visual expression of line charts (Zhao et al. 2018b; Muelder et al. 2016; Shi et al. 2012). For example, Andrienko et al. (2010) designed two graphs (Fig. 1d) to indicate that the calling behavior on Saturday and Sunday differs from that in the working days. The upper graph is a traditional line chart, in which the temporal records of 238 areas are depicted as lines, overlapping each other. The lower graph is a statistical aggregation of lines, in which the significant statistical variables are presented to better depict the feature trends of original dataset. Other than aggregation solutions, Kincaid (2010) proposed SignalLens (Fig. 1e), in which a Focus+Context approach was provided to get deeper insights into the low-level signal details in the context of the entire signal trace. Liu et al. (2018a) employed blue noise sampling to reduce the number of series while preserving major patterns.

3.1.3 Facilitating expression and tasks

Driven by various requirements of analytical tasks, the traditional line charts are enhanced in different manners. For example, Guo et al. (2019) combined pixel map and line chart to visualize details of variable correlations. Hao et al. (2011) proposed a visual analytics approach for peak-preserving prediction of large seasonal time series, in which color cues are presented to show the difference between the actual and predicted data; the certainty ban shows the confidence of prediction, and the most significant data points are highlighted in the dark shaded area. Zhao et al. (2011) proposed a novel visualization technique called ChronoLenses (Fig. 1f). Users can construct an interactive lens on the span of line chart and perform various transformations on the data. Furthermore, a flexible and reusable time-series visual analysis interface would be created through changing the parameters of lenses.

3.2 Parallel coordinates

Parallel coordinates are a common means of visualizing multivariate data (Inselberg 1985; Al-Dohuki et al. 2017; Xia et al. 2018a). In parallel coordinates, the axes of an n-dimensional space are represented as n parallel lines (see Fig. 2a). A data item in n-dimensional space is visualized as a polyline with vertices on the axes. The position of the vertex on the i-th axis encodes the i-th coordinate of the data item. With this visual encoding, parallel coordinates support analyzing the distribution of data items in each axis and the correlation between neighboring axes. In the past, various enhancements of parallel coordinates have been proposed to facilitate tasks and handle challenges. While there are tremendous variations in the literature, we only review distinct encoding of polylines and enhanced layouts of axes.
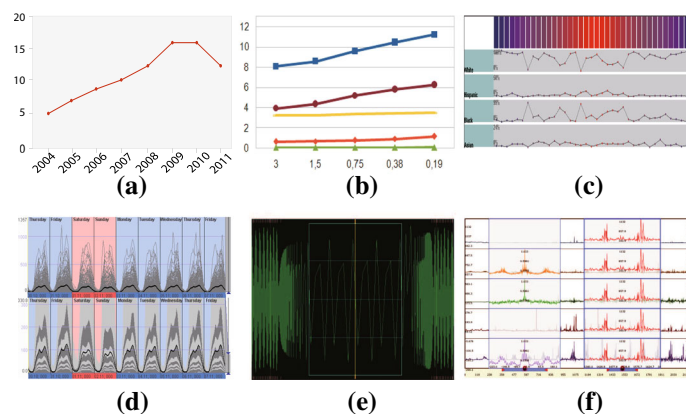

Fig. 1 Standard line chart and its enhancements. **a** A traditional line chart (Munzner 2014); **b** a line chart with color and shape encoding (Pagot et al. 2011); **c** comparison through small multiples (Chang et al. 2007); **d** statistical aggregation (Andrienko et al. 2010); **e** Focus+Context approach (Kincaid 2010). **f** Interactive lenses on line chart (Zhao et al. 2011)

3.2.1 Encoding of polylines

Because parallel coordinates are highly related to line charts, many enhancing techniques for line charts can be performed on parallel coordinates, such as using of color and opacity (Zhao et al. 2019; Holten and Van Wijk 2010) and sampling (Ellis and Dix 2006). These methods are proposed to encode additional information or handle the large-scale data problem. When dealing with category data, Kosara et al. (2006) proposed parallel sets, which encode the number of data items into the width of polylines (see Fig. 2d). Different from line charts, the position and shape of lines in parallel coordinates are flexible. Therefore, many designs replace polylines with smooth curves (Graham and Kennedy 2003) and bundled curves (Palmas et al. 2014) (see Fig. 2b) to facilitate the visually tracing of data items.

3.2.2 Layout of axes

Visualizing the correlation between adjacent axes is one of the major analysis tasks in parallel coordinates. An appropriate dimension ordering is critical to reveal patterns in dealing with multi-dimensional data. A traditional solution is to enable interactive ordering or order axes according to some measures. For instance, Peng et al. (2004) reordered the axes by calculating outliers between neighboring dimensions to reduce the visual clutter. Furthermore, Zhou et al. (2018c) proposed cluster-aware method for parallel coordinate plots to achieve semantic dimension ordering. Another popular solution is integrating scatterplots into parallel coordinates and yielding new layouts. Yuan et al. (2009) combined scatterplots with parallel coordinates to take advantages of both visualizations (see Fig. 2e). The visualization of converting two neighboring axes into a scatterplot shows relationships among multi-dimensions. Claessen and van Wijk (2011) proposed flexible linked axes and integrated scatterplots into the visualization to present multivariate data (see Fig. 2c). Viau et al. (2010) proposed Parallel Scatterplot Matrix that combines a scatterplot matrix and parallel coordinates to visualize and select features within a network (see Fig. 2f).

3.3 Bar charts

A bar chart (Fig. 3a) presents counts of categorical data items with bars, whose length encodes the counts. In another word, it presents a two-dimensional data, where the key attribute is categorical and the other attribute is quantitative (Munzner 2014). The bars can be plotted vertically or horizontally (Gu et al. 2018; Wu et al. 2017; Zhou et al. 2018b). A bar chart supports value comparison of different categories.

3.3.1 Handling multiple dimensions

Stacked bar charts and grouped bar charts (Fig. 3b) are the most common variations when there are two key attributes (Chen et al. 2018b; Streit and Gehlenborg 2014; Xie et al. 2014). Generally, they present sub-bars
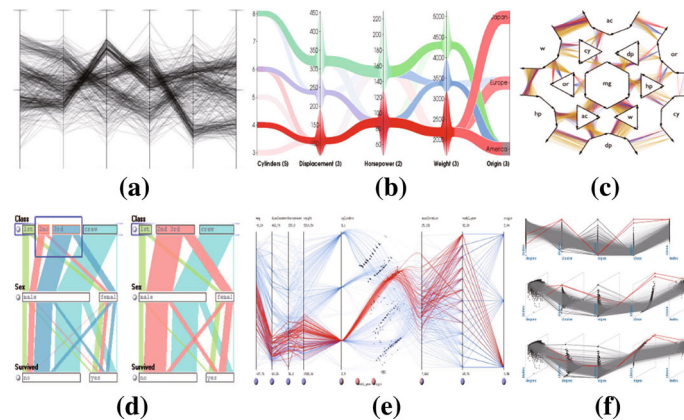
Fig. 2 Standard parallel coordinates and its enhancements. **a** Standard parallel coordinates (Holten and Van Wijk 2010); **b** bundled curves (Palmas et al. 2014); **c** flexible linked axes with integrated scatterplots (Claessen and van Wijk 2011); **d** parallel sets (Kosara et al. 2006); **e** parallel coordinates with integrated scatterplots (Yuan et al. 2009); **f** parallel coordinates combining with scatterplot matrix (Viau et al. 2010)

corresponding to the second key attribute and encode sub-bars, e.g., encoding with color (Liu et al. 2018b; Wang et al. 2018b; Chen et al. 2017). In stacked bar charts, each bar is stacked by multiple sub-bars to present the values of sub-categories (Zhou et al. 2018e; Liao et al. 2015; Huang et al. 2019). In comparison, grouped bar charts plot sub-bars in the category axis to compare the values of sub-categories (Wang et al. 2014; Zhou et al. 2019; Kamw et al. 2019). Other than these two variations, Taher et al. (2016) layouted two key attributes in a squared bottom, and presented each bar as a physical 3D stack. Similarly, Meuschke et al. (2017) presented 3D stacks while the two key attributes represent spatial information (Fig. 3f). Another solution to handle multiple dimensions is to use area rather than length to encode the value. In such case, bar charts are transferred into mosaic plots (Wickham and Hofmann 2011) (Fig. 3e). It can support more than two key attributes. Chen et al. (2016) proposed another solution to layout the bar charts of different dimensions in a matrix, which is similar to scatterplots matrix.

3.3.2 Handling composite attributes

When analysts are interested in not only the value but also the distribution statistics of each bar, such as the maximum and minimum, level lines are added to each bar to show its statistics (Hajizadeh et al. 2013). However, it may be misleading since the bar is encoded with length, while the level lines are encoded with position (Streit and Gehlenborg 2014).

3.3.3 Facilitating expression and tasks

A great number of approaches are proposed to facilitate the expressiveness of bar charts and analysis tasks (Zhou et al. 2018d). Usually, color channels are used to distinguish data of different categories (Xie et al. 2014). To highlight the part–whole relationship, a part of bars or sub-bars can be encoded in different colors (Hajizadeh et al. 2013; Wang et al. 2018a). To emphasize the relative contributions of each sub-bar, normalized bar charts normalize each bar to a uniform length. To facilitate the comparison among bars, Unger et al. (2018) attached level lines to bar charts. While the conventional rectangular bar works well in bar chart, the general public would appreciate more expressive design. Kim et al. (2017) proposed an approach to generate data-driven graphics, in which the bars are represented as expressive graphics.

3.3.4 Handling individual items

While the bars only represent the counts of each category, designers would like to look for individual items in bar charts. Wang et al. (2019) employed bar chart to visualize degree distribution with discontinuous *x* axis for degree. Dot plots (Wilkinson 1999) stack data items as points in bars (Fig. 3c). Recently, Rodrigues and Weiskopf (2018) presented nonlinear dot plots allowing a dynamic size of points. Ren et al. (2017) presented each bar as stacked glyphs of people to provide an expressive presentation.
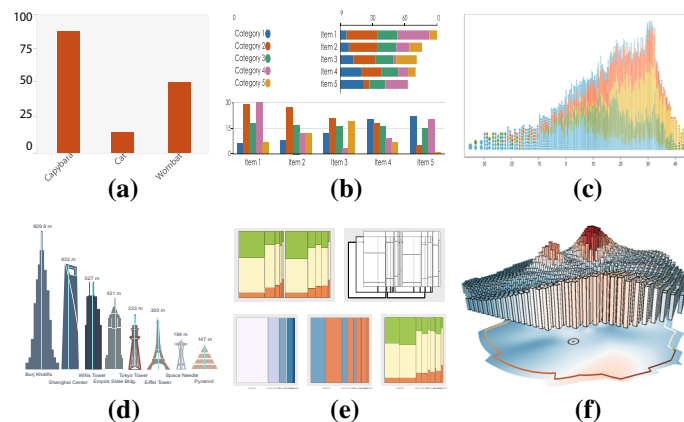

Fig. 3 Standard bar chart and its enhancements. **a** Standard bar chart (Munzner 2014); **b** stacked bar chart (top) and grouped bar chart (bottom) (Streit and Gehlenborg 2014); **c** nonlinear dot plots (Rodrigues and Weiskopf 2018); **d** an expressive bar chart (Kim et al. 2017); **e** mosaic plots (Wickham and Hofmann 2011); **f** A 3D bar chart (Meuschke et al. 2017)

3.3.5 Histograms

Histograms (Pearson 1895) can be considered as a variation of bar charts. It uses the lengths of bars to encode the frequency or frequency density of values. Figure 4a shows an example of the original histograms. Although it has a similar presentation with bar charts, the key attribute for a histogram is continuous rather than discrete. Usually, the first step to construct a histogram is to aggregate the key values into a set of bins.

Having the similar shape with bar charts, histograms also have enhancement approaches alike. The bars or partial of bars can be encoded to highlight the distribution of partial data items (Unger et al. 2018; Chen et al. 2015a) (Fig. 4d). Similarly, stacked histograms are developed to show part–whole relationships (Wickham and Hofmann 2011; Andrienko et al. 2018) (Fig. 4e). van den Elzen and van Wijk (2011) compared different variations of histograms, including stacked histograms, smoothed histograms and streaming graphs. They concluded that smoothed histograms prevent discontinuities for easier interpretation, and streaming graphs are best suited to see individual class distributions as well as quantities. Different layouts of histograms have also been proposed for specific scenarios. For high-dimensional data visualization, Fan et al. (2017) used a color-encoded smoothed histogram to present the entropy information in the event stream. Wan and Hansen (2017) added extra axes for higher-dimension data. Geng et al. (2011) presented angular histograms to present the frequency of data in each axis (Fig. 4b). In a radial-layout visualization, the bars of a histogram can be shaped as arcs (Alsallakh et al. 2014, 2012) (Fig. 4f).

3.4 Scatterplots

Scatterplots encode objects with two quantitative attributes as marks in a two-dimensional space. Figure 5a shows an example of original scatterplots. The two attributes are encoded as positions in the two axes, respectively. Munzner (2014) summarized that original scatterplots are suitable for hundreds of data items. When the scale and complexity of data grow, the design of traditional scatterplots is enhanced to handle new challenges (Sarikaya and Gleicher 2018; Zhou et al. 2018a; Ma et al. 2018).

3.4.1 Handling multiple dimensions

Usually, designers enhance the scatterplots with additional visual channels to encode additional attributes (Wu et al. 2015). Sarikaya and Gleicher (2018) summarized that possible channels include color, size, symbols, opacity, texture, depth of field, and blurriness. Among these channels, Gleicher et al. (2013) showed that symbols are weaker than color in identifying multiclass scatterplots (Fig. 5c). Li et al. (2009) evaluated the performance of symbols as well as the size, and Li et al. (2010) studied the discrimination of opacity in scatterplots, respectively.

For categorical attributes, such as class, a great many of works take color as the first choice (Ma et al. 2017). For instance, Brown et al. (2012), Aupetit et al. (2014), and Chen et al. (2015a) encoded the class of
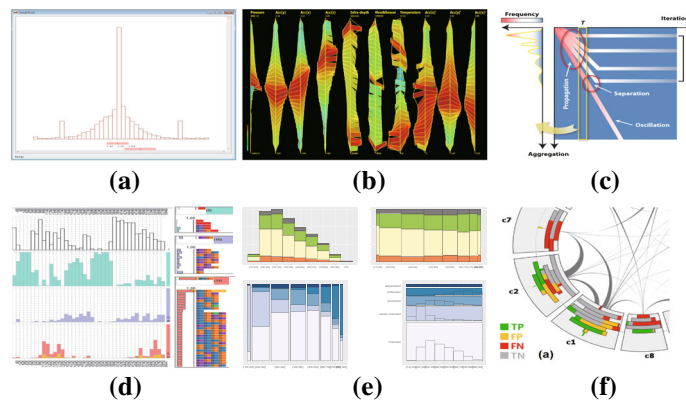

Fig. 4 Standard histogram and its enhancements. **a** Guo et al. (2011): standard histogram. **b** Geng et al. (2011): angular histogram. **c** Wan and Hansen (2017): the histogram with an extra axis. **d** Unger et al. (2018): the sequence of partial encoded histograms. **e** Wickham and Hofmann (2011): mosaic plots. **f** Alsallakh et al. (2014): the histogram shaped as arcs

points into color channel, respectively. Xia et al. (2017) allowed users to set color of points to identify their classes. Redundant channels have also been used to emphasize an attribute (Kanjanabose et al. 2015).

For quantitative attributes, designers often choose size, opacity, and blurriness to encode them. Usually, uncertainty is encoded to opacity channel (Xia et al. 2018b) or blurriness (Feng et al. 2010) (Fig. 5d). Inspired by the concept of depth of field in optics, Staib et al. (2016) encoded the distance to the current focus into blurriness. Choo et al. (2014) encoded citation counts into size channel.

When there are multiple additional attributes, designers often encode them into multiple channels. For instance, Choo et al. (2014) encoded three attributes into color, symbols and size, respectively. When the dimensionality of data continues to grow, the increasing of visual channels results in perception burden rapidly. To address this issue, one choice is to use a fan-like glyph to represent attributes (Liao et al. 2018; Kwon et al. 2017; Zhao et al. 2018a). When the data item represents an image, designers can directly plot the images in the scatterplots (Tenenbaum et al. 2000; Dang and Wilkinson 2014; Chen et al. 2018a).

3.4.2 Handling large-scale data

When the number of data points grows, visual clutter happens, i.e., marks overlap each other. To address this issue, strategies can be categorized as reducing the data, simplifying the visual representation, and modifying the space of the plot (Sarikaya and Gleicher 2018). An example to reduce the data is subsampling (Bertini and Santucci 2006; Chen et al. 2014). Mayorga and Gleicher (2013) addressed the overlapping issue by abstracting dense regions as smooth shapes and subsampling outlying points in the remaining regions. Generalized scatterplots (Keim et al. 2010) distorted the representation to take advantage of unused space. Similarly, continuous scatterplot approaches (Bachthaler and Weiskopf 2008; Lehmann and Theisel 2010; Heinrich et al. 2011) transfer discrete data items into a continuous field and generate a color map to show the field (Fig. 5f). In the same line, Kernel Density Estimation (KDE) approaches (Lampe and Hauser 2011) estimate the density of data items and create a color map to represent the density distribution.

3.4.3 Handling composite attributes

To represent composite attributes, such as scalar fields, directions, and relationships, graph amenities are added into the scatterplots (Zhu et al. 2019). Cheng et al. (2016) used iso-contours and kernel density estimation to encode scalar field information. Similarly, Cheng and Mueller (2016) encoded scalar-filed to iso-contours. Chen et al. (2014) encoded the trend of data with a line upon each point to identify the direction. The BubbleSet approach (Collins et al. 2009) represents the group relationship among points by bubble-like iso-contours (Fig. 5g).
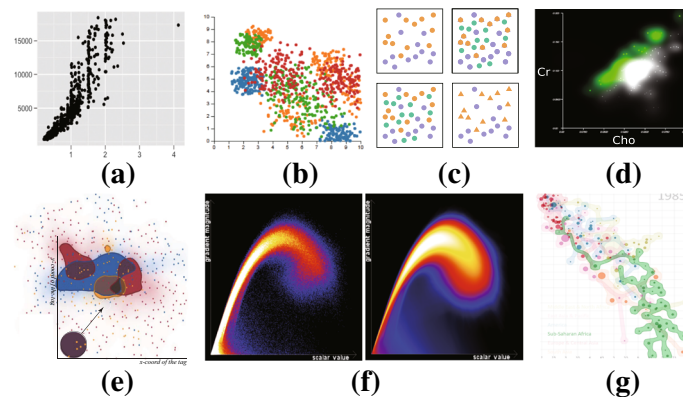
Fig. 5 Standard scatterplot and its enhancements. **a** A standard scatterplot (Munzner 2014); **b** a scatterplot with color encoding (Sarikaya and Gleicher 2018); **c** scatter plots with color encoding and shape encoding (Gleicher et al. 2013); **d** the blurred scatterplot (Feng et al. 2010); **e** the visual abstraction of scatterplot (Chen et al. 2014); **f** progressive splatting of continuous scatterplots (Heinrich et al. 2011); **g** BubbleSet (Collins et al. 2009)

Table 1 The enhancement techniques of statistical charts are clustered into four categories which are similar to Sarikaya and Gleicher (2018)

Enhancement techniques	Line charts	Parallel coordinates	Bar charts	Scatterplots
Encoding	Color	Color	Color	Color
	Width	Width	–	Size
	–	–	Symbols	Symbols
	Opacity	Opacity	Opacity	Opacity
	Texture	–	Texture	Texture
	–	–	–	Blurriness
Grouping	–	–	–	Representation type
	–	–	–	Positional binning
	–	–	–	Enclosure
	–	–	–	Shape abstraction
	–	–	Grouped bar	
Position	–	–	–	Subsampling
	–	–	–	Displacement
	–	–	–	Zooming
	–	Curves	–	–
	–	–	Mosaic plots	–
	–	Flexible layout	Radial layout	–
	Logarithmic scaling	Logarithmic scaling	Logarithmic scaling	–
	Scale break	Scale break	Scale break	–
Graph amenities	Grid lines	Grid lines	Grid lines	Grid lines
	Axis ticks	Axis ticks	Axis ticks	Axis ticks
	Legend	Legend	Legend	Legend
	Lines	Lines	Lines	Lines
	Annotations	Annotations	Annotations	Annotations

4 Discussion

In this section, we would like to shed light on the distribution of enhancement techniques among the four types of charts and reason the distribution we found in the literature. Besides that, we also discuss our framework by listing the challenges and tasks in an exemplar manner.

4.1 The enhancement techniques

Enhancement techniques are referred to added or varied encodings for strengthening visual presentation of statistical charts. Inspired by Sarikaya and Gleicher (2018)'s work, we categorize enhancement techniques into four types, i.e., encoding, grouping, position, and graph amenities. In Table 1, we fill the concrete enhancement techniques, which are found in the literature, in corresponding cells. We found that some techniques are feasible for all four kinds of charts, such as graph amenities (we consider histograms as a variant of bar charts). On the other hand, encoding, grouping, and position are found in a part of types of enhanced charts only. For instance, size can encode an additional attribute in scatterplots. Similarly, in line charts, the width of line can be used to encode an additional attribute. However, in bar charts and histograms, we have not found such an encoding using size or width channel. The reason behind this difference is the inherent feature of primary marks of these charts. It is also the main reason of the difference of grouping and position methods among the four charts.

It is worth noting that there are not available techniques for the options in blank cells. Besides indicating a possible mismatching between enhancement techniques and charts, it may also suggest potential research opportunities. For instance, the shape abstraction is used in scatterplots, and could be used in line charts to show the trend and distribution of a group of polylines.

4.2 Challenges and tasks

We present this survey in a challenges-and-tasks-driven manner. Table 2 presents typical examples of approaches to handle challenges and analysis tasks. We have identified seven typical challenges and tasks for four types of charts in the literature. Our survey indicates that high dimensionality is the most frequent challenge and the primary driven force of statistical charts enhancement. The other two major challenges, large data size and large data range, are mainly identified in line charts and scatterplots.

Table 2 Representative enhancement techniques and corresponding challenges and tasks

Authors/techniques	Challenges and tasks							Chart types				
	Too many data items	Too many dimensions	Large-scale value	Part–whole relationship	Comparison	Trend	Individual analysis	Line charts	Parallel coordinates	Bar charts	Histograms	Scatterplots
Color and shape encoding (Pagot et al. 2011)		✓			✓			✓				
Small multiples (Chang et al. 2007)		✓			✓			✓				
Grouping (Andrienko et al. 2010)	✓					✓		✓				
Graph amenities (Hao et al. 2011)			✓		✓			✓				
Displacement (Zhao et al. 2011)		✓					✓	✓				
Width encoding (Kosara et al. 2006)	✓			✓	✓							
Bundled curves (Palmas et al. 2014)	✓								✓			
Integrating scatterplots (Yuan et al. 2009)			✓	✓			✓		✓			
Flexible layout (Claessen and van Wijk 2011)		✓			✓				✓			
Integrating scatterplots matrix (Viau et al. 2010)		✓							✓			
Stacked bar and grouped bar (Streit and Gehlenborg 2014)		✓			✓					✓		
3D stack (Taher et al. 2016)		✓								✓		
Mosaic plots (Wickham and Hofmann 2011)		✓									✓	
Partial encoding (Hajizadeh et al. 2013)				✓			✓			✓		
Stacked points (Wilkinson 1999)										✓		
Graph amenities (Kim et al. 2017)				✓	✓					✓		
Stacked histogram (Andrienko et al. 2018)		✓		✓							✓	
Stacked histogram (Unger et al. 2018)		✓		✓						✓	✓	
Extra axes (Wan and Hansen 2017)		✓					✓			✓	✓	
Angular layout (Geng et al. 2011)		✓			✓						✓	
Radial layout (Alsallakh et al. 2014)		✓									✓	
Color and opacity encoding (Xia et al. 2018b)												✓
Subsampling (Bertini and Santucci 2006)	✓		✓									✓
Shape abstraction (Mayorga and Gleicher 2013)	✓		✓									✓
Graph amenities (Chen et al. 2014)						✓						✓
Shape abstraction (Heinrich et al. 2011)	✓										✓	✓
Graph amenities (Collins et al. 2009)				✓							✓	✓

Table 2 provides suggestions for designers to select proper charts and designs. In the first step, designers can choose the proper type of chart according to the data type and the major analysis task. Subsequently, designers can identify the data characteristics and derived tasks, e.g., the number of data items, the number of dimensions, and the range of values. The characteristics and tasks, which may yield challenges, lead to the choice of design space. This table gives examples when the chart type and challenges are identified. Although the supported data types and tasks of different types of charts may overlap with each other, we suggest making the decision of chart type and designs following the above process. In this way, the performance of the major tasks could be maximized.

5 Conclusion

Statistical charts are widely used in exploratory data analysis and are the origins of many visual forms. While there are many available enhancement techniques, understanding the design strategies and making the right choice are valuable. In this paper, we have presented a challenge-and-task-driven framework to help design decisions. We have provided abundant examples and indicated potential areas for innovation in the design of the four statistical charts.

References

Al-Dohuki S, Wu Y, Kamw F, Xin L, Xin L, Ye Z, Ye X, Wei C, Chao M, Fei W (2017) Semantictraj: a new approach to interacting with massive taxi trajectories. IEEE Trans Visual Comput Graph 23(1):11–20

Alsallakh B, Aigner W, Miksch S, Groller ME (2012) Reinventing the contingency wheel: scalable visual analytics of large categorical data. IEEE Trans Visual Comput Graph 18(12):2849–58

Alsallakh B, Hanbury A, Hauser H, Miksch S, Rauber A (2014) Visual methods for analyzing probabilistic classification data. IEEE Trans Visual Comput Graph 20(12):1703–1712

Andrienko G, Andrienko N, Mladenov M, Mock M, Pölitz C (Oct 2010) Discovering bits of place histories from people's activity traces. In: 2010 IEEE symposium on visual analytics science and technology, pp 59–66. https://doi.org/10.1109/VAST.2010.5652478

Andrienko G, Andrienko N, Fuchs G, Garcia JMC (2018) Clustering trajectories by relevant parts for air traffic analysis. IEEE Trans Visual Comput Graph 24(1):34–44. https://doi.org/10.1109/TVCG.2017.2744322

Aupetit M, Heulot N, Fekete J (Oct 2014) A multidimensional brush for scatterplot data analytics. In: 2014 IEEE conference on visual analytics science and technology (VAST), pp 221–222. https://doi.org/10.1109/VAST.2014.7042500

Bachthaler S, Weiskopf D (2008) Continuous scatterplots. IEEE Trans Visual Comput Graph 14(6):1428

Bertini E, Santucci G (2006) Give chance a chance: modeling density to enhance scatter plot quality through random data sampling. Inf Visual 5(2):95–110

Brown ET, Liu J, Brodley CE, Chang R (Oct 2012) Dis-function: Learning distance functions interactively. In: 2012 IEEE conference on visual analytics science and technology (VAST), pp 83–92. https://doi.org/10.1109/VAST.2012.6400486

Chang R, Wessel G, Kosara R, Sauda E, Ribarsky W (2007) Legible cities: focus-dependent multi-resolution visualization of urban relationships. IEEE Trans Visual Comput Graph 13(6):1169–1175

Chen H, Chen W, Mei H, Liu Z (2014) Visual abstraction and exploration of multi-class scatterplots. IEEE Trans Visual Comput Graph 20(12):1683–92

Chen H, Zhang S, Chen W, Mei H, Zhang J, Mercer A, Liang R, Qu H (2015a) Uncertainty-aware multidimensional ensemble data visualization and exploration. IEEE Trans Visual Comput Graph 21(9):1072–1086

Chen W, Guo F, Wang FY (2015b) A survey of traffic data visualization. IEEE Trans Intell Transp Syst 16(6):2970–2984

Chen W, Lao T, Xia J, Huang X, Zhu B, Hu W, Guan H (2016) Gameflow: narrative visualization of NBA basketball games. IEEE Trans Multimed 18(11):2247–2256

Chen W, Lu J, Kong D, Liu Z, Shen Y, Chen Y, He J, Liu S, Qi Y, Wu Y (2017) Gamelifevis: visual analysis of behavior evolutions in multiplayer online games. J Visual 20(3):1–15

Chen W, Huang Z, Wu F, Zhu M, Guan H, Maciejewski R (2018a) Vaud: a visual analysis approach for exploring spatio-temporal urban data. IEEE Trans Visual Comput Graph 24(9):2636–2648. https://doi.org/10.1109/TVCG.2017.2758362

Chen W, Xia J, Wang X, Wang Y, Chen J, Chang L (2018b) Relationlines: visual reasoning of egocentric relations from heterogeneous urban data. ACM Trans Intell Syst Technol 10(1):2:1–2:21. https://doi.org/10.1145/3200766

Chen W, Guo F, Han D, Pan J, Nie X, Xia J, Zhang X (2019) Structure-based suggestive exploration: a new approach for effective exploration of large networks. IEEE Trans Visual Comput Graph 25(1):555–565. https://doi.org/10.1109/TVCG.2018.2865139

Cheng S, Cui P, Mueller K (2016) Extending scatterplots to scalar fields. In: IEEE visualization conference (Scivis poster)

Cheng S, Mueller K (2016) The data context map: fusing data and attributes into a unified display. IEEE Trans Visual Comput Graph 22(1):121–130

Choo J, Lee C, Kim H, Lee H, Liu Z, Kannan R, Stolper CD, Stasko J, Drake BL, Park H (Oct 2014) Visirr: visual analytics for information retrieval and recommendation with large-scale document data. In: 2014 IEEE conference on visual analytics science and technology (VAST), pp 243–244. https://doi.org/10.1109/VAST.2014.7042511

Claessen JH, van Wijk JJ (2011) Flexible linked axes for multivariate data visualization. IEEE Trans Visual Comput Graph 17(12):2310

Collins C, Penn G, Carpendale S (2009) Bubble sets: revealing set relations with isocontours over existing visualizations. IEEE Trans Visual Comput Graph 15(6):1009–1016

Dang TN, Wilkinson L (March 2014) Scagexplorer: exploring scatterplots by their scagnostics. In: 2014 IEEE Pacific visualization symposium, pp 73–80. https://doi.org/10.1109/PacificVis.2014.42

Ellis G, Dix A (2006) Enabling automatic clutter reduction in parallel coordinate plots. IEEE Trans Visual Comput Graph 12(5):717–724

Fan X, Peng Y, Zhao Y, Li Y, Meng D, Zhong Z, Zhou F, Lu M (2017) A personal visual analytics on smartphone usage data. J Vis Lang Comput 41:111–120. https://doi.org/10.1016/j.jvlc.2017.03.006

Feng D, Kwock L, Lee Y, Taylor R (2010) Matching visual saliency to confidence in plots of uncertain data. IEEE Trans Visual Comput Graph 16(6):980

Friendly M (2008) The golden age of statistical graphics. Stat Sci 23(4):502–535

Geng Z, Peng Z, Laramee RS, Roberts JC, Walker R (2011) Angular histograms: frequency-based visualizations for large, high dimensional data. IEEE Trans Visual Comput Graph 17(12):2572–2580

Gleicher M, Correll M, Nothelfer C, Franconeri S (2013) Perception of average value in multiclass scatterplots. IEEE Trans Visual Comput Graph 19(12):2316

Graham M, Kennedy J (July 2003) Using curves to enhance parallel coordinate visualisations. In: Proceedings on 7th international conference on information visualization, 2003. IV 2003, pp 10–16. https://doi.org/10.1109/IV.2003.1217950

Gu T, Zhu M, Chen W, Huang Z, Maciejewski R, Chang L (2018) Structuring mobility transition with an adaptive graph representation. IEEE Trans Comput Soc Syst 5(4):1121–1132. https://doi.org/10.1109/TCSS.2018.2858439

Guo Z, Ward MO, Rundensteiner EA, Ruiz C (Oct 2011) Pointwise local pattern exploration for sensitivity analysis. In: 2011 IEEE conference on visual analytics science and technology (VAST), pp 131–140. https://doi.org/10.1109/VAST.2011.6102450

Guo F, Gu T, Chen W, Wu F, Wang Q, Shi L, Qu H (2019) Visual exploration of air quality data with a time-correlation-partitioning tree based on information theory. ACM Trans Interact Intell Syst 9(1):4:1–4:23. https://doi.org/10.1145/3182187

Hajizadeh AH, Tory M, Leung R (2013) Supporting awareness through collaborative brushing and linking of tabular data. IEEE Trans Visual Comput Graph 19(12):2189

Hao MC, Janetzko H, Mittelstädt S, Hill W, Dayal U, Keim DA, Marwah M, Sharma RK (2011) A visual analytics approach for peak-preserving prediction of large seasonal time series. Comput Graph Forum 30(3):691–700

Heinrich J, Bachthaler S, Weiskopf D (2011) Progressive splatting of continuous scatterplots and parallel coordinates. In: Eurographics/IEEE—vGTC conference on visualization, pp 653–662

Holten D, Van Wijk JJ (2010) Evaluation of cluster identification performance for different pcp variants. Comput Graph Forum 29(3):793–802

Huang Z, Lu Y, Mack E, Chen W, Maciejewski R (2019) Exploring the sensitivity of choropleths under attribute uncertainty. IEEE Trans Visual Comput Graph. https://doi.org/10.1109/TVCG.2019.2892483

Inselberg A (1985) The plane with parallel coordinates. Vis Comput 1(2):69–91

Kamw F, Al-Dohuki S, Zhao Y, Eynon T, Sheets D, Yang J, Ye X, Chen W (2019) Urban structure accessibility modeling and visualization for joint spatiotemporal constraints. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2018.2888994

Kanjanabose R, Abdul-Rahman A, Chen M (2015) A multi-task comparative study on scatter plots and parallel coordinates plots. In: Eurographics conference on visualization, pp 261–270

Keim DA, Hao MC, Dayal U, Janetzko H, Bak P (2010) Generalized scatter plots. Inf Visual 9(4):301–311. https://doi.org/10.1057/ivs.2009.34

Kim NW, Schweickart E, Liu Z, Dontcheva M, Li W, Popovic J, Pfister H (2017) Data-driven guides: supporting expressive design for information graphics. IEEE Trans Visual Comput Graph 23(1):491–500. https://doi.org/10.1109/TVCG.2016.2598620

Kincaid R (2010) Signallens: Focus+Context applied to electronic time series. IEEE Trans Visual Comput Graph 16(6):900

Kosara R, Bendix F, Hauser H (2006) Parallel sets: interactive exploration and visual analysis of categorical data. IEEE Trans Visual Comput Graph 12(4):558–568

Kwon BC, Kim H, Wall E, Choo J, Park H, Endert A (2017) Axisketcher: interactive nonlinear axis mapping of visualizations through user drawings. IEEE Trans Visual Comput Graph 23(1):221–230

Lampe OD, Hauser H (Mar 2011) Interactive visualization of streaming data with kernel density estimation. In: 2011 IEEE Pacific visualization symposium, pp 171–178. https://doi.org/10.1109/PACIFICVIS.2011.5742387

Lehmann DJ, Theisel H (2010) Discontinuities in continuous scatter plots. IEEE Trans Visual Comput Graph 16(6):1291–1300. https://doi.org/10.1109/TVCG.2010.146

Li D, Mei H, Shen Y, Su S, Zhang W, Wang J, Zu M, Chen W (2018) Echarts: a declarative framework for rapid construction of web-based visualization. Vis Inf 2(2):136–146

Li J, van Wijk JJ, Martens J (April 2009) Evaluation of symbol contrast in scatterplots. In: 2009 IEEE Pacific visualization symposium, pp 97–104. https://doi.org/10.1109/PACIFICVIS.2009.4906843

Li J, van Wijk JJ, Martens J (March 2010) A model of symbol lightness discrimination in sparse scatterplots. In: 2010 IEEE Pacific visualization symposium (PacificVis), pp 105–112. https://doi.org/10.1109/PACIFICVIS.2010.5429604

Liao H, Wu Y, Chen L, Hamill TM, Wang Y, Dai K, Zhang H, Chen W (Oct 2015) A visual voting framework for weather forecast calibration. In: 2015 IEEE scientific visualization conference (SciVis), pp 25–32. https://doi.org/10.1109/SciVis.2015.7429488

Liao H, Wu Y, Chen L, Chen W (2018) Cluster-based visual abstraction for multivariate scatterplots. IEEE Trans Visual Comput Graph 24(9):2531–2545. https://doi.org/10.1109/TVCG.2017.2754480

Liu S, Chen Y, Wei H, Yang J, Zhou K, Drucker SM (2015) Exploring topical lead-lag across corpora. TKDE 27(1):115–129

Liu M, Shi J, Cao K, Zhu J, Liu S (2018a) Analyzing the training processes of deep generative models. IEEE Trans Visual Comput Graph 24(1):77–87. https://doi.org/10.1109/TVCG.2017.2744938

Liu S, Xiao J, Liu J, Wang X, Wu J, Zhu J (2018b) Visual diagnosis of tree boosting methods. IEEE Trans Visual Comput Graph 24(1):163–173. https://doi.org/10.1109/TVCG.2017.2744378

Ma Y, Lin T, Cao Z, Li C, Wang F, Chen W (2016) Mobility viewer: an Eulerian approach for studying urban crowd flow. IEEE Trans Intell Transp Syst 17(9):2627–2636

Ma Y, Chen W, Ma X, Xu J, Huang X, Maciejewski R, Tung AKH (2017) Easysvm: a visual analysis approach for open-box support vector machines. Comput Vis Media 3(2):1–15

Ma Y, Tung AKH, Wang W, Gao X, Pan Z, Chen W (2018) Scatternet: a deep subjective similarity model for visual analysis of scatterplots. IEEE Trans Visual Comput Graph. https://doi.org/10.1109/TVCG.2018.2875702

Mayorga A, Gleicher M (2013) Splatterplots: overcoming overdraw in scatter plots. IEEE Trans Visual Comput Graph 19(9):1526–1538

Mei H, Ma Y, Wei Y, Chen W (2018) The design space of construction tools for information visualization: a survey. J Vis Lang Comput 44:120–132

Meuschke M, Voss S, Beuing O, Preim B, Kai L (2017) Combined visualization of vessel deformation and hemodynamics in cerebral aneurysms. IEEE Trans Visual Comput Graph 23(1):761

Muelder C, Zhu B, Chen W, Zhang H, Ma KL (2016) Visual analysis of cloud computing performance using behavioral lines. IEEE Trans Visual Comput Graph 22(6):1694–1704

Munzner T (2014) Visualization analysis and design. AK Peters, Natick

Pagot C, Osmari D, Sadlo F, Weiskopf D, Ertl T, Comba J (2011) Efficient parallel vectors feature extraction from higher-order data. Comput Graph Forum 30(3):751–760. https://doi.org/10.1111/j.1467-8659.2011.01924.x

Palmas G, Bachynskyi M, Oulasvirta A, Seidel HP, Weinkauf T (March 2014) An edge-bundling layout for interactive parallel coordinates. In: 2014 IEEE Pacific visualization symposium, pp 57–64. https://doi.org/10.1109/PacificVis.2014.40

Pearson K (1895) Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Philos Trans R Soc A Math Phys Eng Sci 186:343–414

Peng W, Ward MO, Rundensteiner EA (2004) Clutter reduction in multi-dimensional data visualization using dimension reordering. In: IEEE Symposium on information visualization, pp 89–96

Playfair W, Wainer H, Spence I (2005) The commercial and political atlas and statistical breviary (Original version was published in 1786). Cambridge University Press, Cambridge

Ren D, Lee B, Höllerer T (2017) Stardust: accessible and transparent GPU support for information visualization rendering. Comput Graph Forum 36(3):179–188

Rodrigues N, Weiskopf D (2018) Nonlinear dot plots. IEEE Trans Visual Comput Graph 24(1):616–625. https://doi.org/10.1109/TVCG.2017.2744018

Sarikaya A, Gleicher M (2018) Scatterplots: tasks, data, and designs. IEEE Trans Visual Comput Graph 24(1):402–412

Schulz H-J, Nocke T, Heitzler M, Schumann H (2013) A design space of visualization tasks. IEEE Trans Visual Comput Graph 19(12):2366–2375

Shi C, Cui W, Liu S, Xu P, Chen W, Qu H (2012) Rankexplorer: visualization of ranking changes in large time series data. IEEE Trans Visual Comput Graph 18(12):2669–2678

Shneiderman B (Sep. 1996) The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE symposium on visual languages, pp 336–343. https://doi.org/10.1109/VL.1996.545307

Staib J, Grottel S, Gumhold S (2016) Enhancing scatterplots with multi-dimensional focal blur. Comput Graph Forum 35(3):11–20

Streit M, Gehlenborg N (2014) Bar charts and box plots. Nat Methods 11(2):117

Taher F, Jansen Y, Woodruff J, Hardy J, Hornbaek K, Alexander J (2016) Investigating the use of a dynamic physical bar chart for data exploration and presentation. IEEE Trans Visual Comput Graph 23(1):451–460

Tenenbaum JB, Silva Vd, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323

Unger A, Dräger N, Sips M, Lehmann DJ (2018) Understanding a sequence of sequences: visual exploration of categorical states in lake sediment cores. IEEE Trans Visual Comput Graph 24(1):66–76. https://doi.org/10.1109/TVCG.2017.2744686

van den Elzen S, van Wijk JJ (Oct 2011) Baobabview: interactive construction and analysis of decision trees. In: 2011 IEEE conference on visual analytics science and technology (VAST), pp 151–160. https://doi.org/10.1109/VAST.2011.6102453

Viau C, McGuffin MJ, Chiricota Y, Jurisica I (2010) The FlowVizMenu and parallel scatterplot matrix: hybrid multidimensional visualizations for network exploration. IEEE Trans Visual Comput Graph 16(6):1100–1108

Wan Y, Hansen C (2017) Uncertainty footprint: visualization of nonuniform behavior of iterative algorithms applied to 4D cell tracking. Comput Graph Forum 36(3):479–489

Wang F, Chen W, Wu F, Zhao Y, Hong H, Gu T, Wang L, Liang R, Bao H (2014) A visual reasoning approach for data-driven transport assessment on urban roads. In: 2014 IEEE conference on visual analytics science and technology (VAST). IEEE, New York, pp 103–112

Wang X, Chou J, Chen W, Guan H, Chen W, Lao T, Ma K (2018a) A utility-aware visual approach for anonymizing multi-attribute tabular data. IEEE Trans Visual Comput Graph 24(1):351–360. https://doi.org/10.1109/TVCG.2017.2745139

Wang X, Gu T, Luo X, Cai X, Lao T, Chen W, Wu Y, Yu J, Chen W (2018b) A user study on the capability of three geo-based features in analyzing and locating trajectories. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2018.2875021

Wang X, Chen W, Chou J, Bryan C, Guan H, Chen W, Pan R, Ma K (2019) Graphprotector: a visual interface for employing and assessing multiple privacy preserving graph algorithms. IEEE Trans Visual Comput Graph 25(1):193–203. https://doi.org/10.1109/TVCG.2018.2865021

Wickham H, Hofmann H (2011) Product plots. IEEE Trans Visual Comput Graph 17(12):2223–2230

Wilkinson L (1999) Dot plots. Am Stat 53(3):276–281

Wu W, Zheng Y, Qu H, Chen W, Groller E, Ni LM (Oct 2015) Boundaryseer: visual analysis of 2D boundary changes. In: 2014 IEEE conference on visual analytics science and technology (VAST), pp 143–152. https://doi.org/10.1109/VAST.2014.7042490

Wu F, Zhu M, Wang Q, Zhao X, Chen W, Maciejewski R (2017) Spatialctemporal visualization of city-wide crowd movement. J Visual 20(2):183–194

Wu X, Chen Z, Gu Y, Chen W, Me Fang (2018) Illustrative visualization of time-varying features in spatio-temporal data. J Vis Lang Comput 48:157–168. https://doi.org/10.1016/j.jvlc.2018.08.010

Wu Y, Xie X, Wang J, Deng D, Liang H, Zhang H, Cheng S, Chen W (2019) Forvizor: visualizing spatio-temporal team formations in soccer. IEEE Trans Visual Comput Graph 25(1):65–75. https://doi.org/10.1109/TVCG.2018.2865041

Xie C, Chen W, Huang X, Hu Y, Barlowe S, Yang J (2014) Vaet: a visual analytics approach for e-transactions time-series. IEEE Trans Visual Comput Graph 20(12):1743–1752. https://doi.org/10.1109/TVCG.2014.2346913

Xia J, Jiang G, Zhang Y, Li R, Chen W (2017) Visual subspace clustering based on dimension relevance. J Vis Lang Comput 41:79–88. https://doi.org/10.1016/j.jvlc.2017.05.003

Xia J, Gao L, Kong K, Zhao Y, Chen Y, Kui X, Liang Y (2018a) Exploring linear projections for revealing clusters, outliers, and trends in subsets of multi-dimensional datasets. J Vis Lang Comput 48:52–60. https://doi.org/10.1016/j.jvlc.2018.08.003

Xia J, Ye F, Chen W, Wang Y, Chen W, Ma Y, Tung AKH (2018b) LDSScanner: exploratory analysis of low-dimensional structures in high-dimensional datasets. IEEE Trans Visual Comput Graph 24(1):236–245. https://doi.org/10.1109/TVCG.2017.2744098

Yuan X, Guo P, Xiao H, Zhou H, Qu H (2009) Scattering points in parallel coordinates. IEEE Trans Visual Comput Graph 15(6):1001–1008

Zhang T, Wang X, Li Z, Guo F, Ma Y, Chen W (2017) A survey of network anomaly visualization. Sci China (Inf Sci) 60(12):121101

Zhao J, Chevalier F, Pietriga E, Balakrishnan R (2011) Exploratory analysis of time-series with chronolenses. IEEE Trans Visual Comput Graph 17(12):2422–31

Zhao X, Wu Y, Cui W, Du X, Chen Y, Wang Y, Lee DL, Qu H (2018a) Skylens: visual analysis of skyline on multi-dimensional data. IEEE Trans Visual Comput Graph 24(1):246–255

Zhao Y, She Y, Chen W, Lu Y, Xia J, Chen W, Liu J, Zhou F (2018b) Eod edge sampling for visualizing dynamic network via massive sequence view. IEEE Access 6:53006–53018. https://doi.org/10.1109/ACCESS.2018.2870684

Zhao Y, Luo F, Chen M, Wang Y, Xia J, Zhou F, Wang Y, Chen Y, Chen W (2019) Evaluating multi-dimensional visualizations for understanding fuzzy clusters. IEEE Trans Visual Comput Graph 25(1):12–21. https://doi.org/10.1109/TVCG.2018.2865020

Zhou Z, Li H, Liu F, Liu Y, Huang C, Tao Y, Lin H, Su W (2018a) Visual analytics of economic features for multivariate spatio-temporal GDP data. J Visual 21(2):337–350

Zhou Z, Shi C, Hu M, Liu Y (2018b) Visual ranking of academic influence via paper citation. J Vis Lang Comput 48:134–143. https://doi.org/10.1016/j.jvlc.2018.08.007

Zhou Z, Ye Z, Yu J, Chen W (2018c) Cluster-aware arrangement of the parallel coordinate plots. J Vis Lang Comput 46:43–52. https://doi.org/10.1016/j.jvlc.2017.10.003

Zhou Z, Yu J, Guo Z, Liu Y (2018d) Visual exploration of urban functions via spatio-temporal taxi OD data. J Vis Lang Comput 48:169–177. https://doi.org/10.1016/j.jvlc.2018.08.009

Zhou Z, Zhu X, Liu Y, Ren Q, Wang C, Gu T (2018e) Visupi: visual analytics for university personality inventory data. J Visual 21(5):885–901. https://doi.org/10.1007/s12650-018-0499-x

Zhou Z, Meng L, Tang C, Zhao Y, Guo Z, Hu M, Chen W (2019) Visual abstraction of large scale geospatial origin-destination movement data. IEEE Trans Visual Comput Graph 25(1):43–53

Zhu M, Chen W, Xia J, Ma Y, Zhang Y, Luo Y, Huang Z, Liu L (2019) Location2vec: a situation-aware representation for visual exploration of urban locations. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2019.2901117