

Research on Visualization Method and Application of Government Statistical Data

Bian Zhifeng

Xi 'An University of Finance and Economics, Xi 'An 710100, China

Keywords: Government statistics, Data, Visualization

Abstract: The convergence of information technology and economy and society has led to the rapid growth of government statistics. From the perspective of government statistics, this paper discusses the significance of visualization of government statistics, and introduces the related technologies, visualization methods and main tools of visualization of government statistics. An empirical study is made with practical cases, and it is pointed out that the visualization technology of modern government statistical data can discover and reveal the deep meaning of data, and mine the hidden relationship between data, which is of great significance to improve the government statistical service function.

1. Introduction

In recent years, with the rapid development of social economy, the government statistical data and statistical information have increased rapidly in geometric progression. How to deal with data information and make better use of data information has become a problem faced by national statistical departments, government decision-making departments and the public. The traditional application of statistical data is limited to publishing statistical indicators and statistical reports. The basic statistical methods (average, median, range, etc.) do not conform to the cognitive nature of human beings. It is difficult to see the laws according to statistical methods, which can no longer meet the needs of social development.

The visualization technology of government statistical data can use computer graphics and image processing technology to transform invisible data phenomena into visible graphics or images and display them on the screen to enhance the presentation effect of data; The complicated and seemingly unintelligible data are connected and related, and the rules and characteristics are discovered. It is convenient for the government and people to observe, understand and interpret statistical data in a more intuitive way, and then discover the hidden information, knowledge and wisdom in statistical data, helping people to draw the required value from the vast data, integrating information from massive, dynamic, uncertain and even conflicting data, gaining a deeper understanding of complex statistics and economic situation, and realizing the purpose of data self-interpretation and making data speak. At the same time, it provides rapid, verifiable, easy-to-understand evaluation and more effective means of communication. It is of great significance to improve the government's work efficiency, effectively help the public sector to optimize decision-making and improve the government's statistical service function.

2. Statistical Data Visualization Related Technologies

2.1 Web Crawler Technology

Web crawler technology, also known as “web spider”, refers to the technology that a computer program automatically obtains target web page information or scripts in accordance with a certain established process. He can help search engines download the required information on the WEB page, and collect and analyze data in an automatic loop. Finally, the unified storage of data is completed. When the computer program is in the normal operation stage, the crawler program will download the content of the web page from one or more initial target URLs, and then rely on the search method or content matching method to accurately “crawl” the required content of the web page. At the same time Crawlers will also continuously obtain new URLs from web page hyperlinks. When the information retrieved by the crawler meets the stop condition, the retrieval will automatically stop. At this time, it will automatically enter the processing link of collected data. The crawler program builds a database, indexes and properly stores the collected data, so that users can extract and consult various data in the database at any time according to their actual needs.

2.2 Python Language

Python is an interpreted language, which has the characteristics of simple and easy to learn, high-level language, free and open source, strong portability, rich library, object-oriented, extensible, embeddable and standardized code. With the rapid growth of statistical data, using Python to process statistical data is the most popular way. Compared with traditional statistical tools SAS, SPSS, etc., Python has more prominent features and advantages for statistical data visualization: First, it is simple and easy to learn, and has a high degree of popularity. Python has been included in middle and high school education at home and abroad for popularization. The second is that Python is faster than traditional tools in terms of big data processing performance. It does not need to split big data into several small data to load and process like traditional tools, but can directly load and process GB-sized data. Third, the open source ecosystem is active and rich in functions. With the continuous development of Python extension libraries, Python has powerful libraries, frameworks and solutions in scientific research, electronics, government, data analysis, web, finance, image processing, AI technology and other aspects.

The Python-based web crawler technology, due to the use of Python to write programs, can abandon the traditional cumbersome IDE, and only use a text editor to basically complete the development of web crawler technology functions, which provides great convenience for the programming work of technicians. In addition, Python itself has a relatively complete crawler framework, which can support program tasks such as rapid mining, extraction and processing of information and data. With the full use of Python’s powerful network capabilities, even in the face of massive information and data retrieval requirements, only a few codes can be used to effectively complete web page downloads and use its web page parsing library to accurately interpret and express each web page tag to effectively improve crawling data quality level.

2.3 Database Storage

MySQL is a relational database developed by a Swedish company. It is open source, compact, fast, fully functional, and low in cost of ownership. Generally, before opening a database connection, make sure to open the MySQL service function, otherwise the database connection will fail. SQLyog is an easy-to-use, fast and concise graphical management tool for MYSQL databases. It can effectively manage your database at any location, with features of simplicity, ease of use, and

graphics. In order for us to successfully connect to the database in Python, we have to set up the relevant configuration of the database connection, including the address, port number, user name, password, etc.

3. Statistical Data Visualization Methods and Main Tools

Visualization methods for government statistical data are generally divided into two categories: downloading and importing data for visualization and web crawling data to organize and visualize. Here we mainly discuss the visualization after web crawlers crawl the statistical data.

3.1 Statistical Data Visualization Method

According to the goal of statistical data, we need to determine the website and data type to be crawled; analyze the web page through Python crawler technology; store the data characteristics in the database after describing the data, so as to extract, query, and add data at any time; for the acquired data perform visual analysis and get our conclusions.

3.2 Main Tools for Statistical Data Visualization

Government statistics are generally provided in the form of database tables. Commonly used statistical visualization libraries in Python include ECharts, HighCharts, G2, Chart.js, etc. Most of these tools are free and open, and can basically meet the needs of government statistics visualization requirements for the display and analysis of statistical data.

1). Echarts

Echarts (<https://echarts.baidu.com/>) abbreviated from Enterprise Charts, is a pure Javascript data visualization library of enterprise-level charts, open source from Baidu data visualization team. It is usually used as a statistical chart module of web pages. It can run smoothly on PCs and mobile devices, and is compatible with most current browsers (IE8/9/10/11, Chrome, Firefox, Safari, etc.), and the underlying dependency is also on the other the lightweight Canvas library ZRender independently developed by the team provides intuitive, vivid, interactive, and highly personalized data visualization charts. Innovative features such as drag-and-drop recalculation, data view, and range roaming greatly enhance the user experience and give users the ability to mine and integrate data. Users can customize the visualization chart on the Web as needed. Covering mainstream front-end frameworks and extensions of 8 programming languages, there are various types of charts and dynamic visualization effects.

Echarts supports line chart (area chart), histogram (bar chart), scatter chart (bubble chart), candlestick chart, pie chart (ring chart), radar chart (filled radar chart), chord chart, force-oriented layout 12 types of charts such as graph, map, dashboard, funnel chart, event river chart, etc., and 7 interactive components such as title, detail bubble, legend, value range, data area, timeline, toolbox, etc. are provided at the same time, and multiple charts and components are supported the linkage and mash up show. In addition to the built-in charts that contain rich functions, ECharts also provides custom series. You only need to pass in a renderItem function to map the data to any graphics you want. What's better is that these are still and have some preexisting interactive components used in combination without the need to worry about other things. You can download the build file that contains all the charts on the download interface. If you only need one or two of the charts, and the build file that contains all the charts is too large, you can also customize the build after selecting the desired chart type in the online build.

2). HighCharts

HighCharts (<https://www.highcharts.com.cn/>) is a chart library written in pure JavaScript, which

can easily and conveniently add interactive charts to web sites or web applications. This is the most widely used chart on the Web, free for personal study, personal website and non-commercial use. HighCharts supports all mainstream browsers and mobile platforms (iOS, Android, etc.) and multiple devices (devices, tablets, etc.) with simple configuration to support dynamic multi-dimensional charts, statistical visualization results support exporting PDF/PNG/JPG/SVG, etc. various formats, and you can zoom in on the selected part of the chart to support close observation of the chart. The time HighCharts on the timeline can be accurated to milliseconds, which is the simplest and most flexible chart library on the market. HighCharts supports bullet.js, histogram-bellcurve.js, parallel-coordinates.js, pareto.js, sankey.js, streamgraph.js, sunburst.js, tilemap.js, variable-pie.js, variable.js, vector.js, windbarb.js, wordcloud.js, xrange.js and other 14 charts. HighCharts contains four major components: Highcharts, Highcharts Stock, Highcharts Maps, Highcharts Gantt, and supports the linkage and mashup display of multiple charts and components.

4. An Empirical Study on Visualization of Government Statistics

The extended library provided by Python itself contains a complete set of statistical functions and data processing methods, such as bar function for drawing bar graph, barh function for drawing bar graph, hist function for drawing histogram, pie chart function, etc. These functions can efficiently process massive statistical data matrix, and can easily slice, slice, aggregate and resample statistical data. Its rich functions and powerful algorithms have become the preferred solutions for data processing tasks. Next, we use Python and extended library to do empirical research on typical application scenarios in statistical data processing.

4.1 Grab Data

Taking domestic trade, the quarterly data from the homepage of National Bureau of Statistics, as an example, this paper makes a self-defined index query on the retail sales of consumer goods above designated size in 2020. The index includes retail sales of supermarkets above designated size, retail sales of department stores above designated size, retail sales of specialty stores above designated size and retail sales of specialty store above designated size. Set Python crawler code, save the data crawled by crawler as data file and store it in MySQL relational database.

4.2 Data Preprocessing

Data preprocessing refers to the process of data merging, data cleaning, data standardization and data transformation, which is directly used for statistical analysis and modeling. Among them, data merging can merge multiple interrelated tables into one; Data cleaning can remove duplicate, missing, abnormal and inconsistent data; Data standardization can remove dimensional differences between features; Data exchange can meet the data requirements of later analysis and modeling through discretization and dummy variable processing. In the process of statistical data analysis, the processes of data preprocessing cross each other, and there is no fixed sequence.

4.3 Statistical Description of Data

An important link in the visualization processing of government statistical data is the statistical description of relevant data of statistical targets, which mainly includes frequency analysis, concentrated trend analysis, discrete degree analysis, distribution and graphic visualization. Frequency analysis can be used to test abnormal values; The general level of data is reflected by

trend analysis in data set. The commonly used indicators are average, median and mode, etc. The dispersion degree analysis of data is used to reflect the difference degree between data. The commonly used indexes are variance and standard deviation.

4.4 Data Visualization

We use python-highcharts, a third-party library based on Python, to draw the analysis histogram.



4.5 Data Analysis

From the bar graph, the domestic retail industry has been affected by the outbreak of COVID-19 in 2020. Department stores, specialty stores, and specialty store all experienced negative growth, and only supermarkets achieved less positive growth, but it was far below a normal level. It directly illustrates the great impact of COVID-19 outbreak on domestic economy.

5. Summary

Modern government statistical work analyzes the economic and social development situation from a large number of statistical data through visualization technology and visual interpretation tools such as information charts, which enables the public to understand and recognize statistical data more deeply, strengthens the transparency of statistical data and improves the credibility of statistical data. Then it helps the government and the public to make more wise decisions, discover and reveal the deep meaning of each data in the data relationship, and mine the hidden data relationship between the data, presenting us with a new perspective in the context of economic development. And in the process of statistical data visualization, statistical data is not boring, but becomes more plastic, feasible and more “humanized”. Visualization of government statistical data represents an important direction of the development of government statistical data in the information age.

References

[1] Python crawler ultra-simple practical tutorial (I) | Crawling the data of National Bureau of Statistics [EB/OL]. <https://www.cnblogs.com/ifever/p/14275040.html>.
 [2] Web crawler and data visualization analysis based on Python [EB/OL]. https://blog.csdn.net/weixin_

43917062/article/details/107590191.

- [3] Xia Dehong. *Research on the Design and Implementation of JAVA Data Visualization [J]. Electronic World*, 2021,(06): 178-179.
- [4] Hu Qianfang, Lian Pengwei, Chen Qiankun. *Application of Python in Statistical Data Processing [J]. Market Research*, 2019(08):33-35.
- [5] Lu Xuan. *A preliminary study on the construction of government data visualization public service platform [J]. Electronic World*, 2018,(18): 79-80.
- [6] Guo Wei. *Research on Visualization of Statistical Data under Big Data Background [J]. Inner Mongolia Statistics*, 2017(06):23-27.