

PAPER • OPEN ACCESS

Research on Python Data Visualization Technology

To cite this article: Shengjia Cao *et al* 2021 *J. Phys.: Conf. Ser.* **1757** 012122

View the [article online](#) for updates and enhancements.

You may also like

- [Application Analysis of Artificial Intelligence in Library Network Security](#)
Sheng Li
- [Graph Structure-based Clustering Algorithm for Android Third-party Libraries](#)
Shuyuan Liu and Gang Gan
- [Library Development Strategy for The Community at Coastal Areas](#)
Putut Suharso, Bani Sudardi, Sahid Teguh Widodo et al.

Research on Python Data Visualization Technology

Shengjia Cao, YunhanZeng, Shangru Yang, Songlin Cao*

Hunan University of Finance and Economics, Changsha 410205, China

*1345450833@qq.com

Abstract. In recent years, researchers at home and abroad have accumulated a lot of experience in the research of data visualization technology, and they have played an important role in scientific discovery, medical diagnosis, business decision-making, and engineering applications. As a library developed using Python language, Matplotlib has a concise language, high drawing accuracy, and simple and easy-to-understand code. This article first introduces data visualization and related technologies used and then uses Python's Matplotlib library and pyecharts library to realize data visualization. Through representative examples, combined with the use of correct charts, visual processing of data in different fields, so as to further analyze the effect of visualization.

Keywords: Python, Data visualization, matplotlib, pyecharts

1. Introduction

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 [1]. 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.

The purpose of data visualization is to communicate data and image information more clearly and efficiently with the help of suitable graphical representations, such as histograms, bar charts, column charts, and other statistical charts. Python is a good choice, providing a wealth of third-party libraries, open-source communities, and continuously optimized documentation for data visualization. In this article, Python's Matplotlib library and pyecharts library are used to realize data visualization.

2. Related works

In recent years, with the continuous improvement and development of computer technology, more and more visualization tools have appeared. Many experts and scholars in different fields have used relevant visualization tools to process and analyze the data obtained from their hands. For example, Qing Tao and others collected the rainstorm data of the historical rainy season for statistical prediction and analysis, in order to make preparations for flood prevention and disaster reduction in advance [2]; Zheng Yanping developed a visualization platform for numerical forecast product analysis to provide more practical graphic product services for meteorological research [3]; Liu Peining and



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

others did the statistical analysis and visualization processing on Network Common Data Form data, and demonstrated the application of data processing and visualization in the field of meteorology[4]; He Quanjun and other platforms have developed a visualization system to realize the three-dimensional display of meteorological data [5].

As open-source and free software, Python has developed more than 100,000 toolkits. Matplotlib is one of the most commonly used data visualization tools in Python. Using Matplotlib can easily draw high-quality different types of 2D charts and some relatively basic 3D charts.

IBM has developed an online visual processing tool-ManyEyes[6], which supports users to upload data and create interactive visual websites. Tableau has developed a business intelligence software-Tableau, which supports users to perform interactive operations on visual graphics, and can perform intuitive analysis and processing of data; the software has a built-in memory data engine, which can realize data accelerated processing[7].

3. Application of movie data visualization

The movie data used in this chapter comes from <http://dataju.cn/Dataju/web/datasetInstanceDetail/226>, the file data includes 28 attributes of 5043 movies grabbed from the Internet Movie Database website Information, the film time span is 1916-2017 and 66 countries and includes 2399 directors and thousands of actors. Attributes include movie name, number of comments, rating, director, release time, release country, etc.

(1) Import data package

```
import numpy as np
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import pylab as mpl
```

```
mpl.rcParams['font.sans-serif'] = ['SimHei']
```

(2) Analysis of movie types

Looking at the development trend and situation of movies as a whole:

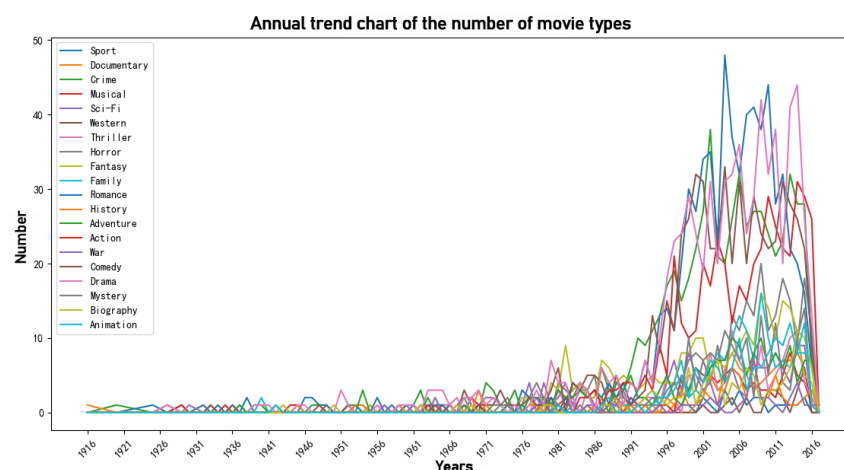


Fig. 1 Annual trend chart of the number of movie types

The line chart can reflect the trend of things, so that people can find the law of change, and can more accurately see and predict the future development trend. A single line chart can understand the development law of a certain thing over time.

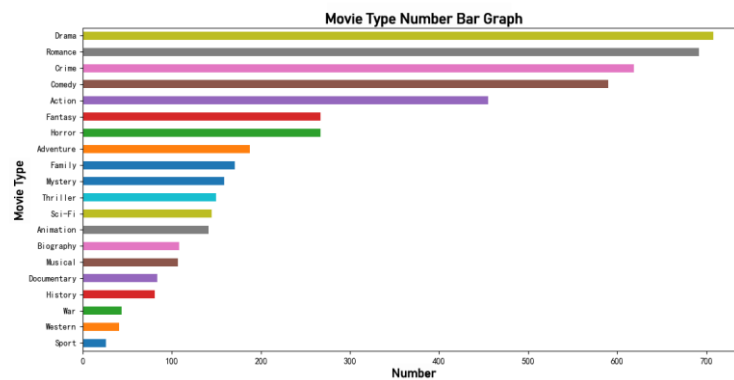


Fig. 2 Movie type number bar graph

The movie genres of Drama, Romance, Crime, Comedy, and Action rank among the top five. Documentary, History, War, Western, and Sports films are the bottom five.

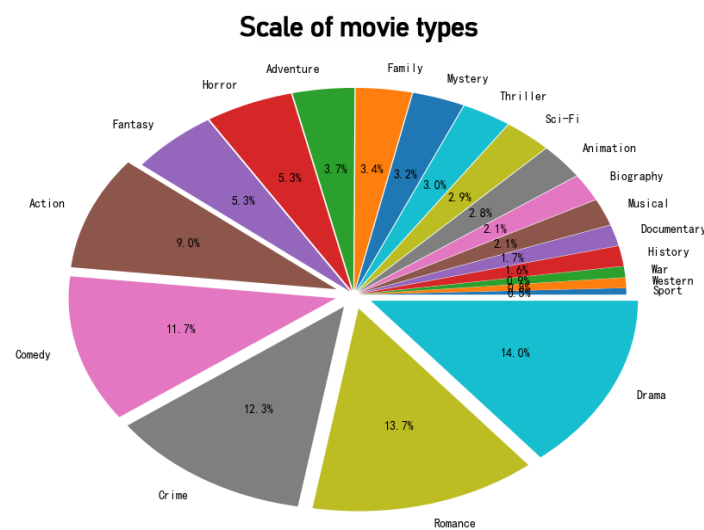


Fig. 3 Movie type of scale chart

From the pie chart, you can not only see the proportion of each type of movie but also clearly get the percentage of each type in the overall.

Both bar graphs and pie graphs can show a certain type of quantitative relationship between totals. When there are fewer variables, the bar chart is more helpful for us to understand the sorting situation of the quantity, but when there are more variables, there will be multi-line overlap, the graph is a bit messy, which is not conducive to observation, the pie chart shows more proportion, which can more accurately let people understand the total proportion of each type.

(3) Movie ticket situation

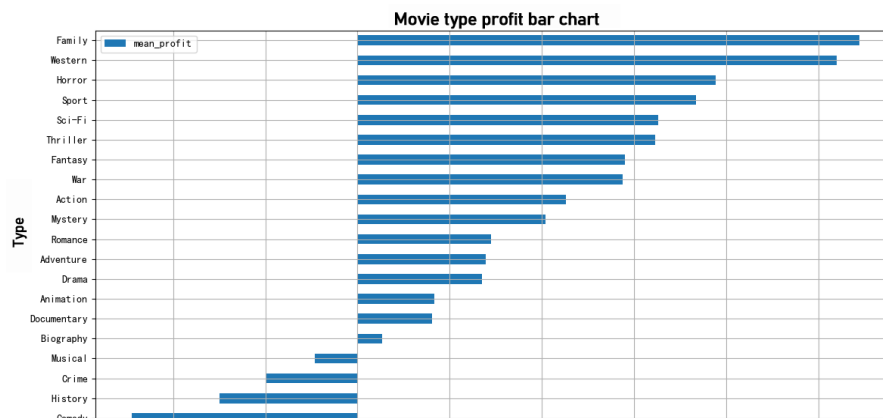


Fig. 4 Movie type profit bar chart

The graph shows the profit of movie types. Among them, Family, Western, and Horror movies have the highest cost, and Musical, Crime, History, and Comedy movies have a negative cost. This type of bar graph intuitively reflects the positive and negative conditions of the value, not only can judge the data sorting situation, but also express the meaning of the data.

4. Application of art quality data to achieve visualization

The main data used in this example are five files, namely Tianjin, Beijing, Shanghai, Guangzhou, and Shenzhen for the full-year air quality index of the five cities in 2018, including attributes such as the AQI index and AQI index ranking of the day. The air quality data of each city comes from www.tianqihoubao.com. The 5 files were crawled down by the Python crawler, with a total of 1436 valid data, and the data was preprocessed.

The visualization process is divided into two parts. The first part separately visualizes the air quality detection situation in Tianjin in 2017, and the latter part comprehensively processes the four cities of Beijing, Shanghai, Guangzhou, and Shenzhen in 2018. The processing realization process is as follows:

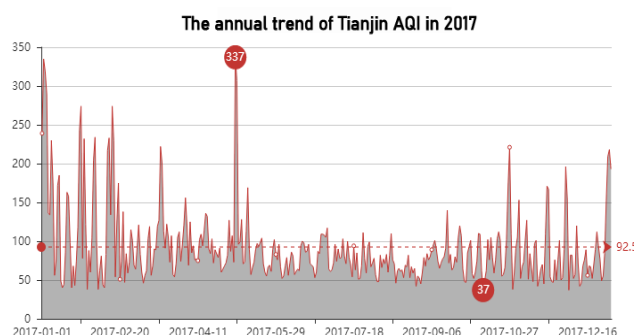


Fig.5 2017 Tianjin AQI annual trend chart

In 2017, the average AQI of Tianjin for the whole year was 92.5, which is close to 100, which is good. This value is very close to light pollution, indicating that Tianjin's air quality in 2017 was poor.

The highest AQI air quality index is now in May, with a value of 337, which is serious pollution. The minimum value of AQI appears in October, with a value of 37, which is excellent. As a whole, Tianjin's AQI index in 2017 was relatively high during the change of seasons between winter and spring and autumn and winter, and the basic AQI index was better in summer and autumn.

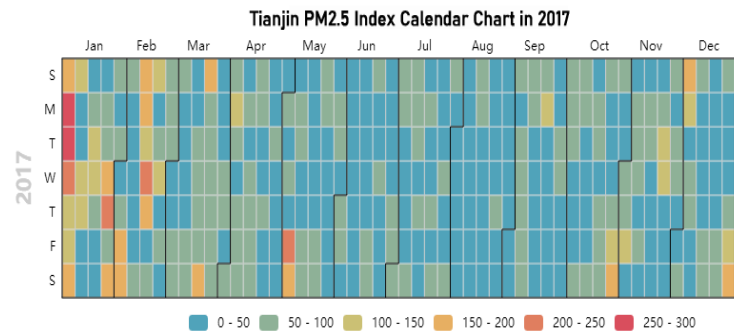


Fig. 6 2017 Tianjin PM2.5 Index Calendar Chart

Judging from the heat map, the PM2.5 index in January was generally high, with half of the days exceeding 100. It was followed by February, accounting for about one-third. The PM2.5 index from June to August was basically stable within 100, of which the index is mostly 0-50, indicating that the overall situation is better. The situation in the remaining months is better, with the index exceeding 100 in just one or two days.

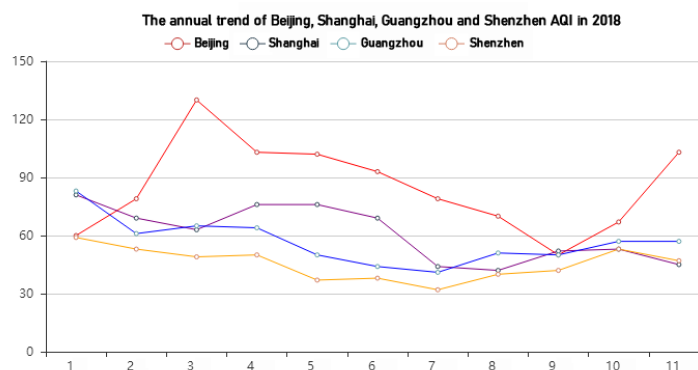


Fig.7 2018 Beijing, Shanghai, Guangzhou and Shenzhen AQI annual trend chart

From the annual AQI trend of the four cities, it can be seen that among the four cities, Beijing has the highest overall AQI index, Shanghai is second, Guangzhou is third, and Shenzhen is the lowest overall.

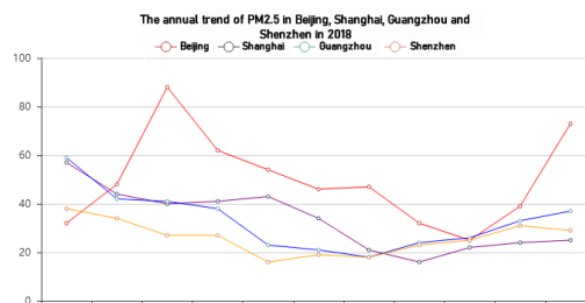


Fig.8 2018 Beijing, Shanghai, Guangzhou and Shenzhen PM2.5 annual trend chart

The change trend of the PM2.5 chart of Beijing, Shanghai, Guangzhou and Shenzhen is basically the same as that of the AQI's annual chart, which once again shows that the PM2.5 index is positively correlated with AQI. Beijing's high PM2.5 index situation has led to an increase in its AQI index.

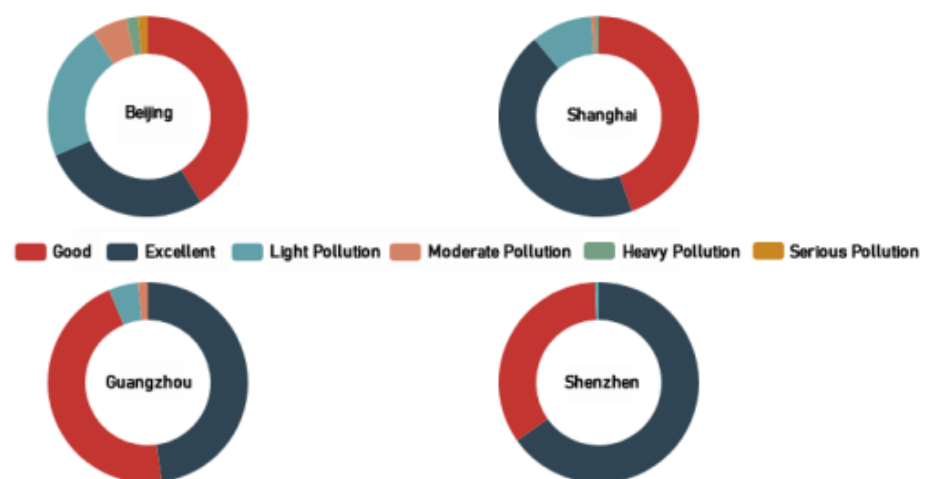


Fig.9 2018 Beijing, Shanghai, Guangzhou and Shenzhen's annual air quality ratio map

From the analysis of the number of polluted days in the total number of days, Beijing has the highest proportion of polluted days in the whole year, followed by Shanghai, and Guangzhou ranked third. The situation in Shenzhen is the best, with less than 10% of the total number of polluted days per year.

Visualization uses Python's pyecharts, which is richer than the charts generated by matplotlib, and the charts generated by pyecharts have a dynamic form. Mouse over the key part of the chart to display the information represented there, which helps readers to interpret the chart information.

5. Summary

In data analysis, data visualization also plays an indispensable role. The emphasis on visualization helps people interpret the data and mine the hidden meaning behind the data faster and more accurately. This article studies the current development of visualization, and demonstrates the idea of visualization for different data through two cases. Select more representative data, use Python's third-party drawing libraries matplotlib and pyecharts to process different data, and select suitable charts for visual analysis.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China, grant number 72073041. Open Foundation for the University Innovation Platform in the Hunan Province, grant number 18K103; 2011 Collaborative Innovation Center for Development and Utilization of Finance and Economics Big Data Property, Universities of Hunan Province. Open project, grant number 20181901CRP03, 20181901CRP04, 20181901CRP05. 2020 Hunan Provincial Higher Education Teaching Reform Research Project under Grant HNJG-2020-1130, HNJG-2020-1124. 2020 General Project of Hunan Social Science Fund under Grant 20B16.

References

- [1] Allen F and Gale D 1994 Limited market participation and volatility of asset prices *J. American Economic Review* **9** 84
- [2] Alexandros V Benos 1998 Aggressiveness and survival of overconfidence traders *J. Journal of Financial Markets* **10** 3
- [3] Benos Alexandros 2000 *Agricultural development in china cambridge* (Harvard University Press) pp1368-1968
- [4] Peining L, Xiao H and Fuxing Y 2014 NetCDF file analysis and visualization application based on R *J. language Meteorological Technology* **04** 629
- [5] Barber Brad M and Odean Terrance 1998 Boys will be boys: Gender, Overconfidence, and Common Stock Investment *J. The Quarterly journal of Economic* **2** pp261-229
- [6] Zhi S 2011 The financial perspective of transforming economic development mode: system and policy evolution and innovation (Dalian: Dongbei University of Finance and Economics Press) pp101-119
- [7] Zhong L 2012 Institutional innovation in the transformation of my country's economic development mode CPC (Beijing : Central Party School Press) pp69-75