



Visualization Techniques for Analyzing Learning Effects – Taking Python as an Example

Keshuang Zhou, Yuyang Li, and Xue Han[✉]

Nanjing Normal University of Special Education, Nanjing 210000, Jiangsu, China
snow@njts.edu.cn

Abstract. With the advent of the information age, data visualization technology has gradually shown its unique features in various information fields, and its importance has gradually been attached importance by various governments and commercial departments. In teaching in our country, most schools simply use office software to create pie chart, histogram or table to realize visualization. This kind of teaching creates a single chart that doesn't change at all. The content of the data visualization course of many school courses is outdated, and the way of visualization is not in line with the needs of The Times, that is, the content has been criticized as abstract, language mechanization, format and public culture [1]. In order to analyze the teaching quality and evaluate and improve it, this paper processes and analyzes the data exported from the teaching administration system based on python, mainly from three aspects: data acquisition, data processing and data analysis. Firstly, the python crawler technology is used to obtain students' grades, secondly, the invalid data is processed, and finally, the matplotlib library is used to visualize the processed data, and the learning status of students in the class is analyzed and evaluated by combining the obtained images. Through the data processing of this paper, it realizes the hiding of the student's name, protects the privacy of the student, and uses the graph to intuitively reflect the student's grade distribution, which makes the grade analysis more convenient.

Keywords: Artificial Intelligence · Python · Visualization · Word Cloud

1 Introduction

Since ancient times, China has always attached great importance to education. With the rapid development of artificial intelligence in today's society, the idea that science and technology are the primary productive forces is fully implemented, based on education. In the daily learning process of students, students' performance in class, study time and grades of online courses, as well as the score statistics of students' final exams all need to deal with a large amount of data. Examination is an important link in the teaching process, and also an important means to reflect the teaching effect. The analysis of examination results can not only reflect students' learning attitude, learning effect and effort level, but also reflect teachers' learning attitude, teaching method, teaching effect and teaching management level to a certain extent. Achievement analysis is an important

part of teaching process [2]. At present, the most commonly used tools are the PivotTable function of Excel or mode, mean value and variance in SPSS to generate visual charts so as to visually reflect the strengths and weaknesses of different subjects and daily learning performance of different students, providing reference for students to adjust their learning status in the future. It also provides a guarantee for teachers and parents to have a clearer understanding of students' learning status. However, once faced with a large number of data, Excel and SPSS will inevitably make people a little confused, still need a more convenient way to carry out data visualization analysis. In the face of massive data, how to analyze and mine the connotation of data and better present useful information from the data has become the main object of current research. Data visualization is an important means of data analysis and information presentation, and has also become the main tool of data research [3]. Python is a commonly used programming language in big data. This paper uses Matplotlib [4] and Numpy [5] libraries in Python [6] and takes the final grades of this class as an example for data analysis. In order to protect personal privacy, the names of students have been processed when taking data in this paper.

2 Current Situation of Analysis for Learning Effect

In the current teaching environment, how to let students have more time to study independently through teaching reform has become a more concerned issue. In the face of the increasing amount of data, the traditional paper analysis method is obviously unable to meet the demand. So how do you turn data into useful information? Through the use of visualization technology, we can use visual charts, graphs and other forms to visually display the performance information. Visual chart analysis of grades is an important reflection of students' learning results in the learning process, reflecting students' mastery of knowledge and learning ability in a period of time. In many cases, students' grades are not particularly ideal, so for them, performance analysis is particularly important. In this way, students can be helped to self-reflect, so as to find their own problems in learning and make targeted adjustments. In daily learning, students can learn the differences between their weak subjects and strong subjects through various channels. But much of this information is indirectly obtained by teachers, which obviously can't meet the needs of students. So visualization allows students to have a more intuitive understanding of subjects they are weak in or interested in, It not only provides reference for students' subsequent development of emphasis, but also provides feedback for teachers and schools' teaching strategies and methods, playing an extremely important role. Traditional analysis of grades wastes time and results are not clear enough. For example, when the author was in high school, he relied on hand-drawn line charts of grades of each subject, which could only reflect the progress trend of the grades of a single subject, and also wasted students' time. Moreover, some teachers will use the data perspective function of Excel and SPSS in statistics to integrate and analyze data and generate visual charts. The efficiency of these two methods in the face of a large number of data is also evident. The use of technical means, such as Python, can be used to visualize the data, so that students can see their own shortcomings intuitively. Through review and consolidation, students can better grasp the content, so that teachers and schools can quickly

grasp students' understanding of knowledge, so as to timely adjust the teaching plan and make relevant teaching plans, so that parents can understand the learning effect of students more clearly. Thus more comprehensive grasp of the students' basic situation.

3 Overview of Tools

3.1 Python

Python is a widely used object-oriented, interpreted, cross-platform computer programming language. It is an open source, free, and powerful language that provides a rich and powerful standard library, as well as a large number of third-party libraries. Functions of third-party libraries involve artificial intelligence, data analysis and processing, machine learning, Web application development and other fields. With the development of these fields in recent years, Python has become more and more popular. It has the advantages of clear and simple, high development efficiency, portability, expansibility, embeddability and so on. The open source and free features also make this language get the support and promotion of many users, and become a new choice for the development of data processing system. 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 [7].

3.2 Matplotlib

Matplotlib is an open source data visualization toolkit that is the most popular drawing library in Python for developers, and it is comprehensive and versatile. Matplotlib is Python's 2D drawing library. It generates public-quality graphics in a variety of hardcopy formats and cross-platform interactive environments, mainly for generating drawings and other visualizations of two-dimensional data. When using this module, it takes a few lines of code to generate the required curve chart, histogram, scatter chart, line chart, pie chart, bar chart, box chart, spectrum chart and radar chart. Not only can you draw 2D graphics, you can draw 3D graphics, you can also draw animations.

3.3 Numpy

As the most important basic package of Python numerical computation, Numpy is an open source numerical computation extension of Python and a basic module of data analysis and Kaohsiung scientific computation. It can be used to store and process large matrices. It not only supports a large number of dimensional arrays and matrix operations, but also provides many advanced numerical programming tools, such as matrix data type, vector processing and precise operation library. Designed for strict digital processing.

4 Data Visualization Analysis

4.1 Concept

Visualization is a theory, method and technology that uses computer graphics and image processing technology to convert data into graphics or images and display them on the screen, and carry out interactive processing. With the development of digital multimedia technology, visualization technology has been widely used in engineering, medicine and other fields. This technology converts relevant information such as data into graphic images for display, which helps users understand the interrelationship and development trend between a large number of abstract data, and improves users' ability to observe things [8]. Visualization tools are widely used in the visualization of graphic images [9], human-computer interaction [10], scientific calculation visualization [11] and other fields. It involves many fields such as computer graphics, image processing and computer vision. For example, some line charts, pie charts and bar charts in Excel can visually display data in the form of charts to help people better understand the meaning of data and make decisions about the next step. This is visualization.

The chart below, for example, is the simplest data analysis (Fig. 1).

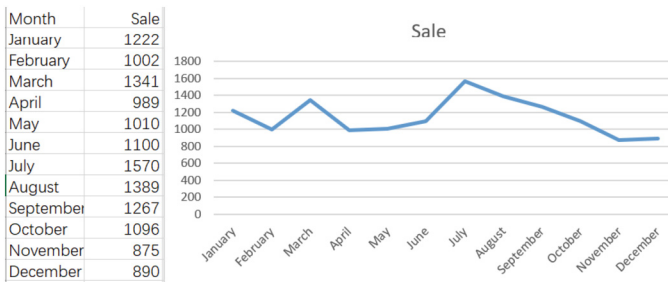


Fig. 1. Simple line graph

Through data visualization technology, text is converted into graphs. The line chart can intuitively see the sales trend of one year. The chart shows that the sales volume changes first, then rises, then rises to the highest value, and finally declines steadily.

4.2 Principle

At present, we are in an era of data and information explosion. No matter when and where we are, we will inevitably face the situation of actively or passively receiving news and feedback. The human eye has powerful pattern recognition capabilities, and more than half of the human brain function can be used to process and feed back visual information [12]. Compared with boring words and numbers, the human brain can be more intuitive and more specific to recognize elements such as graphics, colors, and sizes, and can discover the information contained in the data for the first time from the data visualization graphics [13]. Data visualization refers to the presentation of a large amount of relevant

data in the form of images and charts, such as word cloud map, radar map, percentage pie chart, etc., which can integrate data and visually display graphic images [14]. These charts can communicate and communicate data clearly and effectively, and can analyze data from multiple dimensions to draw deep relevant conclusions. With graphical means, key contents and features can be conveyed intuitively [15]. Word cloud map consists of frequently appearing words, which are cloud-like color graphics, used to display a large amount of text data.

For example, in the topic discussion in the general course of learning, word clouds will be automatically generated when students post their discussions. Words with multiple repetitions are in the center with dark colors and large size [16]. This intuitive visual effect enables teachers to understand the learning effect of students by observing the word cloud. As shown in the picture below (Fig. 2):

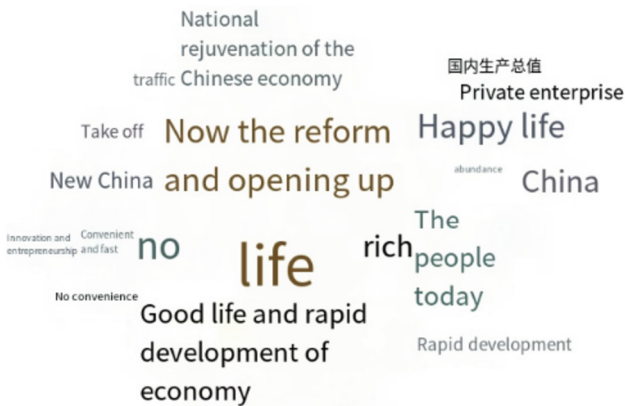


Fig. 2. Topic discussion word cloud map

The implementation code is:

```
def create_ciyun():
    input:text
    open text
    read() text
    list (text)
    Obtain the output final_text
```

Get the data pseudocode

4.3 Steps of Data Visualization Analysis

Any data that can be expressed in charts and images can be analyzed visually. First of all, a large amount of relevant data needs to be acquired. There are many methods of data acquisition, such as questionnaire survey and interview, and crawler and other methods are generally adopted to acquire some open data [16]. Secondly, data screening and processing are carried out. The data we acquire may be messy, complex and repetitive [17]. For such low-quality data, data should be discarded to ensure the effectiveness and reliability of the data. The last is data visualization. This step is to express the processed data information in the form of graphics and images. At present, people generally use Excel, Python, Matlab and other software technologies to realize data visualization analysis [18].

5 Data Visualization Analysis by Python

First of all, climb from the teaching administration system of the final subject scores of the class, calculate the personal score, and then integrate into the class score, and finally the class score for visual analysis.

5.1 Obtain Data

Using the third party library requests this web crawler tool, crawl the student achievement information on the educational administration system, just need to give the website and crawl rules, you can get a lot of information on the website. Based on the analysis of the scores of some students in this class, this paper obtains the score data of the basic course of the final examination in the second semester of the 2022–2023 academic year, and makes a statistical analysis of the information of the students' course scores. The key indicators mainly include the number of students taking the exam, the passing rate of the paper scores, the passing rate of the overall score, the passing rate of the paper scores and the passing rate of the overall score [19]. The code is as follows:

The result information is beautified by BeautifulSoup and saved in a local file, or exported to an Excel workbook for processing [20].

```
Input: the address of the educational administration system
Input: the score address
Request head
Input: the student username and password
def get(url):
x=requests.get(url,headers=head)
return x
if __name__ == '__main__':
    cookie=get (url)
    x=requests (url,headers=head,data=user)
    content=x.text
    gap=requests (url_gap, headers=head)
Obtain the output:gap.text
```

5.2 Data Processing

Check the obtained data, and process invalid duplicate data, modify, replace, or delete it directly [21]. Such programs would also help to reduce error rates by alleviating the burden of manual data processing that hampers the processing of large-scale actigraphy data sets[i].

For example, the data redundancy processing code is as follows.

```
Input:test={'id':[1,2,3,4,5,6,6],'name':['Lisa','Bob','Candy','Mike','John','Gary','Gary'],
           'chinese':[91,90,100,79,98,94,94],'english':[90,95,81,95,95,91,91]}
#View redundant rows
Obtain the output:Null
#Delete redundant rows
Obtain the output:Ture
```

Data processing pseudocode

5.3 Statistical Processing

The results of each subject are counted to obtain the total results of the individual and the class. The weighted average function can be used. The average score, the variance, the highest score and the number of students who failed in each subject of the class were calculated according to the student achievement [22].

5.4 Visualization

Matplotlib library is a third-party library in python language [23]. The most basic charts in daily work include bar charts, bar charts, line charts, pie charts, etc. Matplotlib module provides corresponding drawing functions for these charts [24]. The data used to draw the chart can be directly used in the code provided. The bar chart is taken as an example to draw the performance analysis chart below [25]. The code is as follows:

```
Input:x=[Lisa,Liming,Wanghua,Amy,Tom,Anna,Michelle,Mike]
Input:y = [60, 80, 97, 65, 82, 67, 80, 90]
plt.bar(x, y)
plt.show()
```

Do visual pseudocode

Lines 2 and 3 give the values for the x and y axes of the chart, line 4 uses the bar() function to draw the bar graph, and line 5 uses the show() function to show the graph (Fig. 3).

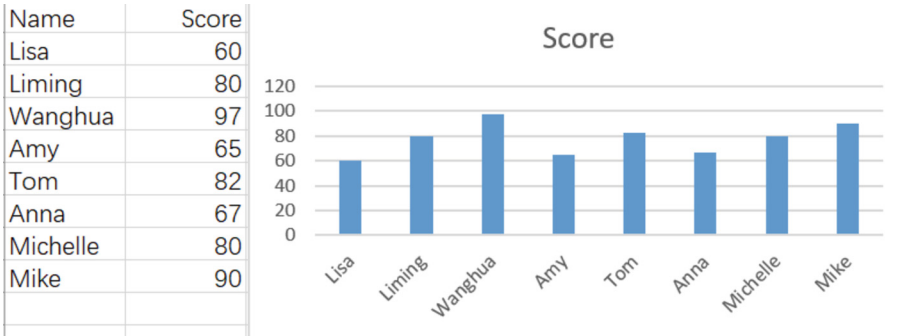


Fig. 3. Bar chart of student achievement

After running, we can intuitively see the distribution of students' grades. In actual operation, we can draw a variety of graphs about grades by modifying some parameters. Teachers can also set up charts of different situations according to the specific needs of analysis, such as individual subject or total score change table (line chart), the percentage of grade in the total number of students in the class (pie chart) and so on [26] (Figs. 4 and 5).

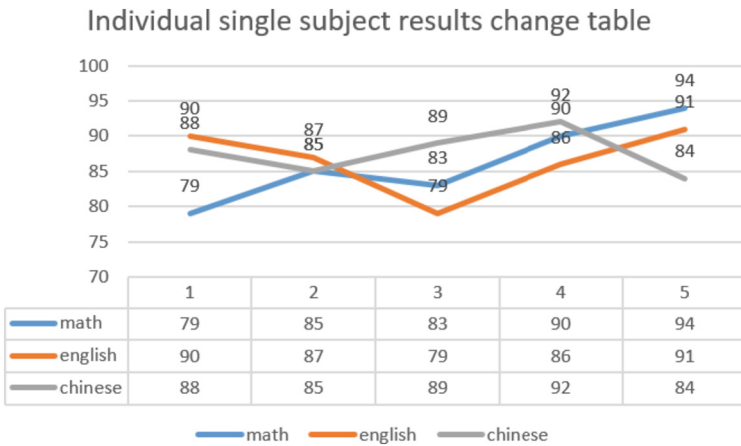


Fig. 4. Line chart of student achievement

It can be seen that python visualization technology has the advantages of processing a large amount of data and high reuse rate [27]. When the original data is changed, information can be automatically updated, which is convenient for the use and observation of data statisticians. Visualization technology based on big data mining existing mass data, extract valuable information, applied to class performance analysis and university teaching management, has become a hot practice, can play a positive role in improving the teaching effect [28].

Class grade ratio chart

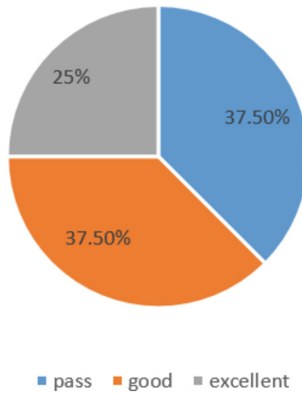


Fig. 5. Pie chart of student achievement

6 Conclusion

With the advent of the era of big data, the development of cloud computing and the continuous progress of big data storage technology, data analysis and visualization analysis are becoming more and more important [29]. This paper emphasizes its application in the field of education. The visual analysis of student achievement data can be generated flexibly by using Python according to the results of part of the final examination of the author's class. Some data can be visualized as a whole, can also be used to show some time series data, and even can be used to show some distribution rules, it is also very convenient to use, can be directly through a code to call another code to achieve, for students, schools, parents in the next deployment of various aspects to provide convenient conditions. In order to make targeted adjustments to the next way of learning and teaching. It can be seen that achievement analysis will continue to be an indispensable part of education in the future [30]. At the same time, we also hope that relevant researchers and scholars can continue to in-depth study on the aspect of performance analysis, and apply it to more aspects in the field of education. Education information has become the mainstream, many schools have their own campus system, performance analysis system, so as to use information management to make school management more scientific and effective. As a result, data analysis and visualization are becoming mainstream. In the future, data visualization will become more popular, which can not only reduce labor costs, but also improve the efficiency of analysis, making analysis more simple and convenient for students. I believe that visual analysis technology will become more and more popular in the near future. However, Python also has some shortcomings, such as Python syntax is relatively simple, resulting in python is not easy to cooperate with other languages [31]. Moreover, python is also based on object-oriented language, so it is difficult to support some complex models, these problems need to be perfected in the future, so Python is a technology that can continue to expand the application.

Funding. This paper is supported by the Fund for Philosophy and Social Sciences of Universities in Jiangsu Province, China, “Research on Integrated Education Based on the Decentralization of Blockchain Technology” (2019SJA0543).

References

1. Fabry, D.L., Higgs, J.R.: Barriers to the effective use of technology in education: current status. *J. Educ. Comput. Res.* **17**(4), 385–395 (1997)
2. Seidel, T., Shavelson, R.J.: Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Rev. Educ. Res.* **77**(4), 454–499 (2007)
3. Few, S.: Eenie, Meenie, Minie, Moe: selecting the right graph you’re your message (2004)
4. Bisong, E.: Matplotlib and seaborn. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, pp. 151–165 (2019)
5. Bressert, E.: *SciPy and NumPy: an overview for developers* (2012)
6. Kelly, S.: *What Is Python? Python, PyGame and Raspberry Pi Game Development*, pp. 3–5 (2016)
7. Diehl, S.: *Software Visualization: Visualizing the Structure, Behaviour, and Evolution of Software*, pp. 11–18. Springer, Heidelberg (2007). <https://doi.org/10.1007/978-3-540-46505-8>
8. Telea, A.C.: *Data Visualization: Principles and Practice*, pp. 91–102. CRC Press, Boca Raton (2007)
9. Shahin, M., Liang, P., Babar, M.A.: A systematic review of software architecture visualization techniques. *J. Syst. Softw.* **94**(5), 161–185 (2014). <https://doi.org/10.1016/j.jss.2014.03.071>
10. Drevelle, V., Nicola, J.: VIBes: a visualizer for intervals and boxes. *Math. Comput. Sci.* **8**(3–4), 563–572 (2014). <https://doi.org/10.1007/s11786-014-0202-0>
11. Allen, F., Gale, D.: Limited market participation and volatility of asset prices. *J. Am. Econ. Rev.* **984** (1994)
12. Cao, S., Zeng, Y., Yang, S., et al.: Research on Python data visualization technology. *J. Phys. Conf. Ser.* **1757**(1), 012122 (2021). IOP Publishing
13. Hammad, G., Reyt, M., Belyi, N., et al.: PyActigraphy: open-source python package for actigraphy data visualization and analysis. *PLoS Comput. Biol.* **17**(10), e1009514 (2021)
14. Dennis, D.R., Meredith, J.R.: An analysis of process industry production and inventory management systems. *J. Oper. Manag.* (2000)
15. Sambasivam, S., Theodosopoulos, N.: Advanced data clustering methods of mining web documents. *Issues Inf. Sci. Inf. Technol.* (2006)
16. Freitas, C.M.D.S., et al.: On evaluating information visualization techniques. In: *Proceedings of the Working Conference on Advanced Visual Interfaces* (2002)
17. Wehrend, S., Lewis, C.: A problem-oriented classification of visualization techniques. In: *Proceedings of the First IEEE Conference on Visualization: Visualization90*. IEEE (1990)
18. Chi, Ed.H.: A taxonomy of visualization techniques using the data state reference model. In: *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings. IEEE* (2000)
19. Klerkx, J., Verbert, K., Duval, E.: Enhancing learning with visualization techniques. In: Spector, J., Merrill, M., Elen, J., Bishop, M. (eds.) *Handbook of Research on Educational Communications and Technology*, pp. 791–807. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-3185-5_64

20. Cao, N., Cui, W.: Overview of text visualization techniques. In: Cao, N., Cui, W. (eds.) *Introduction to Text Visualization*. Atlantis Briefs in Artificial Intelligence, vol. 1, pp. 11–40. Atlantis Press, Paris (2016). https://doi.org/10.2991/978-94-6239-186-4_2
21. Keim, D.A., Kriegel, H.-P.: Visualization techniques for mining large databases: a comparison. *IEEE Trans. Knowl. Data Eng.* **8**(6), 923–938 (1996)
22. Kamat, V.R., et al.: Research in visualization techniques for field construction. *J. Constr. Eng. Manag.* **137**(10), 853–862 (2011)
23. Al-Kodmany, K.: Using visualization techniques for enhancing public participation in planning and design: process, implementation, and evaluation. *Landsc. Urban Plann.* **45**(1), 37–45 (1999)
24. Kucher, K., Kerren, A.: Text visualization techniques: taxonomy, visual survey, and community insights. In: 2015 IEEE Pacific visualization symposium (pacificVis). IEEE (2015)
25. White, S., Feiner, S.: SiteLens: situated visualization techniques for urban site visits. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009)
26. Tatu, A., et al.: Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In: 2009 IEEE Symposium on Visual Analytics Science and Technology. IEEE (2009)
27. Zammito, V.: Visualization techniques in video games. *Electron. Vis. Arts (EVA 2008)*, 267–276 (2008)
28. Vallat, R.: Pingouin: statistics in Python. *J. Open Sour. Softw.* **3**(31), 1026 (2018)
29. Sousa da Silva, A.W., Vranken, W.F.: ACPYPE-Antechamber python parser interface. *BMC Res. Notes* **5**, 1–8 (2012)
30. Millman, K.J., Aivazis, M.: Python for scientists and engineers. *Comput. Sci. Eng.* **13**(2), 9–12 (2011)
31. Ari, N., Ustazhanov, M.: Matplotlib in Python. In: 2014 11th International Conference on Electronics, Computer and Computation (ICECCO). IEEE (2014)