



A Requirements-Driven Framework for Automatic Data Visualization

Tong Li¹(✉), Xiang Wei², and Yiting Wang¹

¹ Beijing University of Technology, Beijing, China
litong@bjut.edu.cn

² Institut Polytechnique de Paris, Palaiseau, France
xiang.wei@ip-paris.fr

Abstract. Data visualization is an essential method for analyzing big data. Regarding the increasing demands on data visualization generation and understanding, more professional knowledge and skills are required, which are difficult to meet in practice. In most cases, people visualize data using existing templates which might not fit their requirements. We believe that it is essential to establish the connections between users' visualization requirements and visualization solutions. In this paper, we propose a four-layer visualization framework to systematically and automatically map user requirements to data visualization solutions. Specifically, the framework is designed based on typical visual features and attributes and establishes mappings based on their semantics. Based on this framework, we have implemented a web-based prototype, which can automate the generation of visualization solutions from user visualization requirements. To evaluate the framework, we conducted a case study with one participant using the developed prototype and received positive feedback and suggestions.

Keywords: Data visualization · Visualization requirements · Visual variable · Prototype tool

1 Introduction

With the growth of science and the economy, we are entering an era where vast amounts of data are being generated at our fingertips. According to Domo¹, on average, each person creates at least 1.7 MB of data per second in 2020. The power of technology has penetrated deeply into our daily lives, allowing even ordinary users to access vast amounts of data. Big data analysis is increasingly influencing decision-making in various fields, such as economics and politics, and the conclusions drawn through this analysis have the potential to be game-changing [10]. Therefore, it is essential to know how to perform data analysis effectively. According to a scientific survey², the human brain processes images

¹ <https://www.domo.com/>.

² <https://www.t-sciences.com/news/humans-process-visual-data-better>.

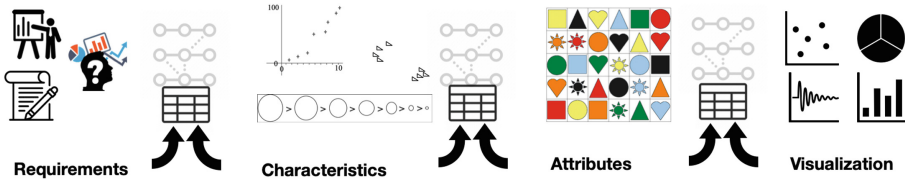


Fig. 1. Overview of the proposed four-layer framework.

60,000 times faster than text. Thus, data visualization is a popular method of processing and presenting data [12]. By presenting various features of data in the form of graphs and charts, data visualization provides a more intuitive and deeper understanding of the data compared to traditional data presentation methods. This is particularly useful for big data, where the visualization results can more effectively represent the information and patterns inherent in the data due to its volume and complexity.

Despite the proliferation of data visualization solutions, generating appropriate visualizations that match specific requirements remains challenging. Most visualization generation tools available today require users to have the expertise and make choices to obtain an appropriate visualization solution [2]. However, the multitude of choices and the lack of clarity on the goals of the requirements make it challenging for users to make effective connections between the two for data visualization. Even obtaining an explicit visualization requirement, such as comparing data, is still difficult to translate into a final visual presentation, such as using line or bar charts to show the comparison results.

In this paper, we propose a four-layer framework that aims to automate the data visualization process by establishing mappings between visualization requirements and visualization solutions. The proposed framework comprises a requirements layer, a visual characteristic layer, a visual attribute layer, and a visualization solution layer, as shown in Fig. 1. Specifically, the requirements layer presents typical visualization requirements, which we collect based on ten interviews targeted at studying users' understanding of data visualization. The visual characteristic layer and the visual attribute layer are defined based on Bertin's research [1]. The former is reasonably related to user requirements, while the latter guides the generation of visualization solutions. We have implemented a prototype tool that embodies our proposed framework and can automate the generation of visualization solutions based on users' specific requirements. We conducted a case study with one participant using our prototype tool to evaluate our proposal. The case study received positive feedback, demonstrating that our proposed framework and prototype tool are useful in practice.

2 Related Work

2.1 Data Visualization Techniques

Although visual charts can convey a great deal of information in a very compact manner, telling the story of big data in a logical way is still a difficult task. Bikakis et al. [2] focus on using modern visualization techniques to process larger and more dynamic data. Python is one of the best-known and most widely used languages for processing data. Furthermore, there are several libraries in Python, such as Seaborn [16], Matplotlib [9], and Plotnine (ggplot) [17], that provide rich templates and methods for data visualization, allowing users to have a standardized way to customize data visualization.

Apart from working with the native language, there are also many advanced studies that integrate data processing together with data visualization. Chuai and Yan [5] introduced methods from data collecting to analyzing. Ren et al. [14] is also an intelligent system that analyzes the data collected from multiple sources.

Although these tools and systems are powerful in data processing and can automatically generate visual charts from big data, they always require a high level of technical knowledge for the user to master them effectively.

2.2 Visualization Requirements Analysis

Many researchers have identified the problem of a lack of mapping between user requirements and data visualization in current automated data visualization systems. As a result, several studies have been conducted to address this issue.

One study by Buja et al. [3] categorizes high-dimensional data analytic tasks into three parts and matches them to three classes of interactive view manipulations. Wu et al. [18] define a set of common tasks in the Artificial Intelligence domain and propose visualizations that can be applied to these tasks. Additionally, a survey by Du and Yuan [6] provides an analysis of competitive sports data to help readers find appropriate visualization techniques for different data types and tasks. All these works provide good matching between user tasks and visual charts. Limited in a specific domain, users who build links are always more precise.

Besides, there are two ways to generalize these mappings. One approach is to collect and summarize mappings proposed by researchers in each domain. The other is to apply visualization on a more generalized dataset and abstract the mappings. Byrd and Dwenger [4] offer a data table that can automatically adjust to different tasks to help users practice and find the most appropriate visualization solution. Furthermore, Peña et al. [13] have proposed a more detailed classification of charts based on the analysis of the Linked Open Data (LOD) dataset. They suggest that this approach could help users to identify visualizations that best satisfy their goals.

Current studies aim to bridge the gap between requirements and data visualizations. However, the general mapping may only suit general user requirements, and specific visualizations are more precise for users working in a specific domain.

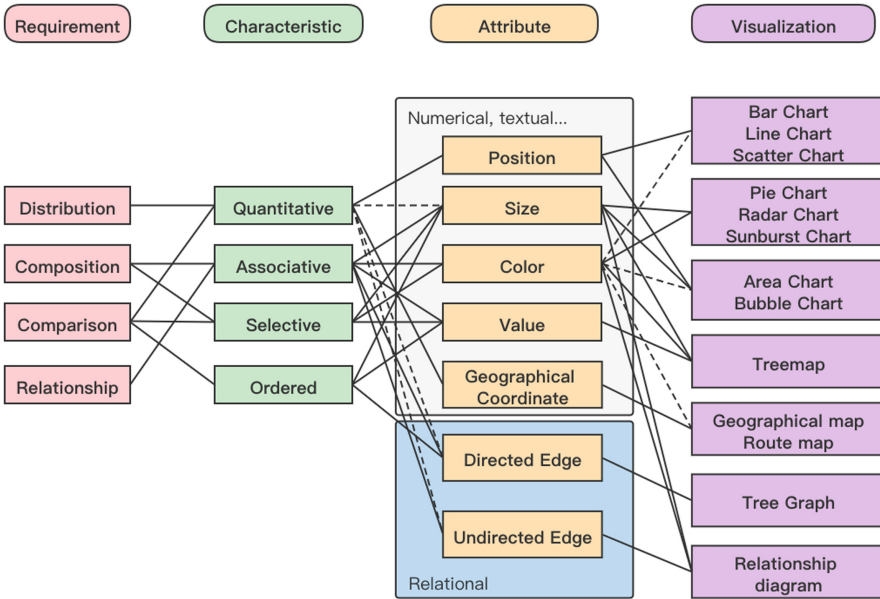


Fig. 2. Four-layer framework in detail.

3 A Four-Layer Framework for Automatically Generating Data Visualization Solutions

In this section, we will detail the design of our four-layer framework. The framework consists of four layers: requirement layer, characteristic layer, attribute layer, and visualization layer, as illustrated in Fig. 2. We will provide further explanation on why we selected these layers and how we built each mapping in the following subsections.

3.1 Understand and Collect User Visualization Requirements

To ensure our framework works automatically, we need to first understand users’ requirements for data visualization. Conducting interviews is an effective empirical method for acquiring straightforward and inspired answers. By designing systematically structured questions, we can obtain as much information about user requirements as possible. The following questions were developed to understand why users need data visualization:

- *Q1:* In what scenarios do you need to create a data analysis diagram?
- *Q2:* What type of data have you worked with?
- *Q3:* What types of diagrams do you usually create?
- *Q4:* What kind of information do you obtain from the charts (shown in Fig. 3), and what tasks do you think they are designed for?

We designed *Question 1 (Q1)* to ask interviewees to describe the scenarios in which they require data visualization. Different types of visualizations have varying abilities to display source data, so we also included *Q2* and *Q3* in asking about the data types and visualization preferences of users. With these three questions, we can gain a general understanding of users' requirements for data visualization. As most current visualization tools require users to choose the appropriate visualization to represent their data, we expect that most users have developed their own approach to building this mapping and selecting the appropriate visualization. Therefore, we included *Q4* in asking users to describe the knowledge they have gained from existing visualization solutions.

We selected examples from six basic chart types (bar, line, area, network, map, and radial). For each chart type example, we hid the chart's title and legend, then encouraged participants to guess the purpose of the chart and the message it conveys. Based on users' responses, we can expect to find an inverse correspondence between the visualization scenarios and users' requirements.

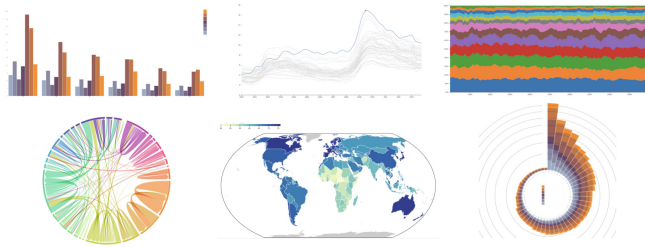


Fig. 3. Example charts provide in the interview.

Because we were conducting this interview to identify patterns from regular users and to gain inspiration from hearing them describe their work patterns, we looked for people who had previous experience with data visualization tools. This way, they could answer our questions more efficiently and clearly. Therefore, we invited 10 candidates to our interview, six of whom were female and four male. They were all aged between 20 and 40, had at least some basic computer skills, and had used data visualization tools more than twice.

After collecting and summarizing the answers we received, we found that they were generally focused, particularly for the first three questions. Although our candidates had different occupations, their answers were surprisingly focused on writing reports and giving presentations (*Answer 1*). In response to the second question, their initial answers were all tables. Therefore, we had to guide them to recall the data types in tables before we received answers such as numerical type, descriptive type, and relational type (*Answer 2*). With regards to the most frequently used charts, we summarized the top five types, which were the bar chart, line chart, scatter chart, pie chart, and geographical map (*Answer 3*).

However, when summarizing the answers to the last question, we found that the responses were more diverse because establishing a connection between user requirements and visualization solutions is highly personalized. Participants' responses typically contained broad directions. For example, in the first bar graph response, a candidate suggested that they might use it to compare the amount of rainfall in different countries or to compare the sales of different brands of cell phones in different months. Even though the two answers given by this participant appeared to be in completely different areas, they both used the same verb, "compare". When we collated the answers we received for the first bar chart, we found that most responses used the verb "compare". For example, it was designed to compare the profits of different products of a company or the incidence of various accidents in the last six years. In fact, we went back to the origin of the graph and found that it was used to compare the distribution of people of different ages across continents in the United States. We, therefore, abstracted the many descriptive answers and used "comparison" as the explanation for the first chart.

Following this pattern, we continued to summarize the speculations of the other charts. We extracted five different purposes: to compare data, to show data trends, to show data composition, to show data relationships, and to show data distribution. Based on the verbs abstracted from the syntax of the user requirements, we categorize them into distribution, composition, comparison, and relationship. We exclude trends from our categorization, as trending is actually a presentation of data relationships. Additionally, we explain how these four requirements apply to the data, based on the complete context, as follows:

- *Distribution*: defines the relative or absolute position of data
- *Composition*: determines the category and hierarchy of data
- *Comparison*: compares the value of data
- *Relationship*: establishes the linkage between data

Based on the conclusions from the interviews, especially the last question, we found that most charts have an inherent meaning in the information they represent. People are surprisingly adept at guessing which type of information matches which type of chart. Moreover, people are also proficient at extracting information from a particular chart. In fact, a proper visualization can convey more information than we expect.

3.2 Mapping User Requirements to Characteristics of Visual Variables

By summarizing user requirements and examining the four extracted words, we found that they match some characteristics mentioned in Bertin's study [7]. One of Bertin's criteria for extracting visualization variables is the expressive power of these attributes in the visualization scheme, i.e., the visual characteristic. We selected the following four characteristics that can be matched to the requirements and adjusted their explanations:

Table 1. Requirements - Characteristics mapping.

| Requirement | Characteristic |
|--------------|----------------------------------|
| Distribution | Quantitative |
| Composition | Associative, Selective |
| Comparison | Ordered, Selective, Quantitative |
| Relationship | Associative |

- *Quantitative*: a single or a group of visual symbols can present absolute value
- *Selective*: a single or a group of visual symbols can stand out from others
- *Ordered*: a series of visual symbols can be arranged in sequence
- *Associative*: a series of visual symbols can be grouped according to similarity

As mentioned above, we found similarities between the characteristics and requirements. However, during the selection of the characteristics, we found that they do not have a one-to-one relationship. Some requirements need to be expressed by multiple expression characteristics. Therefore, we defined a clear mapping between the requirements and characteristics of visual variables. For the *Distribution* goal, whether in the coordinate axis or the map, we can accurately read out the coordinate *quantitative* information. For *Composition*, there are differences between parts, so there is *selective*; each part can be combined into a whole, and there must be an *associative* between them. *Comparison* exists when there are differences among data values, so there is *selective*; values can be distinguished, so there is *quantitative*; if values can be sorted, there is *ordered*. *Relationship* matches the relational chart, so linked data can be grouped together, then there is an *associative*. When we draw a link, there is a weight on a different connection, so *quantitative* is involved. To show the mapping more clearly, we have put them into Table 1.

3.3 Mapping Between Characteristics and Attributes of Visual Variables

Next, we will focus on the attributes with different expressive characteristics and design mapping relations between the characteristics and attributes. Since each selected attribute has different characteristics, Table 2 shows the relationship among them. Specifically, in Table 2, *Y* indicates that the current attribute satisfies this property, *X* indicates that the current attribute does not have this property, and *?* indicates that the situation is not unique.

- *Position*: is a fundamental attribute commonly used in visualization. It refers to the X and Y coordinates, which respectively correspond to the horizontal and vertical addresses of a point in a two-dimensional (2D) space. Together, these coordinates provide precise location information for a point.
- *Size*: is selective because we can easily distinguish symbols with different sizes. Also, we can group symbols with similar sizes. We can order sizes either from small to large or vice versa. If the visualization variable is aligned with the

Table 2. Attributes-Characteristics Mapping.

| Attributes/Characteristics | Selective | Associative | Quantitative | Ordered |
|----------------------------|-----------|-------------|--------------|---------|
| Position | Y | Y | Y | Y |
| Size | Y | Y | ? | Y |
| Value | Y | Y | X | Y |
| Color | Y | Y | X | X |
| Geographical | Y | Y | Y | Y |
| Directed edge | X | Y | ? | Y |
| Undirected edge | X | Y | ? | X |

coordinate axis, we can quickly and directly obtain the area through simple and multi-calculation. But if not, it is difficult to see the value of the area directly. Therefore, the quantitative of the size is uncertain.

- *Value*: is mostly used to represent relative values. Given a group of visual variables with different gray scales, we can easily distinguish, categorize, and sort them. So it has “selective”, “associative” and “ordered”. Since it is a relative value and can’t express the value directly, it does not contain “quantitative”.
- *Color*: is an attribute that is often used to distinguish different variables or identify groups. Although it is very powerful in terms of identification, it is relatively weak in terms of expressing absolute values, and there is no accepted sorting order between different colors. So “color” has “selective” and “associative” but doesn’t have “quantitative” and “ordered” attributes.
- *Geographical*: is an attribute that appears on the geographic map. It is similar to the “position” attribute on the coordinate axes and has four characteristics of expression. For example, the user can select a point on the map individually to get its geographic coordinates, sort it according to its latitude and longitude, or group it according to the pair of continental plates it belongs to.
- *Directed Edge* and *Undirected Edge*: are attributes that appears on a relationship map. Based on the connected node, we can easily select a connected link or group the links together. On edge, there is not always a weight on it. So the “quantitative” is uncertain. For the directed edge, there is a path to follow, so it contains “ordered”, while the undirected edge doesn’t.

According to Tamara [11], although some attributes share the same characteristics, their expressive abilities differ, and there is a hierarchy of these abilities. For instance, position, size, value, geography, and directed edges have ordered characteristics, with the position on a common scale having the best expressive ability. Position on an unaligned scale is slightly lower in the ranking than aligned position, followed by size, then color, and value. Although Tamara’s study did not mention geographic coordinates, we believe there are similarities between geographic coordinates and position on unaligned scales, so we place them between position and size.

Furthermore, attributes with selective characteristics, such as position, size, value, color, and geography, also have different rankings. Ware [15] has proposed

Table 3. Attributes Characteristics Effectiveness Ranking.

| Characteristics | Effectiveness Ranking | |
|-----------------|--|-----------------------------------|
| | Non-Relational data | Relational data |
| Selective | Color > Value > Position > Geographical > Size | |
| Ordered | Size > Value > Position > Geographical | Directed edges |
| Associative | Color > Value > Position > Geographical > Size | Directed edges > Undirected edges |
| Quantitative | Position > Geographical | |

a theory about the ability to combine variables, suggesting that human vision is more attracted to color, value, position, and size. We can thus create a ranking for each attribute based on its relevance, which is carried by all attributes. To determine the ranking, we can refer to selectivity. Although combination and selection work in opposite ways, the expressiveness of the attributes taken should be the same.

Finally, we consider the ranking between attributes that have quantitative characteristics. Since size does not necessarily express quantity, we do not put it in the hierarchy. The coordinate positions in the geographic map correspond to more information than the positions in the 2-dimensional axes. So in terms of the property of simply expressing a quantitative characteristic, it may be slightly inferior. In fact, when ranking, we find it difficult to compare directed edges and undirected edges with other attributes. While some attributes are reflected in many different charts, there are also some attributes that are specific to certain types of charts. The more long-used directed and undirected edges of relational diagrams clearly fall into the latter category. We wanted to find a rubric to classify these two attributes. Specifically, we ranked the expressiveness of different attributes separately and created a validity hierarchy table in Table 3.

3.4 Mapping Visual Variable's Attributes to Data Visualization Solutions

We have successfully established the mapping of user requirements to visual attributes. In the last mapping, we need to link the user requirements to the visualization solutions, i.e., the mapping between attributes and visualizations. Different visualization attributes have different differentiation capabilities among each other, as explained in Table 3. With this ranking in mind, we started analyzing how these attributes function in a complete visualization solution and how they combine with each other to convey different information ranges.

We extracted seven visualization variable attributes, and trying to traverse all the combination cases between them was impossible. On the other hand, detecting all the visual attributes contained in a visualization solution is a feasible way. This does the opposite of extracting visual variables from a complete visualization solution. To this end, we collected some existing cases of common visualization charts and grouped them into various chart types. In the process of finding all the matching attributes, we found that most attributes are usually

expressed in a regular way, no matter what type of chart. In order to improve the efficiency of the analysis, we summarized the following guidelines for reference.

- *Position*: chart with axes of coordinates
- *Size*: chart involving the display of area
- *Value*: chart involving different light and dark variations of the same color
- *Color*: chart involving the display of multiple colors
- *Geographic location*: chart with map display
- *Undirected edges*: chart involving the display of undirected edges
- *Directed edges*: chart involving the display of directed edges

Table 4. Attribute - Visualization Mapping.

| Attribute | Visualization |
|-----------------------------------|----------------------------|
| Position, (Color) | Bar, Line, Scatter Chart |
| Size, Color | Pie, Radar, Sunburst Chart |
| Position, Size, (Color) | Area, Bubble Chart |
| Color, Size, Value | Treemap |
| Geographical coordinates, (color) | Geographical, Route Map |
| Undirected edge, color, size | Relationship diagram |
| Directed edge | Tree graph |

The mappings between attributes and visualization schemes are established and shown in Table 4. We illustrate the application of the mapping in detail with an example of an area chart.

Area Chart with Attributes. The example chart we have selected is an area chart that displays the unaudited revenue of Apple Inc. The chart presents revenue values of the United States, Europe, China, Japan, and other Asia-Pacific regions for different quarters between 2014 and 2016. Users can easily compare the total revenue amounts in different quarters within a given region or compare the percentage of the total share of different regions at the same time.

To identify the visual variable attributes in this chart, we apply the summary rule described above. First, the chart includes two axes - the x and y axes - which allow users to find individual data and track their values. Therefore, the area chart includes the location attribute. Second, the chart represents the values of the data not only by the y-axis coordinates but also by the area of a region enclosed by the x and y axes. The areas are stacked in the ascending direction of the y-axis to show the total values and percentages in combination. Thus, the area chart includes the area attribute. The colors in the chart are used to distinguish different independent variables. However, when the data has only one independent variable, the color attribute is not meaningful. We also found examples of area charts that use a single color. Therefore, the color attribute is optional. Finally, numerical values, geographic maps, undirected edges, and directed edges are not addressed in the area chart.

In summary, area charts always include the location and area attributes and may include the color attribute at some point.

Bubble Chart with Attributes. The second example we chose is the bubble chart. The main purpose of this chart is to compare the life expectancy of different countries on different continents using the Gross Domestic Product (GDP) as a measurement factor. Additionally, the population of each country is also shown on the chart.

In the same way, we apply the general guidelines used to analyze the area chart to the bubble chart. We first focus on the two axes. The x and y coordinates determine the location of each bubble containing its data, so the bubble chart includes the location attribute. Then, the population is represented by the area of the bubble, making it include the area attribute. Countries in different continents are marked with different colors, providing a small visual grouping that is easy to read and understand. However, the grouping does not affect the presentation of the underlying data, so the color is an optional attribute. Finally, values, geographic maps, undirected edges, and directed edges are not addressed in the bubble chart. Therefore, we can conclude that the bubble chart also contains location, area attributes, and color attributes (optional).

Comparing the area chart and the bubble chart horizontally, we find that both share the same attributes. Furthermore, comparing the service purposes of both charts, we find that both are also used for “comparisons”. This further validates that our mapping relationship is potentially valid.

4 Prototype Tool

Our approach is implemented as a tool that automatically generating a series of visualization solutions. The tool follows the four-layer framework and prompts users to input their data and select their requirements. Once the user inputs their data, the tool generates all the matching charts by applying the three mappings.

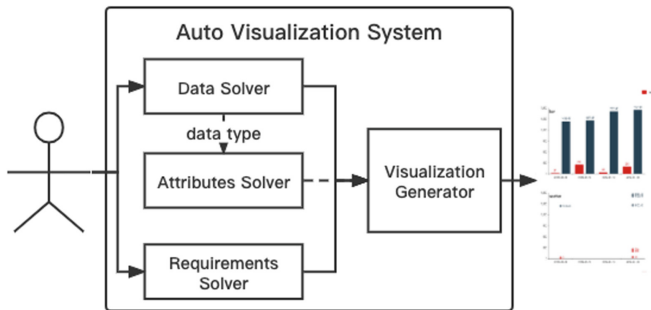


Fig. 4. System architecture.

The tool is a web-based application that runs on Python and Flask. The system architecture is shown in Fig. 4. To run the tool locally, the user can open and run the project in PyCharm, and install the Flask package. After that, the user can open their browser and go to port 5000 to access the tool's homepage, which prompts the user to upload their data and proceed to the next page. The data solver is fully implemented in Python, allowing users to specify the data source and apply selection filters. Some interactive elements, such as HTML components and JavaScript, are used to guide users in entering their requirements. All visualization solutions templates are imported from the Plotly library [8].

5 Evaluation

After thoroughly reviewing related research in this domain, we realized that no other approaches provide the same functionalities as our proposed tool. As such, we cannot compare our approach to any baseline, but we can evaluate it through a suitable case study. To evaluate the effectiveness of our approach, we invited a participant to simulate a practical application of our system and conduct a practical case study. The participant had a clear goal of creating the most suitable visualization scheme.

This participant is a sports events enthusiast who, during his vacation, watched many games of the 2021 Tokyo Olympics. He wanted to create a poster to display the performance of each country in the Olympic Games, and he realized that using a chart would be a condensed and suitable representation for a poster. He searched for a dataset and found one with four columns: the country name, total number of medals, number of gold medals, and number of silver and bronze medals.

The participant uploaded the data and then selected the page, guided by the comparison hint to check the difference and size of the data, which is ideally suited for datasets with two or three groups of numerical data.

Based on the mapping shown in Fig. 2, the comparison matched quantitative, selective, and order characteristics. As mentioned in the previous section, we only consider the mapping of four attributes: position, size, color, and value, based on the non-relational data type recognized by the data solver. Therefore, quantitative mapping to position, selection, and order all correspond to size and value. Since the comparison has quantitative *AND* selective *AND* ordered characteristics, we perform an *AND* operation with the corresponding attributes as well, resulting in position, size, and value as qualified attributes. In the last layer of mapping, we don't consider charts containing color, geographical coordinates, directed edges, or undirected edges. Ultimately, Bar charts, Line charts, Scatter charts, Area charts, and Bubble charts are chosen to apply user data and present it to the users.

This time, our participant received five graphs (shown in Fig. 5). From the bar chart, this participant can directly compare the number of medals. Also, to compare the number of gold, silver, and bronze medals. However, our participant

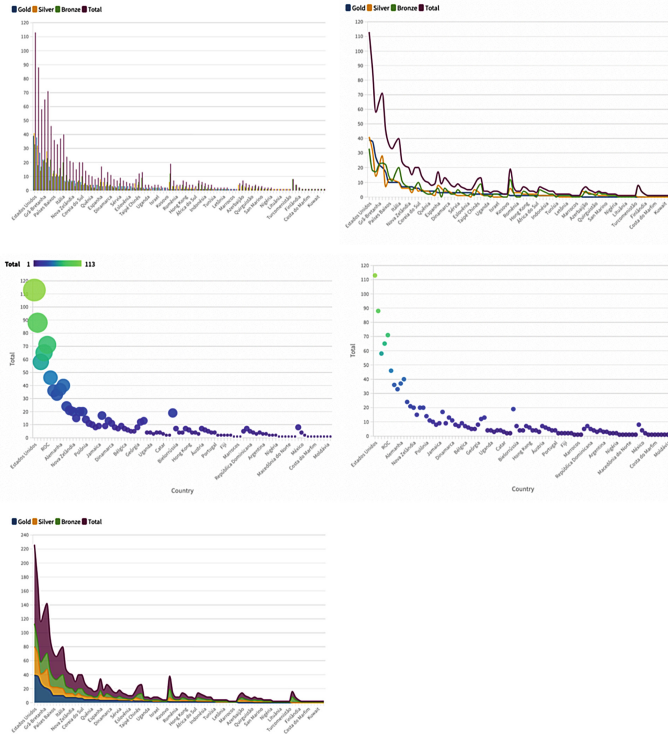


Fig. 5. Visualization solutions on the case study.

thinks it is not clear since there are too many countries, and each of them takes four column bars, making the x-axis too long. While in the line chart, he does not like the lines going through all the points because it makes the value of the points vague. Then in the scatter chart and bubble chart, he received similar results. The only difference between them is that there is one more dimension: “size”, used to present the value on top of “color”. Yet, in these two charts, only the total medal number is presented. Based on these two charts, our participant suggests using a different color to represent different medals. So we can have all the information separately shown in the charts. The reason why our system did not generate this chart is that it does not contain a color attribute. Lastly, the area chart uses a 2-dimensional visual symbol to present the number, so it expands the comparison value. The user commented that with the help of the stacked area, we greatly decreased the length of the x-axis. But ideally, in the area chart, we do not need to separately show the total number because the stacked area from all medals already reveals the value.

In the end, the participant put the area chart onto his poster. He preferred it because all the information is covered in the chart, and it suited better to the

poster size. He left high praise for the system because it helped him to get the most suitable visualization solution quickly.

6 Conclusion

In this paper, we propose an approach based on user requirements for generating visualization models to assist non-expert users in clarifying their data analysis goals and deriving the most appropriate visualizations to facilitate data analysis. Compared to other approaches, our framework covers the entire process, from defining and modeling user requirements to implementing visualizations. The greatest advantage of our framework is that it effectively links user requirements with visualization solutions, making it easier for non-technical users to obtain appropriate visualization solutions without requiring in-depth knowledge of visualization techniques or data source descriptions. Additionally, we list all visualization solutions that meet the requirements, giving the user flexibility in selecting the best solution that meets their needs.

Based on the insights gained from user interviews, we plan to address additional user requirements in our future work, such as predicting future trends in the data. Additionally, we can expand the visualization scheme by including more visualizations and establishing mapping relationships. Moreover, we intend to conduct more large-scale empirical evaluations of our approach to further assess its validity and usability.

Acknowledgement. This work is partially supported by the Project of Beijing Municipal Education Commission (No. KM202110005025), the National Natural Science Foundation of China (No. 62162051), and the Beijing Natural Science Foundation Project (No. Z200002).

References

1. Bertin, J.: *Semiology of Graphics*. University of Wisconsin Press (1983)
2. Bikakis, N.: Big data visualization tools. arXiv preprint [arXiv:1801.08336](https://arxiv.org/abs/1801.08336) (2018)
3. Buja, A., Cook, D., Swayne, D.F.: Interactive high-dimensional data visualization. *J. Comput. Graph. Stat.* **5**(1), 78–99 (1996)
4. Byrd, V.L., Dwenger, N.: Activity worksheets for teaching and learning data visualization. *IEEE Comput. Graph. Appl.* **41**(6), 25–36 (2021). <https://doi.org/10.1109/MCG.2021.3115396>
5. Chuai, Y., Yan, H.: Achieving the success of sustainability systemic design through data visualization approach. In: Trzcielinski, S., Mrugalska, B., Karwowski, W., Rossi, E., Di Nicolantonio, M. (eds.) *AHFE 2021*. LNNS, vol. 274, pp. 215–222. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-80462-6_27
6. Du, M., Yuan, X.: A survey of competitive sports data visualization and visual analysis. *J. Vis.* **24**, 47–67 (2021). <https://doi.org/10.1007/s12650-020-00687-2>
7. Griffin, T.: *Semiology of graphics: diagrams, networks, maps*. Bertin, Jacques (trans. W. j. Berg). The University of Wisconsin Press, Madison, 1983. 416 pages. ISBN 0 299 09060 4. US\$75.00, cloth. *Cartography* **16**(1), 81–82 (1987). <https://doi.org/10.1080/00690805.1987.10438353>

8. Hossain, S.: Visualization of bioinformatics data with dash bio. In: Calloway, C., Lippa, D., Niederhut, D., Shupe, D. (eds.) Proceedings of the 18th Python in Science Conference, pp. 126–133 (2019). <https://doi.org/10.25080/Majora-7ddc1dd1-012>
9. Hunter, J.D.: Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**(03), 90–95 (2007)
10. Manyika, J., et al.: *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute (2011)
11. Munzner, T.: *Visualization Analysis and Design*. CRC Press, Boca Raton (2014)
12. Ou, X., Zhu, Z., Chen, J., Xiao, W.: A study of data visualization of the neo-coronavirus pneumonia epidemic. In: 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), pp. 315–317 (2020). <https://doi.org/10.1109/ECICE50847.2020.9301964>
13. Peña, O., Aguilera, U., López-de Ipiña, D.: Exploring LOD through metadata extraction and data-driven visualizations. *Program* **50**(3), 270–287 (2016)
14. Ren, P., et al.: Intelligent visualization system for big multi-source medical data based on data lake. In: Xing, C., Fu, X., Zhang, Y., Zhang, G., Borjigin, C. (eds.) WISA 2021. LNCS, vol. 12999, pp. 706–717. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87571-8_61
15. Ware, C.: *Information Visualization: Perception for Design*. Morgan Kaufmann, Cambridge (2019)
16. Waskom, M.L.: Seaborn: statistical data visualization. *J. Open Source Softw.* **6**(60), 3021 (2021)
17. Wickham, H.: ggplot2. *Wiley Interdisc. Rev. Comput. Stat.* **3**(2), 180–185 (2011)
18. Wu, A., et al.: AI4VIS: survey on artificial intelligence approaches for data visualization. *IEEE Trans. Vis. Comput. Graph.* **28**(12), 5049–5070 (2022). <https://doi.org/10.1109/TVCG.2021.3099002>