2021

Bubble-Wall Plot: A New Tool for Data Visualization

Robert M.X. Wu
Central Queensland University, robert_wumx@hotmail.com

Yongwen Wang
Shanxi Fenxi Mining Industry (Group) Co. Ltd, hxkwyw@163.com

Wanjun Yan
Shanxi Fenxi Mining Industry (Group) Co. Ltd, Ywj198306@163.com

Jinwen Gou
Shanxi Fenxi Mining Zhongxing Coal Industry Co. Ltd, 1002351048@qq.com

Bao Liu
Shanxi Fenxi Mining Zhongxing Coal Industry Co. Ltd, 13935843689@139.com

See next page for additional authors

Authors
Robert M.X. Wu, Yongwen Wang, Wanjun Yan, Jinwen Gou, Bao Liu, Yong Shi, Haijun Zhao, Ergun Gide, Xinxin Cui, Peilin Wang, Ya Wang, Jianfeng Fan, and Qi Li

Bubble-Wall Plot: A New Tool for Data Visualization

Full research paper

Robert M.X. Wu
Shanxi Normal University, China
School of Engineering and Technology
Central Queensland University
Sydney, Australia
Email: Robert_wumx@hotmail.com

Yongwen Wang, Wanjun Yan
Shanxi Fenxi Mining Industry (Group) Co. Ltd
Jiexiu, China
Email: hxkwyw@163.com, Ywj198306@163.com

Jinwen Gou, Bao Liu, Yong Shi
Shanxi Fenxi Mining Zhongxing Coal Industry Co. Ltd
Lvliang, China
Email: 1002351048@qq.com, 13935843689@139.com, 287499914@qq.com

Haijun Zhao
Shanxi Fenxi Mining Zhongxing Coal Industry Co. Ltd
Email: 339186548@qq.com

Ergun Gide
School of Engineering and Technology
Central Queensland University
Sydney, Australia
Email: e.gide1@cqu.edu.au

Xinxin Cui
Zhejiang University, China
Email: cui2xin@126.com

Peilin Wang, Ya Wang
Shanxi Kailain Technology Co. Ltd, China
Email: auswinston888@gmail.com, 524659190@qq.com

Jianfeng Fan, Qi Li
Shanxi Normal University, China
Email: 1135490359@qq.com, 670197058@qq.com

Abstract

This research aimed to design a new tool for data visualization with performed features - named Bubble-Wall Plot and assumed that it could be an effective tool for developing data visualization systems. This research reviewed seven data visualization approaches for identifying the outliers, including Line Charts, Glyphs, Parallel Coordinates Plot, Pixel-based technique, Redial visualizations, Scatter Plots, and TreeMap. The challenges for current data visualization approaches were also summarized. Two principles were addressed to design the new tool- keep it simple strategy with the smallest strategy. As a result, the newly designed Bubble-Wall Plot has successfully been adopted to develop a warning system for identifying the outliers in a Case Study company, which was deployed for user acceptance testing in May 2021. The main contribution is that this newly designed tool with the simplest style was well-designed and proven to develop a warning visualization system effectively.

Keywords Bubble-Wall Plot, Cartographic Symbolization, Data Visualization, Visualization Tool, Warning Visualization System

1 Introduction

Data visualization is not something new and is a graphical representation of an object, situation, or set of information in a diagram, photograph, or other sorts of image and form a mental image (Metze 2020, p.745; Tableau, 2021). Over the last 30 years, various approaches, techniques, and concepts have been developed to help the user find a suitable data transformation and visual mapping by iterating and evaluating every possible visualization design combination (Behrisch et al. 2018, p.626). As a tool, data visualization was used to help end-users see the data and better understand data by quickly and uniquely conveying information (Paul et al. 2019, p.339; Walker et al. 2020, p.2; Lundkvist et al. 2021, p.7). As a valuable method, data visualizations across any number of variables help end-users make sense of abstract concepts that are particularly useful for problems involving large amounts of data that cannot be easily or quickly understood (Paul et al. 2019, p.339; Walker et al. 2020, p.2). It leverages the human ability to visually identify or detect patterns and recognition (Paul et al. 2019, p.339; Saket, Endert & Demiralp 2019, p.2505) and prompts actions by grabbing their attention providing illustrative data (Lundkvist et al. 2021, p.7). High-quality visualizations of integrated and multi-scale modeling results may strengthen research reliability (Saka, Oshika & Jimichi 2019, p.703) and even offers a new perspective on data (Irwin, Robinson & Belt 2017, pp.520-521; Saka, Oshika & Jimichi 2019, p.697, p.717).

In the literature, safety management researchers have used data visualization to aid job hazard identification for controlling and improving on-site safety monitoring (Guo, Yu & Skitmore 2017, p.135). Researchers used data visualization to display the measures of the warnings, help users understand hazard behavior, and inform responsive teams of a hazard event or disaster warning (Webley 2011, p.25, p.35; Keon et al. 2014, p.989). It may enable to make the prediction (Saka, Oshika & Jimichi 2019, p.696). For example, as a "rapid response system," early warning systems with data visualization help detect clinical deterioration and improve hospital patient safety; if the system detected an anomaly data, it would trigger an alert (Fang, Lim & Balakrishnan 2020, p.2). However, data analysis without generating informative visualizations often does not help the analysts and readers, especially those without solid statistical backgrounds, understand complex research questions, nuanced relationships between variables, and statistical outputs (Avraam et al. 2021, pp.1-2).

Up-to-data literature presented that no single approach was well-performed to cover all feature metrics for data visualization. Researchers and practitioners also faced several critical challenges in identifying the best visualization approach for a given dataset and task. There are, therefore, necessary to introduce interactive and user-centered visualizations to their possible audience much more frequently (Perkhofer et al. 2019, p.517). The research highlighted that there was a need to find more effective ways to integrate and understand the intersectionality of models in various filed; at the same time, the datasets from diverse fields vary from spatial to relational data, especially as the volume, diversity, and complexity of data increase (Christensen et al. 2018, p.344). On the other hand, many studies in data visualization have focused on leveraging the human visual system to offload cognitive work and visualize information effectively (Horton, Nowak & Haegeli 2020, p.1560). Visual analytics is also rapidly emerging as an interdisciplinary scientific field based on information visualization and cognitive and perceptual sciences (Stenliden, Bodén & Nissen 2019, p.105). However, scientists and researchers were rarely trained to develop visualization (Saka, Oshika & Jimichi 2019, p.702).

This research aimed to design a new tool for data visualization with performed features - named "Bubble-Wall Plot." This research assumed that this newly designed Bubble-Wall Plot could effectively develop data visualization systems as an effective tool. The following sections include related works, review on outlier data visualization, challenges for data visualization approaches, Bubble-Wall Plot, conclusions, limitations, and further research.

2 Related Works

Data Visualization has spread with different terms over an extended period, such as infographics, information graphics, informative graphs, statistical graphics, graphical visualizations, etc. (called data visualization in this research). In the literature, a few studies reviewed research on data visualizations between 1974 and 2016. Bertini, Tatu & Keim (2011) reviewed data visualization techniques in 20 papers published between 1974 and 2010. Dimara & Perin (2020) synthesized a review of interaction in visualization about 59 articles published between 1991 and 2017. Friedman (2021) conducted empirical research for the role of visualization in scientific discovery in the field of ecology between 1996 and 2016.

This research reviewed data visualization approaches discussed in Q1 publications between Jan 2017 and May 2021. 25 visualization approaches were found, including Bar Chart, Bubble Map, Dimension

Hierarchies, Geo-Spatial maps, Glyphs-based techniques, Heatmaps, Histograms, Line Charts, Network Diagrams, Parallel Coordinates Plot, Pie Chart, Pixel-based techniques, Polar Coordinate Plots, Radial visualizations, Radar Chart, Sankey Chart, Scatter Plot, Stacked Graph, Sunburst, Table Lens, Topological Hierarchies, TreeMap, Typographic, Tag Clouds, and Weather Map.

However, research indicated that most data visualization techniques missed identifying outliers (Johansson & Forsell 2016, p.585). Among the above approaches, eight data visualization approaches or tools were used for determining the outliers. They include Line Charts, Geo-Spatial maps, Glyphs-based techniques, Parallel Coordinates Plot, Pixel-based techniques, Redial visualizations, Scatter Plots, and TreeMap (Behrisch et al. 2018, pp.635-647). In contrast with other approaches, Geo-Spatial maps represent concepts with a semantic closer to humanity, such as cities, lands, roads, etc. (Behrisch et al. 2018, p.646).

3 Review on Outlier Data Visualization

This section discussed seven data visualization techniques for identifying outliers, including Line Charts, Glyphs-based techniques, Parallel Coordinates Plot, Pixel-based techniques, Redial visualizations, Scatter Plots, and TreeMap.

Line Charts were commonly used for visualizing time series to scientific data (Wang et al. 2018, p.3096). They may analyze the temporal aspects of data and put their exclusive focus on the task of trend analysis, and used two axes in x- and y-direction to reference each data point in the coordinate system (Behrisch et al. 2018, p.647). Line Chart is significantly more accurate and speed than other charts for Correlation and Distribution tasks (Saket, Endert & Demiralp 2019, p.2508, p.2511), and higher users' preferences than using Scatter Plots (Saket, Endert & Demiralp 2019, p.2511). However, Line Charts face several limitations. The main limit is that Line Chart is slower with response time for retrieving value (Zhang, Sarvghad & Miklau 2021, p.1791). The second limitation is that Line Charts have low performance for Derived Value and Finding Cluster; The axes values were drawn at uniform intervals, making it difficult to precisely identify a specific data point (Saket, Endert & Demiralp 2019, p.2511). One more limitation is that an inappropriate aspect ratio of the height to width may affect the visual perception of the accuracy of value judgments (Heer & Agrawala 2006, p.701; Behrisch et al. 2018, p.647).

Glyphs-based approaches utilize shape, color, opacity, size, location, etc., to encode high-dimensional information by rendering "small graphical symbols" and have been used to provide statistical and sensitivity information to present trends in the data (Liu et al. 2017, p.1256). The most significant advantage of the glyph is its flexibility in the layout. All graphics are designed to represent data points, positioned independently and efficiently combined with other established visualizations opening space for various application areas with additional information about the data (Behrisch et al. 2018, p.641). The main limitation is that this visual feature does not carry any information about the data to add color to the plot or highlight certain visual features and better solve the analysis task (Behrisch et al. 2018, p.641).

A Parallel Coordinates Plot is one of the most popular visualizations for multi- and high-dimensional data: equally-spaced vertical axis represents the dimensions of the dataset; the top of the axis corresponds to the highest, while the bottom to the lowest value in each dimension; data points are mapped to polylines across the axis, such that the intersection between an axis and a polyline marks the data value (Behrisch et al. 2018, p.636). It is strongly recommended visualization in the information visualization community and highly cited in scientific research (Perkhofer, Walchshofer & Hofer 2020, p.62). The main advantage of this technique is that it gained popularity by enabling analysts to explore patterns across a large set of dimensions (Behrisch et al. 2018, p.636) and is even suitable for the visualization of the unlimited number of data points (Abi Akle, Yannou & Minel 2019, p.234). Parallel Coordinates Plots were used for displaying a general overview of the multi-dimensional monitoring data to provide an interface for users to explore and filter the high-dimensional source data (Johansson & Forsell 2016, p.585; Zhang, Gong & Koyamada 2020, p.1089, p.1091, p.1098). Compared with other visualization techniques, Parallel Coordinates Plot could quickly identify correlation patterns at a glance, such as positive, negative, and non-trivial (multiple) correlations (Johansson & Forsell 2016, p.585; Abi Akle, Yannou & Minel 2019, p.235). This visual mapping allows analysts to spot high-level patterns, as well as single data points of interest (Behrisch et al. 2018, p.636). Parallel Coordinates Plot also performs better in accuracy and response time for each visualization task and sensitivity analysis tasks to the clustering, outlier detection, and change detection in subsets of the data (Kanjanabose, Abdul-Rahman & Chen 2015, p.268; Johansson & Forsell 2016, p.585, p.587; Netzel et al. 2017, p.118, p.130).

However, Parallel Coordinates Plots face at least seven challenges. The main challenge is that a perceived pattern in Parallel Coordinates depends primarily on ordering the dimension axis; a proper ordering can reveal unknown patterns; in contrast, a non-useful ordering may hide them (Behrisch et al. 2018, p.636). It requests the applicable order of dimensions to enhance correlation and clustering (Bertini, Tatu & Keim 2011, p.2209). The second challenge a large number of dimensions decreases the available screen space between two axes and results in cluttered plots (Behrisch et al. 2018, p.636). The third challenge is that the patterns disappear due to overplotted lines with an increasing number of data records (Behrisch et al. 2018, p.636). The fourth challenge is that Parallel Coordinates Plot does not exhibit much strength in supporting value retrieval (Kanjanabose, Abdul-Rahman & Chen 2015, p.268). The fifth challenge is that when display time is limited, people spend more time interpreting the perceived signal and generating an output response when using Parallel Coordinates Plots instead of Scatter Plots, which seems to point at an increase in cognitive load (Li, Martens & van Wijk 2010, p.26). The sixth challenge is that all the evidence hence implies that the accuracy of judgment for Parallel Coordinates Plot is much lower than that for Scatter Plots (Li, Martens & van Wijk 2010, p.25). One more challenge is that positive and negative correlation perception is not symmetric: negative correlations are visible more clearly (Behrisch et al. 2018, p.636).

The Pixel-based technique creates a separate view (called sub-window) for every dimension of a dataset, mapped to precisely one pixel within each sub-window, and colored according to the value in the respective dimensions (Behrisch et al. 2018, p.639). It encodes individual data values as pixels and focuses on arranging them in meaningful ways (Liu et al. 2017, p.1255). The Pixel-based visualizations can also be used to identify clusters and outliers on extensive high-dimensional data (Behrisch et al. 2018, p.639). The main advantages are that Pixel-oriented visualizations do not face overplotting issues and display large amounts of data without aggregation (Behrisch et al. 2018, p.638). However, this requires a manual setting of the interestingness thresholds (Behrisch et al. 2018, p.639).

Radial visualizations are two-dimensional projections of high-dimensional data into a circle and arrange the data in a circular or elliptical fashion (Behrisch et al. 2018, p.439, p.640). The main limitation is that a Radial visualization did not be recommended for an interactive data visualization as it was difficult for participants to identify and compare values needed and drill down more into the data rather than visually perceive the differences in the alternative visualization (Albo et al. 2016, p.575, p.577). Another limitation is that radial visualizations are highly dependent on the ordering of dimensions, which is, in turn, dependent on the user's task: if one data instance has high values in two neighboring dimensions, it is plotted more closely to the circumference, in addition, another data instance with high values in two opposite dimensions is plotted more closely to the center of the circle (Behrisch et al. 2018, p.640). This technique cannot show all details due to the visual limitations inherent in radial layouts (Keim et al. 2006, p.127). One more limit is that Radial visualization approaches are not recommended for interactive data visualization (Nguyen et al. 2020, p.2, p.3).

A Scatter Plot is a handy technique widely used for investigating the relationship between two different variables as x- and y-axis in a Cartesian coordinate view and might be extensible to multi-dimensional data and very appropriate for large data visualization (Urribarri & Castro 2017, p.114; Behrisch et al. 2018, p.634). The main advantage of this visual representation is its strength in getting a quick overview of data (Nguyen et al. 2020, p.3). It has the readability of single data instances and patterns to be straightforward and quickly understood, which may help indicate problems, unique properties, and anything interesting about the data (Behrisch et al. 2018, p.634). Scatter Plots have high accuracy, speed, and user preference for finding anomalies (Saket, Endert & Demiralp 2019, p.2511), and better with perceptual accuracy and faster with response time for filter and cluster (Behrisch et al. 2018, p.634; Zhang, Sarvghad & Miklau 2021, p.1791). Scatter Plots have significantly higher accuracy and speed for finding correlations but are lower than Line Charts (Li, Martens & van Wijk 2010, p.26; Harrison et al. 2014, p.1950; Saket, Endert & Demiralp 2019, p.2508, p.2511). Moreover, Scatter Plots outperform Parallel Coordinates Plots in shorter response times (Netzel et al. 2017, p.123) and supporting visual correlation analysis more effectively between two variables (Li, Martens & van Wijk 2010, p.29). Scatter Plots may also outperform Parallel Coordinates Plot to analyze specific relationships such as linear relationships (Johansson & Forsell 2016, p.585).

The main limitation is that Scatter Plot may not provide interactive functions, such as linking, brushing, and zoom-in view, which involve highlighting or de-emphasizing data and have limited in displaying details (Nguyen et al. 2020, p.2, p.3). The second limitation is that Scatter Plot has significant overlapping that presents high superposition when visualizing big datasets (Urribarri & Castro 2017, p.114). The third limitation is that this technique does not exhibit much strength in supporting value retrieval (Kanjanabose, Abdul-Rahman & Chen 2015, p.268). One more limit is that Scatter Plots did not be recommended for interactive data visualization (Nguyen et al. 2020, p.2, p.3).

TreeMap is a standard tool for visualizing trees with associated data in a space-filling and nested layout (Kopp & Weinkauf 2019, p.535), an effective method for visualizing hierarchies (Wang, Wang & Alexander 2015, p.34). This approach is one of the most relevant information visualization techniques to support the analysis of large hierarchical data structures or data clusters (Soares et al. 2020, p.1). Each node is drawn in size related to its data and serves as a container for drawing its children (Kopp & Weinkauf 2019, p.535). All nodes were allocated into square horizontal and vertical rows (Bostock, Ogievetsky & Heer 2011, p.2306). Generally, TreeMap approaches allow one to remain comprehensible at much higher data densities (Behrisch et al. 2018, p.645). As a space-filling visualization, it makes efficient use of the limited screen space and allows for high-dimensions to be displayed efficiently to depict hierarchical data aspects (Behrisch et al. 2018, p.645; Songer, Hays & North 2004, p.182). Large branches in the hierarchy are given large areas (Behrisch et al. 2018, p.645; Songer, Hays & North 2004, p.182). Each rectangle in a TreeMap represents a node in a tree that parent node rectangles enclose child rectangles while its area is designed to be proportional to the node's value (Behrisch et al. 2018, p.645). The size of each sub-rectangle represents one measure, while color or transparency is often used to describe another measurement of data (Wang, Wang & Alexander 2015, p.34; Cheong et al. 2016, p.1394). The TreeMap allows nodes to be resized smoothly, without shuffling or occlusion that would impede the perception of changing values (Bostock, Ogievetsky & Heer 2011, p.2306). Another advantage is that when fully populated with all levels of detail in the hierarchy, TreeMap has a high data/ink ratio (Songer, Hays & North 2004, p.184). TreeMap presents several challenges for data representations. The main challenge is that TreeMaps lacks a value message (Cheong et al. 2016, p.1394) and cannot represent zero and negative values (Soares et al. 2020, p.2). The second challenge is that TreeMap can be applied only to hierarchical data (Wang, Wang & Alexander 2015, p.37). The third challenge is that users find it challenging to compare regions with extreme aspect ratios (Behrisch et al. 2018, p.645). One more challenge is that TreeMaps encode values using an area that is less accurate than judgments of other visual encodings, such as length (Behrisch et al. 2018, p.645).

4 Challenges for Data Visualization Approaches

The above discussions clearly stated that no single approach was well-performed for outlier data visualization. Therefore, this needs to explore a better tool to meet the up-to-data needs for data visualization. However, scholars in scientific disciplines faced unique challenges in creating visualization approaches that required insights from data analysis to data visualization (Friedman 2021, p.35). As business managers and decision-makers lacked familiarity and knowledge concerning interactive visualization options (Perkhofer et al. 2019, p.515), practitioners also faced the challenges of identifying the best visualization system for a given task (Harrison et al. 2014, p.1943; Behrisch et al. 2018, p.625; Dimara & Perin 2020, p.127). The challenges for data visualization approaches were summarized as: requiring an effective interactive visualization tool, being a challenge for interpretation to represented risks, limited to map the correlations between variables, being challenges for designing the user interface to identify data patterns, different viewpoints to current approaches, and challenged by high-dimensional datasets.

5 Bubble-Wall Plot: A New Visualization Tool

The above discussions clearly stated that no single data visualization approach was well-performed to cover all feature metrics for identifying the outliers and summarized the challenges of identifying the best visualization approach for a given dataset and task that practitioners face. This research designed a new data visualization approach with performed features for solving such limitations, which may be a simple and effective tool for data visualization.

5.1 A Bubble-Wall Plot Design

Designing visualization involves visual stimulation with a chart, plot, or diagram that can vary in size, color, and display platform to represent information or data (Yoon et al. 2016, p.245), which should consider information visualization principles (Horton, Nowak & Haegeli 2020, p.1560).

Some studies focused on ten elementary perceptual tasks provided by Cleveland & McGill (1984), including standard position scale, non-aligned position scale, length, direction, angle, area, volume, curvature, shading, and color saturation. Based on Cleveland & McGill's research, Saket et al. (2018, p.1329) assessed 12 interactive graphical encodings to motivate data visualization designers. They include Horizontal Distance, Vertical Distance, Horizontal Position, Vertical Position, Rectangular Area, Circular Area, Horizontal Length, Vertical Length, Horizontal Curvature, Vertical Curvature, Shading, and Angle. Most studies followed Bertin's cartographic symbolization comprising six visual variables:

color, value, size, shape, pattern, and orientation (Cheong et al. 2020, p.1024; Horton, Nowak & Haegeli 2020, pp.1560). Other visual features were also used in the literature, such as Blur, Crispness, focus, fog, locations or spatial position, map, and texture (Behrisch et al. 2018, p.626; Cheong et al. 2020, p.1024; Horton, Nowak & Haegeli 2020, pp.1560).

This research addressed the Keep It Simple Strategy, Keep It the Smallest Strategy, and Bertin's cartographic symbolization principles to design a new visualization tool - named **"Bubble-Wall Plot"** (see Figure 1) as:

- Addressing Keep It the Smallest Strategy (KISS).

The newly designed visualization tool was well-addressed the Keep It the Smallest Strategy. Visualization idioms should present data with the smallest spatial dimensions, avoid three-dimensional visualizations, and use one-dimensional lists where possible (Horton, Nowak & Haegeli 2020, p.1561). The Bubble-Wall Plot was designed with the smallest number of spatial dimensions - a one-dimension design.

- Addressing Keep It Simple Strategy (KISS).

The newly designed visualization tool was well-addressed the Keep It Simple Strategy. Effective and efficient visualizations follow a simple mantra; The most information in the simplest possible form as the simple visualizations help communicate research outcomes to the public (Behrisch et al. 2018, p.625; Christensen et al. 2018, p.344). Research verified that users who viewed simple graphs perceived the results as more credible and aesthetically pleasing than users who viewed complex graphs (Wanzer et al. 2021, p.7). The most straightforward visualization may result in the most uncomplicated process for the users.

The Bubble-Wall Plot was designed with one bubble (a) and two horizontal lines (b and c). The bubble (a) stands for the Bubble status. Line b stands for the upper limit value (ULV). Line c stands for the lower limit value (LLV). The value point b stands for a real-time value, which was retrieved from the real-time system. Two horizontal lines (b and c) determine the range (e) to the changes of the bubble (a).

- Following Bertin's Cartographic symbolization

This research followed Bertin's cartographic symbolization comprising six visual variables: color, value, size, shape, pattern, and orientation (Cheong et al. 2020, p.1024; Horton, Nowak & Haegeli 2020, pp.1560). According to cartographic theory, different color schemes have particular applications and recommended uses, depending on the type of data displayed and the kind of relationship represented (Klockow-McClain, McPherson & Thomas 2020, p.318).

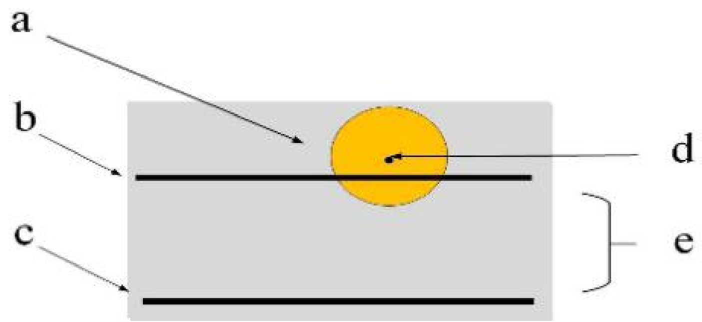


Figure 1: Cartographic Symbolization of A Bubble-Wall Plot

Five types of Bubble-Wall Plots are designated to different statuses in practice: Normal Status (see Figure 2-a), Normal Status Closed to the Upper Limit Value (see Figure 2-b), Normal Status Closed to the Lower Limit Value Figure (see Figure 2-c), Exceeding the Upper Limit Value (see Figure 2-d), and Being Below the Lower Limit Value (see Figure 2-e).

The Bubble-Wall Plot was designed with blue to state the normal status. The Bubble-Wall Plot with yellow indicated the anomalous status – outliers identified. If d exceeds line b, the ULV specifies that the retrieved real-time value exceeds the ULV. The Bubble will be changed to a yellow color from a blue (see Figure 2-d). If the value of d is below the line c - the LLV, which identifies that the retrieved real-time value is below the LLV. The Bubble will also be changed to a yellow color from a blue (see Figure 2-e).

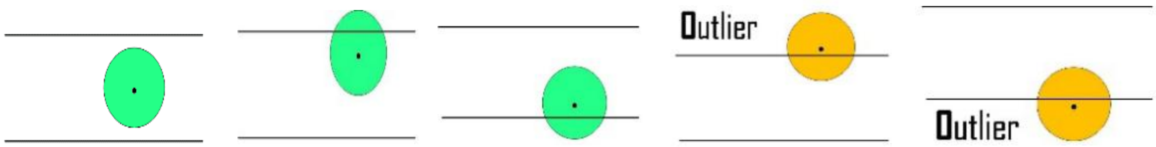


Figure 2-a *Figure 2-b* *Figure 2-c* *Figure 2-d* *Figure 2-e*

Figure 2-a: Normal Status; Figure 2-b: Normal Status Closed to the Upper Limit Value; Figure 2-c: Normal Status Closed to the Lower Limit Value; Figure 2-d: Exceeding the Upper Limit Value; Figure 2-e: Being Below the Lower Limit Value.

5.2 Summary of Significant Features of Bubble-Wall Plot

Table 1 illustrates the comparison between the Bubble-Wall Plot and the other seven visualization tools adopted to identify outliers. Three features were remarkable to the design Bubble-Wall Plot. The simple style is the first significant feature. With the simplest symbolization and smallest spatial dimension, one bubble and two lines comprised the Bubble-Wall Plot. Line Chart and Scatter Plot are simple symbolizations for symbolization style but still complex than the Bubble-Wall Plot. All other five symbolizations are complex. For symbolization dimension, Line Chart, Scatter Plot, and Radial symbolizations need two dimensions. Parallel Coordinate Plots, Pixel, Glyphs, and TreeMap symbolizations use multi-dimensions. The second notable feature is easily visual. The Bubble-Wall Plot can promptly visualize any anomaly changes between two variables. Yellow color signs any outliers or anomaly data to the emergency responsive management team, and easy for the safety-responsive team to manage without knowing the mechanisms of anomaly data. One more significant feature is straightway. The scope (e) to the status of the bubble (a) is directly demonstrated by the two horizontal lines (b and c). If d exceeds the line b or is below the c, it will be the outlier. The Bubble will be changed to a yellow color from a blue (see Figure 2-d, Figure 2-e).

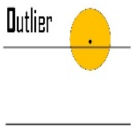
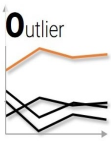
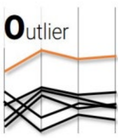
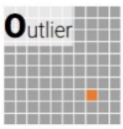
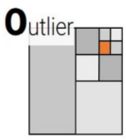
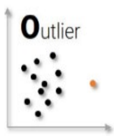
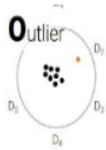
	Bubble-Wall	Line Chart	Parallel Coordinates	Pixel	Glyphs	TreeMap	Scatter Plot	Radial
Symbolization *								
Style	simplest	simple	complex	complex	complex	complex	simple	complex
Dimension	one	two	multi	multi	multi	multi	two	two

Table 1. Comparison of Bubble-Wall Plot and Other Symbolizations to Outliers
*(Note: * provided by Behrisch et al. (2018, pp.635-647))*

6 Conclusions, Limitations, and Further Research

This research aimed to design a new tool for data visualization with performed features - named Bubble-Wall Plot and assumed that it could be an effective tool for developing data visualization systems. This research reviewed seven data visualization approaches for identifying the outliers, including Line Charts, Parallel Coordinates Plot, Scatter Plots, TreeMap, Glyphs, Pixel-based techniques, and Redial visualizations. The challenges for current data visualization approaches were also summarized. Two KISS principles were well-addressed for designing a new data visualization tool -Bubble-Wall Plot.

As a result, the newly designed Bubble-Wall Plot has successfully been adopted to develop a warning system for identifying the outliers in a Case Study company, which was deployed for user acceptance testing in May 2021. The Case Study report is ready as a full research paper for scientific reviewing. The limitation is that this new tool was only adopted in a single case study company for developing a warning system. Further research should be conducted to verify that this newly designed Bubble-Wall Plot might be adopted as an effective data visualization tool to develop different types of data visualization systems. The main contribution is that this newly designed tool with the simplest style was well-designed and proven to develop a warning visualization system effectively.

7 References

Abi, A. A., Yannou, B., and Minel, S. 2019. "Information Visualisation for Efficient Knowledge Discovery and Informed Decision in Design by Shopping," *Journal of Engineering Design* (30:6), pp. 227-253.

Albo, Y., Lanir, J., Bak, P., and Rafaeli, S. 2016. "Off the Radar Comparative Evaluation of Radial Visualization Solutions for Composite Indicators," *IEEE Transactions on Visualization and Computer Graphics* (22:1), pp. 569-578.

Avraam, D., Wilson, R., Butters, O., Burton, T., Nicolaides, C., Jones, E., Boyd, A., and Burton, P. 2021. "Privacy Preserving Data Visualizations," *EPJ data science* (10:1), pp. 1-34.

Behrisch, M., Blumenschein, M., Kim, N. W., Shao, L., El-Assady, M., Fuchs, J., Seebacher, D., Diehl, A., Brandes, U., Pfister, H., Schreck, T., Weiskopf, D., and Keim, D. A. 2018. "Quality Metrics for Information Visualization," *Computer Graphics Forum* (37:3), pp. 625-662.

Bertini, E., Tatu, A., and Keim, D. 2011. "Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization," *IEEE Transactions on Visualization and Computer Graphics* (17:12), pp. 2203-2212.

Bostock, M., Ogievetsky, V., and Heer, J. 2011. "D³ Data-Driven Documents," *IEEE Transactions on Visualization and Computer Graphics* (17:12), pp. 2301-2309.

Cheong, L., Bleisch, S., Kealy, A., Tolhurst, K., Wilkening, T., and Duckham, M. 2016. "Evaluating The Impact of Visualization of Wildfire Hazard Upon Decision-Making Under Uncertainty," *International Journal of Geographical Information Science* (30:7), pp. 1377-1404.

Cheong, L., Kinkeldey, C., Burfurd, I., Bleisch, S., and Duckham, M. 2020. "Evaluating The Impact of Visualization of Risk Upon Emergency Route-Planning," *International Journal of Geographical Information Science* (34:5), pp. 1022-1050.

Christensen, A. J., Srinivasan, V., Hart, J. C., and Marshall-Colon, A. 2018. "Use of Computational Modeling Combined with Advanced Visualization to Develop Strategies for The Design of Crop Ideotypes to Address Food Security," *Nutrition Reviews* (76:5), pp. 332-347.

Cleveland, W. S., and McGill, R. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of the American Statistical Association* (79:387), pp. 531-554.

Dimara, E., and Perin, C. 2020. "What is Interaction for Data Visualization?," *IEEE Transactions on Visualization and Computer Graphics* (26:1), pp. 119-129.

Fang, A. H. S., Lim, W. T., and Balakrishnan, T. 2020. "Early Warning Score Validation Methodologies and Performance Metrics: A Systematic Review," *BMC Medical Informatics and Decision Making* (20:1), pp. 1-7.

Friedman, A. 2021. "Data and Visual Displays in The Journal of Ecology 1996–2016," *Information Visualization* (20:1), pp. 35-46.

Guo, H. L., Yu, Y. T., and Skitmore, M. 2017. "Visualization Technology-based Construction Safety Management: A Review," *Automation in construction* (73), pp. 135-144.

Harrison, L., Fumeng, Y., Franconeri, S., and Chang, R. 2014. "Ranking Visualizations of Correlation Using Weber's Law," *IEEE transactions on Visualization and Computer Graphics* (20:12), pp. 1943-1952.

Heer, J., and Agrawala, M. 2006. "Multi-Scale Banking to 45 Degrees," *IEEE Trans. Visualization and Comp. Graphics* (12:5), pp. 701-708.

Horton, S., Nowak, S., and Haegeli, P. 2020. "Enhancing The Operational Value of Snowpack Models with Visualization Design Principles," *Natural Hazards and Earth System Sciences* (20:6), pp. 1557-1572.

Irwin, W. J., Robinson, S. D., and Belt, S. M. 2017. "Visualization of Large-Scale Narrative Data Describing Human Error," *Human Factors* (59:4), pp. 520–534.

Johansson, J., and Forsell, C. 2016. "Evaluation of parallel coordinates. Overview, Categorization and Guidelines For Future Research," *IEEE Transactions on Visualization and Computer Graphics* (22:1), pp. 579-588.

Kanjanabose, R., Abdul-Rahman, A., and Chen, M. 2015. "A Multi-Task Comparative Study on Scatter Plots and Parallel Coordinates Plots," *Computer Graphics Forum* (34:3), pp. 261-270.

Keon, D., Steinberg, B., Yeh, H., Pancake, C. M., and Wright, D. 2014. "Web-based Spatiotemporal Simulation Modeling and Visualization of Tsunami Inundation and Potential Human Response," *International Journal of Geographical Information Science* (28:5), pp. 987-1009.

Keim, D. A., Mansmann F., Schneidewind, J., and Schreck, T. 2006. "Monitoring Network Traffic with Radial Traffic Analyzer," *2006 IEEE Symposium On Visual Analytics Science And Technology*, pp. 123-128.

Klockow-McClain, K. E., McPherson, R. A., and Thomas, R. P. 2020. "Cartographic Design for Improved Decision Making: Trade-Offs in Uncertainty Visualization for Tornado Threats," *Annals of the American Association of Geographers* (110:1), pp. 314-333.

Kopp, W., and Weinkauf, T. 2019. "Temporal Treemaps: Static Visualization of Evolving Trees," *IEEE Transactions on Visualization and Computer Graphics* (25:1), pp. 534-543.

Li, J., Martens, J. B. O. S., and van Wijk, J. J. 2010. "Judging Correlation From Scatter Plots and Parallel Coordinate Plots," *Information Visualization* (9:1), pp. 13-30.

Liu, S., Maljovec, D., Wang, B., Bremer, P. T., and Pascucci, V. 2017. "Visualizing High-dimensional Data: Advances in The Past Decade," *IEEE Transactions on Visualization and Computer Graphics* (23:3), pp. 1249-1268.

Lundkvist, A., El-Khatib, Z., Kalra, N., Pantoja, T., Leach-Kemon, K., Gapp, C., and Kuchenmüller, T. 2021. "Policy-makers' Views on Translating Burden of Disease Estimates in Health Policies: Bridging The Gap Through Data Visualization," *Archives of Public Health* (79:1), pp. 1-11.

Metze, T. 2020. "Visualization in Environmental Policy and Planning: A Systematic Review and Research Agenda," *Journal of Environmental Policy & Planning* (22:5), pp. 745-760.

Netzel, R., Vuong, J., Engelke, U., O'Donoghue, S., Weiskopf, D., and Heinrich, J. 2017. "Comparative Eye-tracking Evaluation of Scatter Plots and Parallel Coordinates," *Visual Informatics* (1:2), pp. 118-131.

Nguyen, Q. V., Miller, N., Arness, D., Huang, W. D., Huang, M. L., and Simoff S. 2020. "Evaluation on Interactive Visualization Data with Scatter Plots," *Visual Informatics* (4:4), pp. 1-10.

Paul, C. L., Chang, J., Endert, A., Cramer, N., Gillen, D., Hampton, S., Burtner, R., Perko, R., and Cook, K. A. 2019. "TexTonic: Interactive Visualization for Exploration and Discovery of Very Large Text Collections," *Information Visualization* (18:3), pp. 339-356.

Perkhofer, L. M., Hofer, P., Walchshofer, C., Plank, T., and Jetter, H. C. 2019. "Interactive Visualization of Big Data in The Field of Accounting," *Journal of Applied Accounting Research* (20:4), pp. 497-525.

Perkhofer, L., Walchshofer, C., and Hofer, P. 2020. "Does Design Matter When Visualizing Big Data? An Empirical Study to Investigate The Effect of Visualization Type and Interaction Use," *Journal of Management Control* (31:1-2), pp. 55–95.

Saka, C., Oshika, T., and Jimichi, M. 2019. "Visualization of Tax Avoidance and Tax Rate Convergence," *Meditari Accountancy Research* (27:5), pp. 695-724.

Saket, B., Endert, A., and Demiralp, C. 2019. "Task-Based Effectiveness of Basic Visualizations," *IEEE Transactions on Visualization and Computer Graphics* (25:7), pp. 2505–2512.

Saket, B., Srinivasan, A., Ragan E. D., and Endert, A. 2018. "Evaluating Interactive Graphical Encodings for Data Visualization," *IEEE Transactions on Visualization and Computer Graphics* (24:3), pp. 1316-1330.

Soares, A. G. M., Miranda, E. T. C., Lima R. S. d. A. D., Resque, D. S. C. G., and Meiguins, B. S. 2020. “Depicting More Information in Enriched Squarified Treemaps with Layered Glyphs,” *Information* (Basel) (11:2), pp. 1-21.

Songer, A. D., Hays, B., and North, C. 2004. “Multidimensional Visualization of Project Control Data,” *Construction Innovation* (4:3), pp. 173–190.

Stenliden, L., Bodén, U., and Nissen, J. 2019. “Students as Producers of Interactive Data Visualizations-Digitally Skilled to Make Their Voices Heard,” *Journal of Research on Technology in Education* (51:2), pp. 101–117.

Tableau. 2021. “Visualize your Data”. (www.tableau.com, accessed January 11, 2021)

Urribarri, D. K., and Castro, S. M. 2017. "Prediction of Data Visibility in Two-dimensional Scatter Plots," *Information Visualization* (16:2): pp. 113-125.

Walker, J. D., Letcher, B., H, Rodgers, K. D., Muhlfeld, C. C., and D’Angelo, V. S. 2020. “An Interactive Data Visualization Framework for Exploring Geospatial Environmental Datasets and Model Predictions,” *Water* (Basel) (12:2928), pp. 1-20.

Wang, L., Wang, G., and Alexander, C. 2015. “Big Data and Visualization: Methods, Challenges, and Technology Progress,” *Digital Technologies* (1:1), pp. 33-38.

Wang, Y. H., Wang, Z. Y., Zhu, L. F., Zhang, J., Fu, C. W., Cheng, Z. L., Tu, C. H., and Chen, B. Q. 2018. “Is There a Robust Technique for Selecting Aspect Ratios in Line Charts?,” *IEEE Transactions on Visualization and Computer Graphics* (24:12), pp. 3096-3110.

Wanzer, D. L., Azzam, T., Jones, N. D., and Skousen, D. 2021. “The role of Titles in Enhancing Data Visualization,” *Evaluation and Program Planning* (84), pp. 1-10.

Webley, P. W. 2011. “Virtual Globe Visualization of Ash–Aviation Encounters, with The Special Case of The 1989 Redoubt–KLM Incident,” *Computers & Geosciences* (37:1), pp. 25-37.

Yoon, S., Cohen, B., Cato, K. D., Liu, J. F., and Larson, E. L. 2016. “Visualization of Data Regarding Infections Using Eye Tracking Techniques,” *Journal of Nursing Scholarship* (48:3), pp. 244-253.

Zhang, C., Gong, Q. W., and Koyamada, K. 2020. “Visual Analytics and Prediction System Based on Deep Belief Networks for Icing Monitoring Data of Overhead Power Transmission Lines,” *Journal of Visualization* (23:6), pp. 1087-1100.

Zhang, D., Sarvghad, A., and Miklau, G. 2021. “Investigating Visual Analysis of Differentially Private Data,” *IEEE transactions on Visualization and Computer Graphics* (27:2), pp. 1786-1796.

Acknowledgements

The first and second authors contributed to the first authorship equally to this paper.

The third and fourth authors contributed to the second authorship equally to this paper.

Supported by 2021 Shanxi Provincial Education Science "14th Five-Year Plan” (GH-21316).

Supported by the funds of the Shanxi Social Science Federation (SSKLZDKT2019053).

Supported by the funds of the Shanxi Coking Coal Project (201809fx03).

Filed by Australia Provisional Patent Application (2021903202).

Filed by China Patent Application (20211096829.2).

Filed by China Patent Application (20211096839.6).

Filed by China Patent Application (20211097046.6).

Copyright