Comparison and Analysis of Air Quality Visualization between Beijing and Nanjing Based on R

Yan Zhang1,2,*, Peng Zhu1,2, *

1School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China
2Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094, China

*Corresponding author e-mail: pzhou@njust.edu.cn, 834267795@qq.com

Abstract. With the continuous promotion of urban industrialization, urban air pollution problem is more and more serious. Therefore, carrying out real-time environmental testing and analysing the meteorological data have become particularly important for grasping the changes in the urban environment and solving the air pollution problem. This article uses R language to analyse the air quality data of Beijing and Nanjing in 2017 respectively. Compare and explore the reasons for the change in air quality and the differences in air quality between the two cities. The text is divided into four parts: Introduction - Research Status – The Data Acquisition and Analysis - Cause Analysis - Research Conclusions.

1. Introduction
As air quality data grows and accumulates, traditional data analysis and processing methods have been difficult to cope with complex and cumbersome data. Therefore, the effective use of the R language can visualize the complicated data, so as to clearly and clearly display the data characteristics [1]. This paper uses the R language to visualize the collected air quality data of Nanjing and Beijing, and then mines the hidden information and compares the differences between Nanjing and Beijing, and summarizes the reasons for the differences through some geographic knowledge [2]. This is of great significance for the control of urban air quality, and can also provide a reference for the decision-making of environmental management departments, thereby better maintaining the social production and living environment.

2. Research Status
With the continuous advancement of related work, environmental protection work has gradually changed from qualitative management to quantitative management which was more scientific relying on air quality data [3]. And how does such a huge amount of data play its role? This requires us to take more data analysis methods to present this data. Currently, domestic analysis of air quality data is based more on traditional tabular forms [4]. We know that the human brain processes visual information much more easily than written information, while the R language can draw more beautiful and complex charts. We use R language to visualize air quality data, which is a novel attempt in air quality analysis.
3. The Data Acquisition and Analysis

3.1. Data acquisition and ordering
The data source of this paper is from the real gas network, including date, AQI, air quality. The Octopus data acquisition system is based on a fully self-developed distributed cloud computing platform, which get a lot of normalized data from a variety of different websites or web pages in a short amount of time, which reduces the cost of getting information and improves efficiency. Specific steps are as follows:

Step 1: Create a crawler rule through the Octopus collection software and copy the web page into the rule process.
Step 2: Create a loop and then click through to the statistics for each month of the last three years.
Step 3: Loop to crawl the data of each day.

After observing the data for three years, it is found that the analysis with the data of only 2017 is better and the presentation is clearer. Therefore, after obtaining the three-year data, some of the excel forms are manually deleted, leaving only 2017. Beijing’s data is as shown below:

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| A | B       | C        | D | E | F | G | H | I |
| 1 | Date  | AQI     | Grade | PM25 | PM10 | SO2 | CO | NO2 | O3 |
| 2 | 2018/1/1 | 470 | Serious Pollution | 454 | 516 | 8 | 6.7 | 141 | 6 |
| 3 | 2018/1/2 | 248 | Severe Pollution | 198 | 322 | 13 | 3.8 | 102 | 38 |
| 4 | 2018/1/3 | 328 | Serious Pollution | 206 | 398 | 17 | 5.9 | 144 | 17 |
| 5 | 2018/1/4 | 389 | Serious Pollution | 321 | 852 | 18 | 15.5 | 150 | 13 |
| 6 | 2018/1/5 | 270 | Severe Pollution | 225 | 303 | 16 | 5.3 | 109 | 11 |
| 7 | 2018/1/6 | 248 | Severe Pollution | 198 | 199 | 25 | 4.3 | 108 | 14 |
| 8 | 2018/1/7 | 216 | Severe Pollution | 166 | 171 | 18 | 3.3 | 82 | 10 |
| 9 | 2018/1/8 | 54 | Good | 34 | 50 | 10 | 0.9 | 34 | 55 |
| 10 | 2018/1/9 | 62 | Good | 36 | 74 | 7 | 0.8 | 39 | 49 |
| 11 | 2018/1/10 | 45 | Excellent | 25 | 45 | 8 | 0.7 | 36 | 61 |
| 12 | 2018/1/11 | 94 | Good | 57 | 63 | 19 | 1.6 | 75 | 18 |
| 13 | 2018/1/12 | 62 | Good | 41 | 61 | 12 | 0.9 | 49 | 48 |
| 14 | 2018/1/13 | 34 | Excellent | 8 | 22 | 5 | 0.4 | 22 | 68 |
| 15 | 2018/1/14 | 57 | Good | 30 | 42 | 11 | 0.9 | 45 | 50 |
| 16 | 2018/1/15 | 73 | Good | 53 | 71 | 11 | 1 | 50 | 31 |
| 17 | 2018/1/16 | 23 | Mild Pollution | 93 | 106 | 18 | 1.7 | 66 | 8 |
| 18 | 2018/1/17 | 178 | Moderate Pollution | 134 | 152 | 21 | 2.1 | 73 | 19 |

**Figure 1.** Air quality data example

3.2. Data analysis and visualization

3.2.1. Air quality level word cloud map display

Using R language programming, select the highest frequency of 20 words, draw a word cloud map, you can more intuitively see the difference in air quality between Nanjing and Beijing [5]. It can be seen that Beijing has more polluted days. However, the word cloud map in Nanjing is not seriously polluted. This shows that the overall air quality in Beijing is worse than Nanjing.
3.2.2. **AQI frequency histogram.** AQI is an index that reflects air quality more directly and explicitly. Therefore, AQI is also essential for the analysis of air quality in two cities. This paper uses a histogram to show the frequency of the two cities' AQI index.

![AQI frequency histogram](image)

Comparing the data below 200, it is obvious that Nanjing is more concentrated, and data above 200 shows that Beijing is more than Nanjing. The higher the AQI index, the worse the air quality. Therefore, the overall air quality in Nanjing is better.

3.2.3. **Scatter plot of six pollutants.** We use the AQI index and a matrix scatter plot of six pollutants to study the most closely related contaminants that affect air quality:

![Scatter plot of six pollutants](image)

From the above matrix scatter plot, it can be seen that there is a positive correlation between the AQI index and the six pollutants, and PM2.5 is most closely related to the AQI index.
3.2.4. **AQI and PM2.5 calendar heat map.** Above we understand that the most intuitive performance of air quality is the AQI index. Therefore, comparing AQI in one year can clarify the relationship between air quality and seasonal factors such as temperature and humidity. The heat map is a beautiful and intuitive illustration, we show the above through the calendar heat map:

![AQI calendar heat map](image)

From the size of the index and the color depth, it can be seen that Beijing is close to the highest value; then observe the color change of each day to find that the color of Beijing in January and February is generally deep, while Nanjing is higher in December and January than in other months.

4. **Cause Analysis**
Regardless of the analysis, Beijing's air quality is worse than that of Nanjing. The reason for this phenomenon can be analyzed from the human characteristics of the city in addition to the geographical location. Beijing's industrial development is earlier than the south, and it is the capital of China, it is densely populated, industrially developed, and pollutants are discharged for a long time. At the same time, Beijing was seriously polluted in January and February. The cause of this phenomenon is firstly the energy structure problem, centralized heating and dry northwest winds allow the diffusion of pollutants to stay.

In May and June, Nanjing did not enter the rainy season. The temperature was high, rainfall was low, and relative humidity was low, the meteorological conditions were generally accelerated and it was conducive to the generation and accumulation of ozone. However, the weather conditions in Nanjing are much hotter and humid than in Beijing, and this feature is also conducive to the dilution of pollutants. Therefore, Nanjing's air quality is far superior to Beijing.

5. **Research Conclusions**
This paper climbs the data of Beijing and Nanjing air quality 2018, and visually analyzes the two data respectively, and draws the conclusion that Beijing air quality is better than Nanjing, and analyzes the months and quarters of poor air quality in Beijing and Nanjing. Therefore, put forward pollution control opinions
(1). Adjust the energy structure. Improve energy efficiency, vigorously develop green energy, and prohibit straw burning.
(2). Adjust the industrial structure. Increase investment in technology research, strengthen industrial upgrading, and improve factory efficiency.
(3). Manage traffic congestion. Promote public transportation.
(4). Improve the management of the three wastes. Rationally lay out factories, strictly supervise waste treatment and discharge, reduce the use of pesticides, and reduce individual heating.
(5). Improve laws and regulations and deal with pollution problems according to law.

Acknowledgments
This research was supported by National Natural Science Foundation of China (71874082, 71403121), Ministry of Education Humanities and Social Sciences Project of China (18YJA870021), Social Science Fund of Jiangsu Province, China (16TQC005), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX18_0367).

References
[1] Zhuang Wu, Aiming Qin. Visualization of air quality data based on big data, A. Wide-angle lens”Wide-angle lens”. 2015, (03): 249 - 253.
[2] Yajun Hou. The application of R language in data mining, J. Applied technology research. 2014, (02): 63 - 65.
[3] Zhenghai Zhang. Research on Visualization and Prediction Methods of Urban Air Quality Data, D. Shandong University. 2012.
[4] BaoLei. Exploratory air quality data analysis based on big data, J. ELECTRONICS WORLD·Exploration and Observation. 2017, (16): 5 - 6.
[5] Lina Han. Data visualization technology and application prospects, J. Coal mine modernization. 2005, (06): 39 - 40.