INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS

ISSN(print): 2643-9840, ISSN(online): 2643-9875

Volume 05 Issue 05 May 2022

DOI: 10.47191/ijmra/v5-i5-03, Impact Factor: 6.261

Page No. 907-910

# Research on Data Visualization Technology Based on Python

**Feng Li[1], Lingling Wang[2]**

[1,2]School of management science and Engineering, Anhui University of Finance and Economics, Bengbu 233000, China

**ABSTRACT:** This study aims to research on data visualization technology using Python language. Python is an open-source programming language developed with the community-based model. In this paper, different visual presentation types are introduced, such as visualization of text data, network visualization, and visualization of spatial information. Additionally, common data visualization techniques are presented in this paper. Finally, some data visualization tools (Smartbi, D3, Google Chart API, Processing, and Rapidminer) are described in this paper.

**KEYWORDS:** Visualization Technology, Python, Data Visualization Tools

## I. INTRODUCTION

Python is easy to learn as well because of its simple syntax, which has a powerful interactive network visual information management library with numerous information visibility optimization libraries. It is an open-source programming language developed with the community-based model. It's free to use, and since it's open-source supports multiple platforms and can be run on any environment. For example, 2D and 3D information visualization optimization libraries Matplotlib, Seaborn and Pandas, Folium, Basemap, MapBox, GeoPlotlib, PyechartsMap, etc. Information visualization management library of social Service Network networkX, Wordcloud, a library for information visibility optimization of dictionaries and cloud images, and WordCloud from Pyecharts.

The Matplotlib library is Python's third-party data visualization library and the most widely used data visualization drawing library in Python. It is very convenient to get the general information of the data. The Numpy library is Python's base for data processing and is the foundation of high-performance scientific computing and data analysis. Because it provides a large number of mathematical function libraries for data operation, it can be used to store and process large matrices, supporting a large number of dimensional arrays and matrix operations. Pandas is a well-known data analysis library for Python. It includes a large number of databases and standard data models and provides the tools needed to efficiently manipulate large data sets. The main function is to do a lot of data processing but also can draw the efficient completion of the drawing work. The Seaborn library is a graphical visualization Python package based on Matplotlib that allows to creation of beautiful charts with just a few lines of code. The advantage is that it provides a highly interactive interface for statistical visualization. It has a high-level interface that can draw attractive statistical graphics. The visualization results obtained from Seaborn are often faster and more beautiful. The Bokeh library is an interactive visualization library for modern Web browser rendering capabilities. It is the use of Python through a simple and fast way to provide high performance of the interaction of large data sets to get multi-functional visualization. Pyqtgraph is a pure Python graphical GUI library based on PyQt4/PySide and Numpy libraries. It is written entirely in Python. However, due to the internal use of the NumPy library and GraphicsView framework, it can carry out a large number of digital processing and rapid display [1].

## II. VISUAL PRESENTATION TYPE

At present, there are all kinds of data analysis and visualization applications on the network. These visualization technologies can be divided into three categories from the technical perspective: text data visualization, network visualization and spatial information visualization [2].

(1) Visualization of text data

Text visualization covers the process of information collection, text information mining, visual rendering and interaction design. The results of visual analysis such as tag cloud, text map, network graph and overlapping graph can be obtained by using word bag model, feature mapping, text clustering, lexical and syntactic analysis to text content or data

In the text visualization, the general is the visualization type of the chart class, such as bar chart, pie chart, broken line chart, bubble chart, etc., these are the visualization of the chart class. However, there are three types of visualization that are widely used in the field of text, namely, visualization based on text content, visualization based on text relations, and visualization based on multi-level information.

(a) Visualization based on text content. At present, this kind of visualization is mainly applied to word frequency and word distribution, and we are familiar with word clouds and various distribution maps.

(b) Visualization based on text relations. This kind of visualization is mainly to study the internal and external relations of the text so that people can easily understand the content of the text and find certain objective laws. We commonly have tree graph, node connection network graphs, and so on.

(c) Visualization based on multi-level information. This type of visualization mainly helps users to have a deeper understanding of textual data based on multiple aspects of the information. Nowadays, more and more attention is paid to text visualization that contains time information and geographic coordinates. Common visualizations in this category include geo-thermal maps, Spark Clouds, and matrix-based sentiment analysis visualizations.

(2) Network visualization

Network visualization is usually used to show the relationship between data in a network and is generally used to depict interconnected entities, such as social networks.

The most common tool for creating social network graphs is Python, and users need to access the Network library before creating social network graphs. As a graph theory and complex network modeling tool in Python language, the network library contains a variety of commonly used graphs and complex network analysis algorithms. Users can use the Python library to analyses complex network data, and can also use it for simulation modeling.

(3) Visualization of spatial information

Spatial information visualization refers to the process of graphing complex scientific phenomena, natural landscapes and some abstract concepts by using computer graphics and image processing technology. Cartography and computer graphics are often used in the process of spatial information visualization, and the process of spatial information visualization is to input the information of cartography, through inquiry, analysis and processing, and then combined it with the display of graphics and images, to achieve interactive processing and display in a concise and easy to understand visual form.

## III. COMMON DATA VISUALIZATION TECHNIQUES

In the application of information visualization technology, target classification methods are not technology-driven but target-oriented. Data visualization method based on object classification has been widely used in the industry. Abstract the goals of data visualization into comparison, distribution, composition, and relationships [3, 4]. Common data visualization tools can be divided into the following categories:

(1) Parallel visualization

Parallel visualization usually includes three parallel processing methods: task parallelism, pipeline parallelism, and data parallelism. Task parallelism divides the visualization process into independent sub-tasks, and there is no data dependence between the sub-tasks.

The visualization process is divided into several stages. The computer executes each step in parallel to speed up the process. Data parallelism is a parallel processing method of "single program and multiple data". It divides data into several subgroups and then executes programs to process different subdata in parallel with the granularity of subdata.

(2) In situ visualization

The numerical simulation process can intuitively eliminate bottlenecks in large-scale simulation output.

Raw materials can be divided into image, distribution, compression, and characteristics. The output is treated as the raw image, while the data is treated as visual and image storage in digital simulation. Output - Visual display of raw location-allocation data, according to user-determined statistics, calculation and retention of statistics during numerical simulation, and subsequent statistical reports;

Using the compression algorithm to visually display the output of the compressed data reduces the number of analog data output, and the compression reduces the number of input for subsequent visual processing.

The visual method used to show the starting position of the output is to extract characteristics, which are extracted and preserved during the digital simulation and subsequent visual processing of input data.

(3) Time series data visualization

Providing data in chronological order helps people to look at the past from a statistical perspective and predict the future, for example, to build prediction models and analyze user predictions and behaviors.

The area chart shows the changes and changes in numbers over a period of time, which is the most common trend. In the bubble chart, you can set a variable in the animation process of an axis as time or change a data variable. Candle patterns are often used as a tool. Histograms are often used as a project management tool, thermometers show color changes in data, and histograms are suitable for showing data allocation between sequential intervals or specific time intervals. Use line charts to show continuous time intervals or the number of time intervals, trends and relationships most often displayed.

## IV. DATA VISUALIZATION TOOLS

Data visualization analysis tool, since it is a tool to analyze data, it must have the ability to deal with massive data and graphical display and interaction. Therefore, it needs to be able to quickly collect, screen, analyses, summarize and show the information needed by decision-makers, and update it in real-time according to the newly added data [5].

### Smartbi

Smartbi supports Excel as a report designer and is perfectly compatible with Excel configuration items. Support Excel all built-in graphics, background, conditional formatting and other design complex dashboard style. Excel plug-in functions all Excel graphics such as feature graphics: mini, Pareto, bullet, small and many, etc. Commonly used graphics, such as column chart, pie chart, line graph, radar chart, etc., and combined with the dynamic data in the data warehouse for data display.

Smartbi supports a complete ECharts graphics library, supporting a variety of graphics, including dozens of dynamic interaction graphics such as waterfall diagram, relationship diagram, radar diagram, oil volume diagram, thermal map, tree diagram, etc. Support 3D dynamic graphics effect, such as 3D route chart, 3D scatter chart, 3D column chart for data visualization display; Support rich Echarts graphics controls such as wheel cast controls, running lights, TAB controls, URL controls, can be directly using Echarts all options configuration; Integration with other HTML5 graphical controls is also supported.

### D3

D3 is another Java library that supports SVG rendering. However, D3 offers a wide range of complex chart styles beyond linear and bar charts, such as Voronoi charts, trees, circular clusters, and word clouds. D3.js, which stands for Data-driven Documents, uses HTMLCSS and SVG to render amazing charts and analysis diagrams. D3's emphasis on web standards is strong enough to make it possible to use on all major browsers without being tied to other types of architectures, and it can combine visually appealing components with a data-driven approach.

### Processing

Processing is the marquee tool for data visualization. All you need to do is write some simple code and compile it into Java. There is also a processing.js project that will make it easier for websites to use Processing without Java Applets. Since the port supports Objective-C, you can also use Processing on iOS. Although Processing is a desktop application, it can run on almost any platform, and the Processing community has grown over the years to include a large number of instances and code.

### Rapidminer

Rapidminer, a handy data visualization tool, another essential tool for data processing, is an open-source data science platform that works through a visual programming mechanism. Its capabilities include modifying, analyzing, and creating models and quickly integrating the results into business processes. Rapidminer has received a lot of attention and has become a reliable tool in the minds of leading data scientists.

### Google Chart API

The Google Chart API tool set has removed the static image feature and now only provides the dynamic Chart tool. It works in all browsers that support SVGCanvas and VML, but one big problem with Google Chart is that: Diagrams are generated on the client side, which means devices that do not support JavaScript will not be able to use them, nor will they be able to use them offline or save the results to other formats, as was previously the case with static images. Despite the above issues, it's undeniable that the Google Chart API is incredibly rich, and if you don't have a particular need for customization or resistance to Google's visual style, you can start with Google Chart.

## V. CONCLUSIONS

This paper first introduces the data visualization that can present the current types. Visualization technology has been widely used in finance, medical care, geography, e-commerce, and so on, and has played a huge role. Through the visualization application of knowledge maps, users can understand certain knowledge in the shortest time and obtain accurate information. The visual display of data represents abstract "data" in a visible form, which helps people understand the data. Data visualization essentially deals with different data objects compared to traditional data visualization. Data visualization requires large-scale, multi-type, fast-updating, and efficient data processing, which brings a series of new challenges to the research and application of data visualization.

## ACKNOWLEDGMENT

This work was supported in part by the Natural Science Foundation of the Higher Education Institutions of Anhui Province under Grant No. KJ2020A0011, Innovation Support Program for Returned Overseas Students in Anhui Province under Grant No. 2021LCX032. the Science Research Project of Anhui University of Finance and Economics under Grant No. ACKYC20085, Undergraduate teaching quality and teaching reform project of Anhui University of Finance and Economics under Grant No. acszjyyb2021035.

## REFERENCES

1) Enrico G. Caldarola, Antonio M. Rinaldi (2015). Big Data Visualization Tools: A Survey - The New Paradigms, Methodologies and Tools for Large Data Sets Visualization. Proceedings of 4th International Conference on Data Management Technologies and Applications, 2015.
2) Satish Premshankar Yadav, Adarsh S.K Singh (2020).Big Data Analytics with Pandas and Scipy Python Tools, International Research Journal of Engineering and Technology (IRJET), 2020, 1800-1805.
3) Stančin, Igor, and Alan Jović. "An overview and comparison of free Python libraries for data mining and big data analysis." 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2019.
4) Cielen, D. and Meysman, A., 2016. Introducing data science: big data, machine learning, and more, using Python tools. Simon and Schuster.
5) Grover, P. and Kar, A.K., 2017. Big data analytics: A review on theoretical contributions and tools used in literature. Global Journal of Flexible Systems Management, 18(3), pp.203-229.