



# Research on Data Visualization in Different Scenarios

Xin Guo<sup>1</sup>, Mingshu He<sup>1(✉)</sup>, Mo Chen<sup>2</sup>, Xiaojie Zhao<sup>3</sup>, and Ye Tian<sup>2</sup>

<sup>1</sup> Electronic Engineering Institute,  
Beijing University of Posts and Telecommunications, Beijing 100876, China  
hemingshu@bupt.edu.cn

<sup>2</sup> Institute of Network Technology,  
Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>3</sup> Information Science and Technology Institute,  
Beijing Normal University, Beijing 100875, China

**Abstract.** With the development of computer science, the era of big data has arrived. Facing the new era and new challenges, the traditional analytical methods of problems in various fields have been unable to meet the needs. Data visualization is a rapidly developing discipline, it has significant advantages in analyzing problems, so data visualization shines in the era of big data. As people are very concerned about the fields of energy and environment, we choose to conduct data visualization studies in two areas, energy and the environment. According to the different characteristics of data in different fields, we propose targeted data visualization processes and design data visualization solutions. For energy data, we follow the process of data processing, visualization design, and data visualization. Based on the principle of high efficiency and intuitiveness, we add timeline and a combination of various charts to our design, and finally show a dynamic effect. We also propose a multi-dimensional visual mapping visualization scheme. The scheme can refine and enrich the visual results. For environmental data, we follow the process of goal analysis, data processing, visualization and analysis, the work shows the importance of visualization in information analysis and decision-making.

**Keywords:** Data visualization · Timeline · Multi-dimension · Decision-making · Energy · Environment

## 1 Introduction

### 1.1 Data Visualization

At the 2017 IEEE Pacific Visualization Conference, Ebert [1], a visual analysis expert from Purdue University, gave a keynote speech titled “Changing the World with Visual Analytics”. In his speech, he pointed out that in order to solve the challenges in the world, we not only need to advance computer science and big data analysis, but also need a new analysis and decision-making environment. We must effectively combine human decision-making with advanced, guided analysis. And conduct human-computer cooperation discussions and decisions.

This method is data visualization. Data visualization uses intuitively graphical images to present data information in front of people's eyes, enabling people to effectively extract useful information from the original complex data, and perform data analysis in a more intuitive way. It can help humans find the correlation between data in order to make the right decision [2]. Data visualization analysis technology is an important method of big data analysis, which can help data analysts find the rules and patterns implied in data more quickly [3]. Data visualization technology needs the support of modern computer technologies, such as multimedia technologies, and mobile intelligent terminal technologies [4]. With the development of visualization tools, various kinds of visualization works tell us stories of various data [5]. In the field of visualization, science, technology and art are perfectly integrated. Data visualization mainly includes seven steps: acquiring data, analyzing data, filtering data, mining data, displaying data, summarizing data, and human-computer interaction [6].

## 1.2 Application Area

The advent of the computer age has injected new ideas for data visualization and provided performance capabilities and efficiency that cannot be achieved in the hand-painted era. Compared to various numbers, forms, texts and other information, the bright and intuitive graphical forms are more easily accepted. The science of data visualization was quickly applied to various industries.

The air pollution problem in China is becoming more and more serious. The air quality parameters can be acquired through the network. However, we still lack the visual display of air pollution data, lack the comprehensive chart of various parameter information that affects air quality over a period of time, and cannot display the distribution law of air pollution [7].

Based on the above, the data visualization work in this paper will focus on the following two areas: energy and the environment.

## 1.3 Our Work

This paper will basically follow the general steps of data visualization, plan the data visualization process for different application scenarios, and design a visualization scheme based on this scenario. This paper proposes visual design schemes for different data formats and presents them.

In the field of energy, as today's energy shortage, taking the path of sustainable development is an inevitable choice. Visual analysis of energy data is essential [8]. Li [9] and others proposed using the three-dimensional model CityGML to visualize urban energy consumption. Wang [10] and others proposed to visualize the urban power system by using GIS-based visualization. Both programs are based on geographic information, visualize and analyze energy data. For different emphasis, this paper focuses on the characteristics of energy data itself. To present multi-dimensional data, a visual mapping scheme for multi-dimensional data was proposed. In the case that geographic information is very important, according to the point made by Xu [11] and others, the map visualization technology has good geospatial features, which can

ensure that various statistical information is better displayed. This paper will be combined with the map for visual analysis at the right time.

Brehmer [12] proposed that the timeline is often used to describe events. Events can be recorded as data. This paper combines the data and timeline in the research scenario and proposes a dynamic and efficient visualization solution.

In the field of environment, many experts and scholars at home and abroad have used statistical analysis methods on air quality, such as using Spearman rank correlation analysis. Compared to statistical analysis methods, the results of air quality visual analysis are easy to read and understand [13]. This paper will propose a different idea of visualization from the energy field. We will start from the analysis of targets, and then proceed with data processing and data visualization. This case illustrates that data visualization is not only used for the presentation of results, but also promotes research.

The chapters are arranged as follow. Section 2 introduces the visualization based on energy dataset. Section 3 introduces the visualization in environment area. Section 4 summarizes the achievements of this paper.

## 2 Energy

In the field of energy, energy consumption includes direct energy consumption and embodied energy consumption. Direct energy consumption refers to the energy consumed in the manufacture of products and services. Embodied energy consumption refers to the total amount of energy consumed in the process of production, transportation of products, services and their destruction [14]. At present, we have obtained terminal energy consumption, also called direct energy consumption, in 30 provinces and 6 industries over the past 8 years. We also have the input-output table, which is the embodied energy consumption among provinces in monetary expression. For the energy data, the data will be visualized according to the data processing, visualization design, visualization process.

### 2.1 Data Processing

From the energy balance sheet, the direct energy consumption of each of the 6 industries in 30 provinces was obtained. The energy consumption table shows the consumption of various energy sources. In order to analyze the direct energy consumption among provinces and industries, we should have a unified scale. That means, the various types of energy consumption should be uniformly converted to standard coal consumption by using the following formula.

$$d_i = \sum_k C_{ki} \times d_{ki} \quad (1)$$

where  $d_i$  represents the direct energy consumption of department  $i$ , expressed as standard coal consumption.  $C_{ki}$  represents the conversion factor of the  $k$ th energy consumption converted to standard coal.  $d_{ki}$  represents the direct energy consumption of department  $i$ 's  $k$ th kind energy source.

Based on the principle of ecological input-output table, the embodied energy flow between industries is calculated. According to the principle of conservation of energy, embodied energy and direct energy flowing into an industry follow the following relationship, as shown in the Fig. 1.

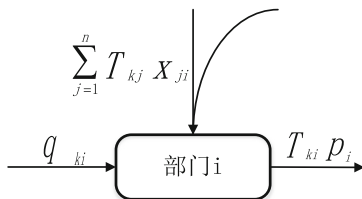


Fig. 1. Direct energy and embodied energy flowing

where  $q_{ki}$  indicates the direct consumption of the  $k$  kind energy in department  $i$ .  $x_{ji}$  denotes the intermediate input of department  $j$  to department  $i$  in the input-output table, which is the monetary expression of embodied energy.  $T_{kj}$  represents the  $k$ th energy consumption intensity of department  $j$  and is defined as the  $k$  energy amount contained in the unit output produced by department  $j$ . Therefore,  $T_{kj} \times x_{ji}$  represents the amount of  $k$  energy implicit in department  $j$ 's money flow to department  $i$ .  $p_i$  is the total output of department  $i$ .  $T_{ki} p_i$  represents the  $k$  energy consumption contained in the total output of department  $i$ . According to the law of conservation of energy, the total input of the  $k$  energy of department  $i$  should be equal to the total output of the  $k$  energy. The formula is as follow.

$$q_{ki} + \sum_{j=1}^n T_{kj} \times x_{ji} = T_{ki} p_i \tag{2}$$

Based on the above, the implicit consumption intensity of the  $k$  energy in department  $j$  can be obtained on the basis of the economic input-output table and the energy balance table. So, we can obtain the amount of embodied energy that flow between all the departments by combing with input-output table. Similar to the problems faced by direct energy consumption, in order to have a unified measurement scale, conversion factors are used to convert various types of energy consumption into standard coal consumption and then sum up. So far, we get results similar to the input-output table, named the embodied energy table. This table shows the energy flow relationship that is expressed in standard coal.

The relationship of energy often reflects the degree of economic development between cities. We divide the provinces into several business circle by clustering algorithms. The four business circles are the Hebei Business Circle, the Guangdong Business Circle, the Anhui Business Circle and the Shaanxi Business Circle.

## 2.2 Visualization Design

**Timeline.** In the face of data visualization in a series of time, we first consider the dynamic display combined with the time axis.

In the dynamic visualization of energy data, histograms are favored for their intuitive and efficient features. In the dynamic display designing, the energy data itself is taken into consideration, and geographical information is weakened. The energy consumption of various industries in each province changes with timeline of the year through the histogram. At the same time, in order to analyze the changes of direct energy consumption from the perspective of the industry of the whole country, a multi-diagram linkage design scheme was proposed. The pie chart and the bar chart were simultaneously displayed and changes with time like the Fig. 2.

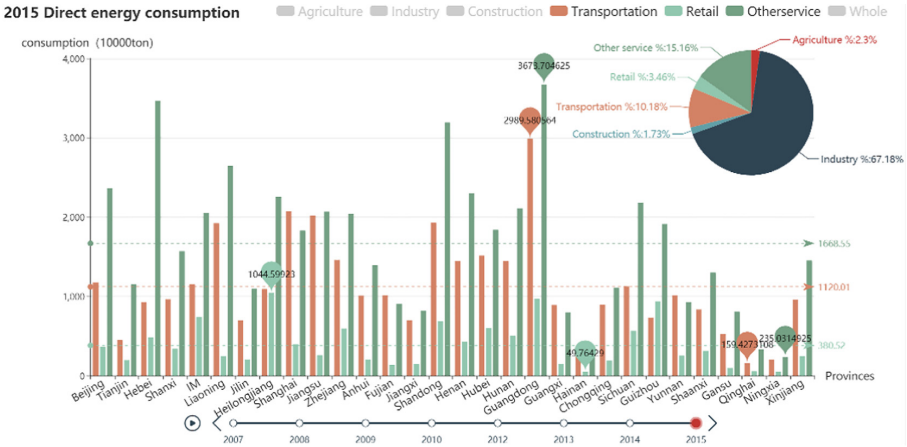


Fig. 2. Data visualization based on timeline

**Combination of Multiple Charts.** Multi-dimensional data information can be displayed through a variety of display methods combined. Like Fig. 3.

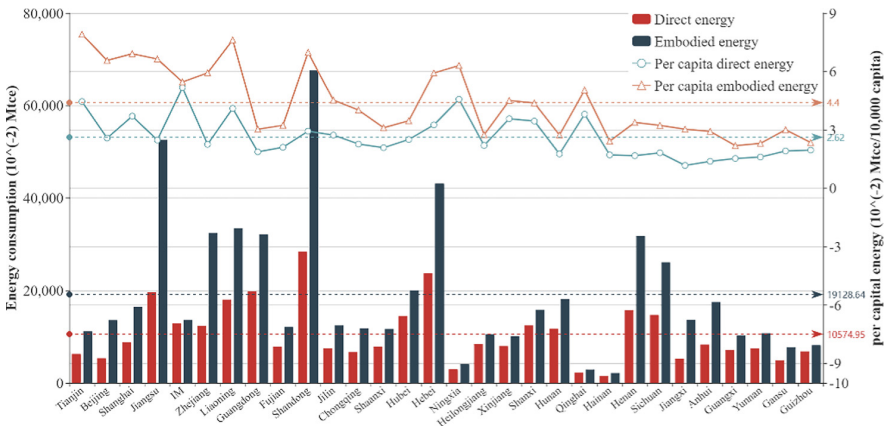


Fig. 3. Combination of bar and line charts

### 2.3 Multi-dimensional Visual Mapping

The visualization of bar and line charts is relatively simple. They are intuitive and efficient in some limited data. However, for data with multiple dimensions, the data visualization based on several axes only appears to be thin. The following will introduce the idea of multi-dimensional data information in visual mapping.

**Scatterplot Multidimensional Visual Mapping.** Although the scatter plot still only has two data axes, multi-dimensional information can be displayed in a scatter plot through visual mapping.

The Fig. 4 shows an example. The horizontal axis represents the embodied energy consumption, and the vertical axis represents the direct energy consumption. The size of the scattered dots can represent the level of GDP per capita. The labels of the scattered dots can distinguish among provinces, even the colors of scattered dots can also be used to display population density or other information (the scatter color in the Fig. 4 is only used to distinguish provinces). The auxiliary lines in the scatter plot can help users to quickly determined the level of the energy consumption of a province compared with the national average. And it's easy to see the proportion of direct energy consumption and embodied energy consumption in a province. Multi-dimensional information is mapped to a variety of visual information, then enrich the visual information.

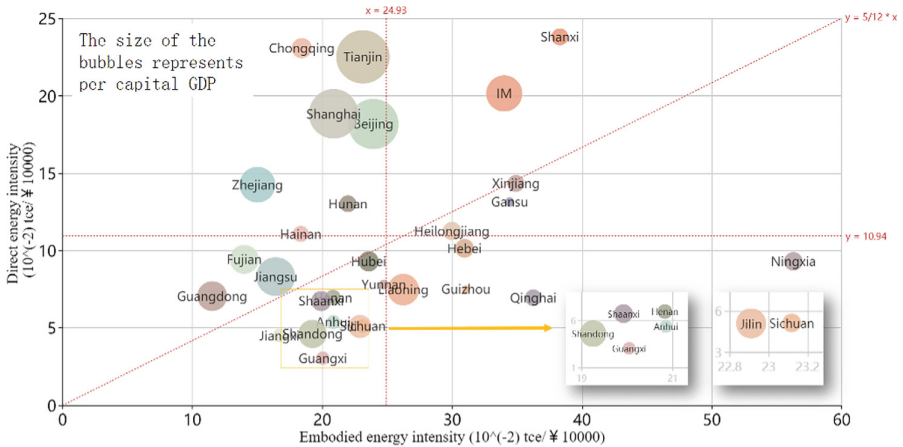
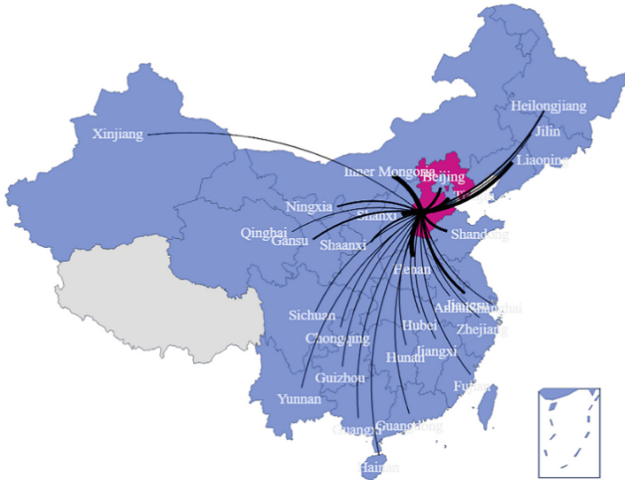


Fig. 4. Scatter plot multi-dimensional visual mapping (Color figure online)

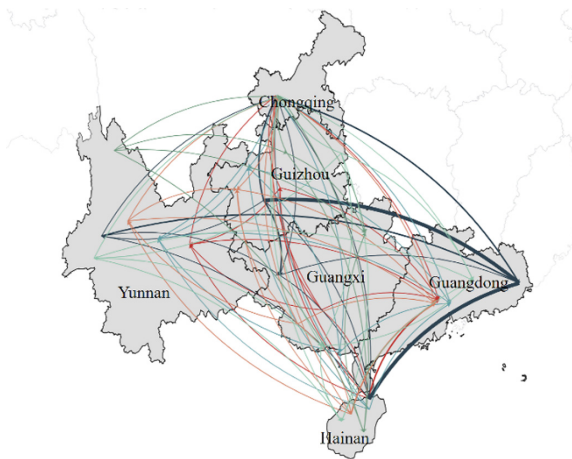
**Map Multi-dimensional Visual Mapping.** The same idea can be applied to map-based visualization.



**Fig. 5.** Map1 multi-dimensional visual mapping

Figure 5 shows the country’s embodied energy transfer. The thickness of the line represents the amount of embodied energy transfer.

In the visualization of the business circle obtained by clustering, the multi-dimensional visualization based on maps is applied more deeply. Figure 6 shows the visualization of the Guangdong Business Circle. The six colors represent the transfer of embodied energy among the six major industries, and the thickness of the lines represents the amount of embodied energy transfer. The multi-dimensional information is displayed in a single picture, making the visual information richer and the visual results concise.



**Fig. 6.** Map2 multi-dimensional visual mapping (Color figure online)

### 3 Environmental Data on Air Pollution

#### 3.1 Target Analysis

At present, environmental monitoring data of some environmental monitoring sites in Beijing, Hebei, Henan, Shandong, Shanxi and so on have been obtained during a certain period of time.

The environmental monitoring data includes the monitoring time, the name of the monitoring site, the concentration of various environmental pollutants and so on. It is a data set with both time and geographic information, and has a variety of data.

The visualization method can be applied to the following three stages in the analysis of environmental data: the early stage (the stage of collating, filtering and cleaning of information), the intermediate stage (information analysis activity), and the later stage (display of information analysis results). Before carrying out these three stages, the target analysis was performed on the PM2.5 data. It can help us find the ways of analyzing the internal laws. The specific process is shown in Fig. 7.

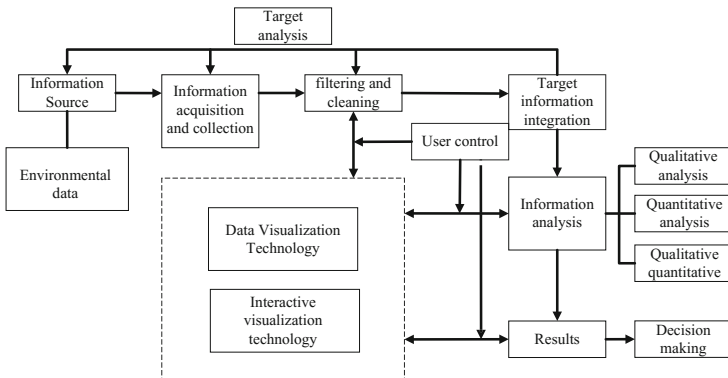
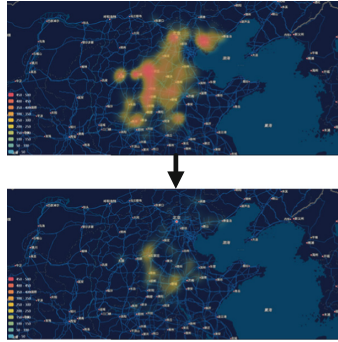


Fig. 7. Process of environmental data visualization

In this paper, the PM2.5 data from the environment monitoring is dynamically visualized to qualitatively analyze the trends of PM2.5 during this period of time. At the same time, we will quantitatively analyze the correlation of PM2.5 among regions and the relationship between the correlation and regional distance. We visualize the results in order to get a prediction. We visualize a variety of environmental data in the form of multiple axes, qualitatively and quantitatively analyze the data.





**Fig. 8.** Visualization of original PM2.5 data

### 3.2 Data Processing

In the original environment data, there is a considerable portion of redundant data, including no information data, duplicate data, contradictory data, and so on. This part of data not only drastically increases the amount of information, but also brings unnecessary trouble to information analysis. The visualization of the PM2.5 original data can be very obvious to show this kind of trouble, and through visualization, we can quickly discover this problem. Therefore, data visualization can help users quickly find a problem or abnormal situation.

From the visualization result (Fig. 8), the monitoring sites at each time in the original data have certain differences, that means, the monitoring sites at this moment are not the same as the moment before. Due to this defect, the intuitive perception of the environmental change trend becomes deceptive. The visualization result shows that the environment is better when compared to the previous time, however, this conclusion may be because of the fact that many sites have not been measured at this time, not because of the real environmental changes.

We filter the data at different times of 200 environmental monitoring stations from the original data. It includes monitoring time, monitoring locations, and environmental pollutant values. For situations that with timeline and geographic information, a map-based timeline visualization scheme is proposed. In order to achieve the dynamic qualitative analysis in the target analysis and the quantitative analysis of the correlation of sites, the following processing and expansion of the data is still required.

We obtained the latitude and longitude coordinate information based on Baidu coordinate system to achieve dynamic visualization analysis with maps qualitatively. The linear distance between sites is calculated from latitude and longitude information by using the Haversine formula to analysis the relationship between the PM2.5 correlations and the sites' straight-line distances quantitatively.

### 3.3 Visual Display

A part of the timeline visualization based on the map is shown as Fig. 9.

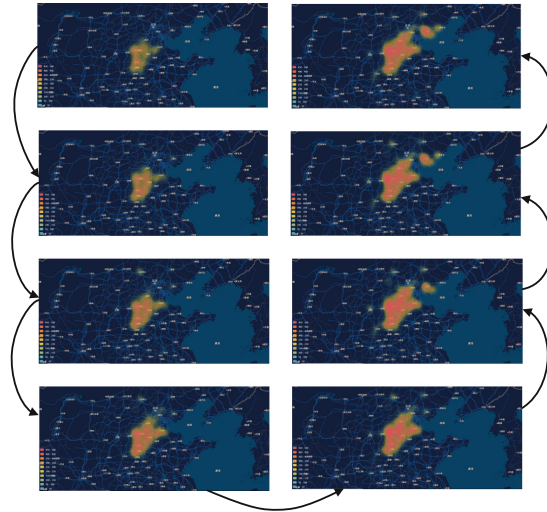


Fig. 9. Dynamic visualization based on map and timeline

Through this dynamic visualization combined with geographical location, we can easily qualitatively analyze the regularity of PM2.5 data. Dynamic visualization combined timeline with geographic information is an efficient and clear visualization solution. We discovered the law of change and further explored the underlying reasons.

We visualize the correlation between a part of the sites as Fig. 10.

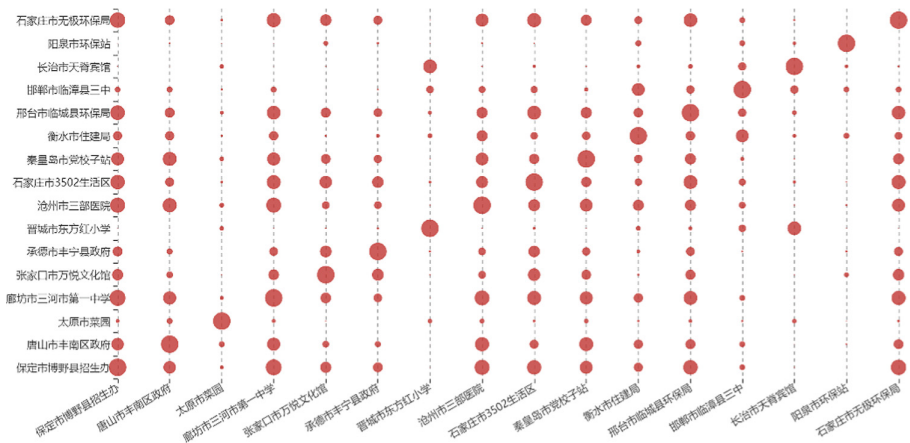
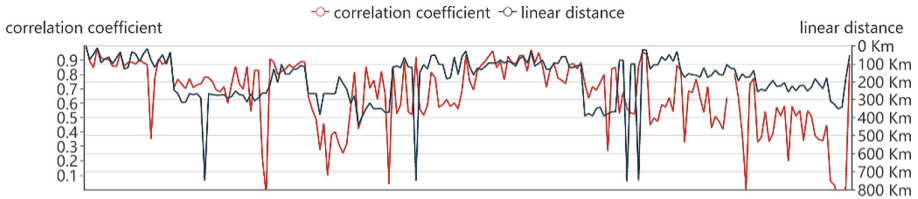


Fig. 10. Correlation between several stations (Color figure online)

In Fig. 10, the correlation between sites is mapped to the size of red circles so that the degree of correlation between a station and another station can be intuitively obtained. We can see the size are different. So, the correlation data visualization also brings us another question: what is the reason for this and what is the relationship between correlation and distance of stations.

In order to further solve this problem, we visualize the correlation coefficients and linear distances in one picture as Fig. 11.



**Fig. 11.** Visualization based on correlation coefficients and linear distances (Color figure online)

From the Fig. 11, the red line that represents the correlation coefficients and the green line that represents the linear distances, although have different positions of the peaks, have the same trend. The visualization results indicate that most sites follow a mathematical rule that the distance between them is longer and the correlation coefficient is smaller.

Thus, in information analysis activities, visualization is not only used as the last step before the analysis of laws, or as a display of research results, but also a key tool to promote the analysis activities. Starting from the information which is presented by visualization on complicated data, new issues are further proposed to promote research.

## 4 Conclusion

This paper concentrate on the areas of energy and environment, puts forward the principle of efficient and rich visualization, adjusts the process of data visualization under different scenarios, and designs visualization schemes under different data characteristics. For types of datasets with time series, timeline-based visualization is more efficient and richer, such as the visualization of direct energy consumption in different provinces and industries in the field of energy. For the datasets that contain the geographic information, if the geographic information is important, the data visualization can be based on maps to display geographic and data relationships intuitively, such as the situation of energy transfer in commercial circles. For multi-dimensional information, make full use of visual mapping, visual coding. Refining and enriching visualization information, such as visualizing multidimensional energy and economic data information. Flexible and appropriate visual design will present great value.

**Acknowledgments.** This research is funded by The National Science Foundation of China (61601053) and National Natural Science Foundation of China (Grant No. 61602051) and the Fundamental Research Funds for the Central Universities under Grant 2017RC11.

## References

1. Ebert, D.: Changing the world with visual analytics. In: 2017 IEEE Pacific Visualization Symposium, Seoul, South Korea, p. 8. IEEE Computer Society (2017)
2. Chang, Y.: The data visualization method in the era of big data. *Electron. Technol. Softw. Eng.* **2018**(5), 156 (2018)
3. Qiao, S., Nurbol, Su, R.: Map analysis for research status and development trend of data visual analysis. *Mod. Electron. Tech.* **41**(14), 161–169 (2018)
4. Chang, Y.: Visual analysis of web data under big data background. *China CIO News* **2018** (5), 148–150 (2018)
5. Lee, B., Riche, N.H., Isenberg, P., Carpendale, S.: More than telling a story: transforming data into visually shared stories. *IEEE Comput. Graphics Appl.* **35**(5), 84–90 (2015)
6. Dai, S., Dong, J., Xue, J.: Big data visualization analysis and application in scientific computing. *J. Eng. Stud.* **6**(3), 275–281 (2014)
7. Ma, C., Li, L., Xue, W.: Research on visualization of air pollution characteristics and distribution. *J. Northwest. Polytechnical Univ.* **35**(6), 1073–1078 (2017)
8. Chen, S.: Analysis of the evolution of energy consumption structure in American four states. *Chem. Enterp. Manag.* **2018**(11), 30–32 (2018)
9. Li, B., Mao, B.: Energy consumption analysis and visualization based on CityGML—a case study of Swedish Smart City. *Geomatics World* **24**(4), 48–52 (2017)
10. Wang, Y., Yu, J., Wu, F.: Data analysis and visualization based on geographic information in the background of Global Energy Internet. *Electr. Power Inf. Commun. Technol.* **14**(3), 49–54 (2016)
11. Xu, N., Luo, J.: Statistical analysis of data map visualization. *Technol. Innov. Appl.* **2018** (22), 67–68 (2018)
12. Brehmer, M., Lee, B., Bach, B., Riche, N.H., Munzner, T.: Timelines revisited: a design space and considerations for expressive storytelling. *IEEE Trans. Visual Comput. Graphics* **23**(9), 2151–2164 (2017)
13. Wang, R., Zhou, M., Wang, Y., Liu, Y.: Mapping spatial and temporal patterns of air condition in Northeast China. *Bull. Surveying Mapp.* **2017**(8), 88–91 (2017)
14. An, Q., An, H., Wang, L.: Analysis of embodied energy flow network between Chinese industries. *J. Syst. Eng.* **29**(6), 754–762 (2014)