Atmospheric pollution analysis based on space-time tendency

Group 13
He Qirui Steven 1830026032
Wei Huanzhe William 1830026116
Zhang Chuliang Solomon 1830026152
Zhang Shichen Harry 1830026154
Data Science
BNU-HKBU United International College

***Abstract*-Nowadays, atmospheric pollution is one of the most severe environmental problems we are facing with. At present, thanks to the development of the air quality monitoring network, our country had achieved remarkable results on preventing and controlling the atmospheric pollution. The air quality monitoring station provides air quality data with high dimensional, temporal characteristics thus in this project, based on these meaningful data set, we use visualization methods to create visualizations to help analyze the space-time tendency of the atmospheric pollution. Our project totally including 4 visualizations, including one scatter plot, two line charts and one choropleth. The content in each of these visualizations can be changed and also can be performed as an animation to intuitively see the space-time tendency of the atmospheric pollution.**

Index Terms-Atmospheric pollution, Air quality, Data Visualization, Pollution Analysis, Animation, Tableau

I. Introduction

In our project, another one is an animation, which shows we create two subprojects. One of them is for mathematical analysis and the variants of the air quality data and it's for visualized purpose. In these two subprojects, we perform the air quality data of China with 4 visualizations. These data are provided by the ChinaVis2021, which contains the air quality data of China from 2013 to 2018, including the volume of several pollutants with the longitude and latitude of the specific location. Since the data size is quite big, we calculate the mean value of each of the attributes in the data set of a whole month to represent the data feature in this month. The data totally contains the air quality data in 5 years, 60 months. According to the data, we do the visualization design and create 4 visualizations, however, considering that the contents in the visualizations can be changed, there are totally more than 4 visualizations in our project. The 4 main visualizations of our project are two line charts, one scatter plot and a choropleth. The two line charts shows the pollutants by month and year respectively, the scatter plot shows the variation of PM2.5 and PM10 in 60 months. The choropleth shows the pollution standard of China by appending color on the map according to the specific pollutant value. Thanks to the help of Tableau, a data analytics software, the pollutant type in top left corner, the pollutant type showing in the visualizations can be changed. Similarly, to achieve the each of the visualizations can be change and the visualizations can also be performed as animations, they directly show the variant of pollutant value during 2013 to 2018. By modifying the choice box on the transformation from graph to animation, find the

symbol → on the top left corner. After doing data analysis to these visualizations, we draw a conclusion to the possible reasons for the atmospheric pollution.

II. Data Description

The data set of this project comes from the 2013 – 2018 China high resolution air pollution reanalysis open data set provided by "ChinaVis 2021" data visualization competition.(http://naq.cicidata.top:10443/chinavis/opendata)

The data is the national air quality reanalysis data based on geospatial grid and corresponding meteorological data, including 13 attributes of six conventional pollutants (PM2.5, PM10, SO2, NO2, Co, O3), wind speed (U&V), temperature (TEMP), air pressure e(PSFC), relative humidity (RH), longitude (lon) and latitude (lat). As shown in the figure below:

	PM2.5(微克每立方米)	PM10(微克每立方米)	SO2(微克每立方米)	NO2(微克每立方米)	CO(毫克每立方米)	O3(微克每立方米)	U(m/s)	V(m/s)	TEMP(K)	RH(%)	PSFC(Pa)	lat	lon
0	20.21	23.56	8.65	4.58	0.29	63.44	-2.48	-2.31	291.54	68.14	100350.11	18.34	109.25
1	20.41	23.96	9.28	5.24	0.29	62.53	-3.43	-3.21	291.96	67.41	100546.63	18.34	109.38
2	21.17	24.78	10.07	5.47	0.30	62.24	-3.43	-4.72	291.37	69.59	100428.13	18.33	109.52
3	22.24	26.04	10.36	5.28	0.30	63.58	-2.92	-5.32	290.74	72.94	100569.27	18.33	109.66
4	19.68	22.34	7.62	3.58	0.29	66.21	-2.51	-3.13	291.67	65.97	101620.34	18.49	108.83

Figure 1 Overview of data

The data includes the data of air pollutants in China for five years. Because of the large amount of data, we integrate the data of each month to calculate and get the average value of each attribute of each month. Once a month for 60 months. Because the original data has a CSV file every month, and the file name contains the date but the date is not included in the data. Therefore, we need to merge the 60 months file data and add a date column into the data to integrate it into the CSV file named "data".

The processed data is shown in the figure:

	PM2.5(微克每立方米)	PM10(微克每立方米)	SO2(微克每立方米)	NO2(微克每立方米)	CO(毫克每立方米)	O3(微克每立方米)	U(m/s)	V(m/s)	TEMP(K)	RH(%)	PSFC(Pa)	lat	lon	Name.4.1.1
0	0.24	0.27	0.04	0.11	0.13	90.05	3.23	3.14	265.14	32.93	52504.92	33.36	85.26	201511
1	0.23	0.26	0.04	0.11	0.13	90.12	3.99	2.68	266.11	30.98	53304.02	33.38	85.42	201511
2	0.20	0.23	0.04	0.11	0.13	84.57	4.17	4.27	268.39	27.57	55118.07	33.43	85.74	201511
3	0.19	0.22	0.04	0.11	0.13	83.43	3.95	4.38	269.58	25.19	55898.88	33.46	85.90	201511
4	0.23	0.27	0.04	0.11	0.13	88.44	3.78	4.40	266.33	29.81	53240.33	33.49	85.23	201511
5	0.23	0.26	0.04	0.11	0.13	90.06	4.26	3.46	266.00	30.19	52992.27	33.52	85.39	201511
6	0.22	0.25	0.04	0.11	0.13	88.57	4.57	3.18	267.34	27.95	53933.20	33.54	85.55	201511
7	0.20	0.23	0.04	0.11	0.13	84.97	3.98	3.04	270.24	24.07	56309.91	33.57	85.71	201511
8	0.18	0.21	0.04	0.11	0.13	82.89	2.78	2.62	271.07	23.54	57244.04	33.59	85.87	201511
9	0.22	0.26	0.04	0.11	0.13	86.79	4.06	4.13	268.45	26.45	54891.38	33.63	85.20	201511

Figure 2 Processed data

Observe the type of data. From the figure below, you can see that all 13 attributes are float data, and the date column is integer.

```
PM2.5(微克每立方米)     float64
PM10(微克每立方米)      float64
SO2(微克每立方米)       float64
NO2(微克每立方米)       float64
CO(毫克每立方米)        float64
O3(微克每立方米)        float64
U(m/s)            float64
V(m/s)            float64
TEMP(K)           float64
RH(%)             float64
PSFC(Pa)          float64
lat               float64
lon               float64
Name.4.1.1          int64
dtype: object
```

Figure 3 Features of data

If there are missing data in the imported data, it will affect the data visualization. So, we can check whether there are missing values in each row and column. From the figure below, we can observe that there are no missing data in all rows and columns.

```
df.isnull().sum(axis=0)#
```

```
PM2.5(微克每立方米)     0
PM10(微克每立方米)      0
SO2(微克每立方米)       0
NO2(微克每立方米)       0
CO(毫克每立方米)        0
O3(微克每立方米)        0
U(m/s)            0
V(m/s)            0
TEMP(K)           0
RH(%)             0
PSFC(Pa)          0
lat               0
lon               0
Name.4.1.1        0
dtype: int64
```

Figure 4 Data preprocessing

III. Visualization Design

A. Graph

The dataset contains the latitude and longitude information, and the longitude and latitude on the content of each of the atmospheric pollutants, and also includes temperature, humidity and atmospheric pressure. Therefore, we need to analyze the changes of various atmospheric pollutants in the whole China, and the relationship between atmospheric pollutants and year and month, as well as the interconnection between atmospheric pollutants.

In order to do the above analysis, we designed 3

kinds of visualizations:

1. Choropleth map: In the dataset, every point has its longitude and latitude, so we present these points in the form of a Chinese map and encode the intensity of the pollutants into the color channel. The severity of various atmospheric pollutants and the geographical distribution of the pollutants can be visually displayed on the map, as shown below.

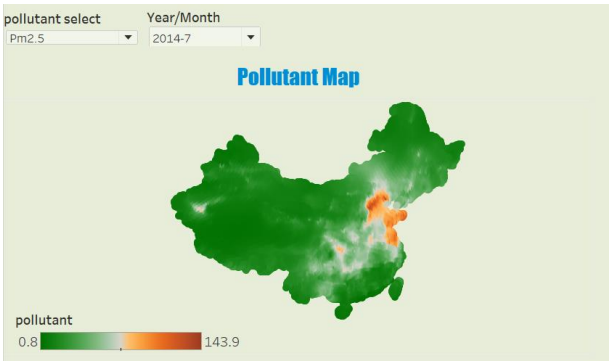


Figure 5 Choropleth map

2. Line chart: The time is encoded in X-axis and the pollutant is encoded in Y-axis, we designed 2 line charts to show the annual and monthly changes of average atmospheric pollutants, as shown below.

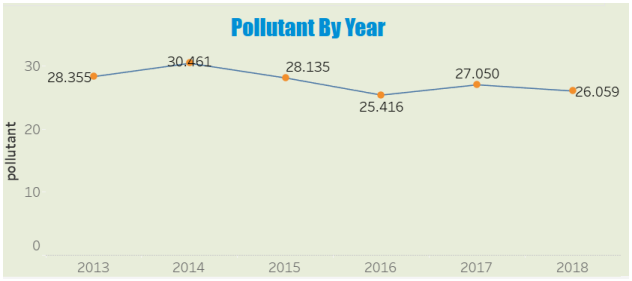


Figure 6 Line chart by year

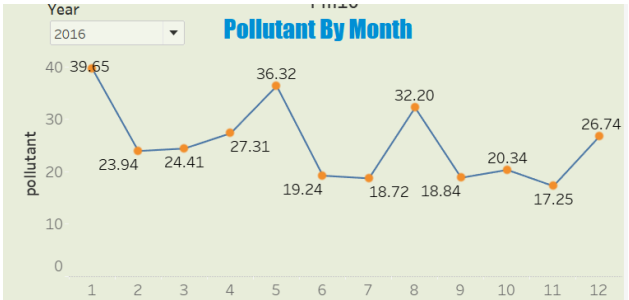


Figure 7 Line chart by month

3. Scatter plot: The scatter plot shows the correlation between 2 atmospheric pollutants, 2 pollutants are encoded in X-axis and Y-axis, respectively, as shown below.

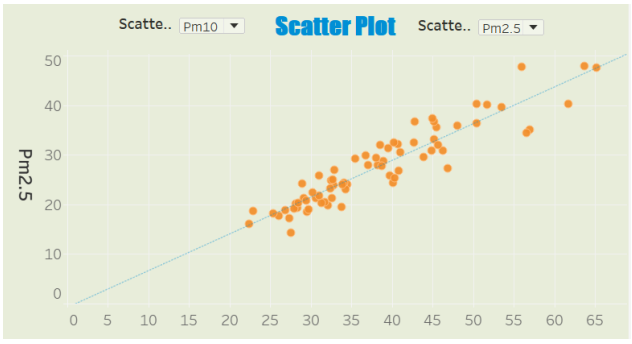


Figure 8 Scatter plot

B. Interaction

Our visualization allows user to do the following interaction:

1. Select: The user can select all the points in the graphs to see their information.
2. Filter: The user can filter data conditionally.
3. Explore: The user can choose to show different kinds of pollutants for all the graphs and different years in the second line chart, selecting a pollutant will update all the graphs.
4. Zooming: The user can zoom in and zoom out in the map.
5. Play animation: We also created a map (ATMOSPHERIC POLLUTANT ANALYSIS V2.twbx) that can be played in the form of animation to show the change of pollutants by month.

C. Dashboard

V1:

No animation in map, only filtering.

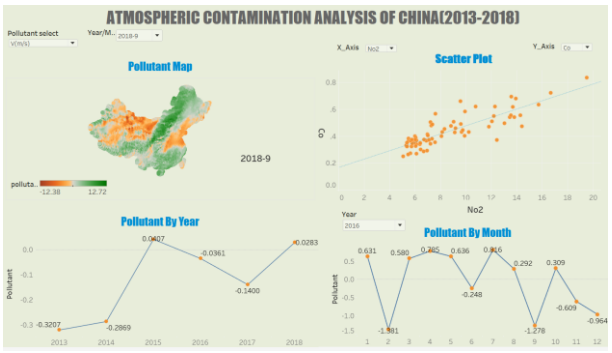


Figure 9 Dashboard V1

V2:

Animation in map.

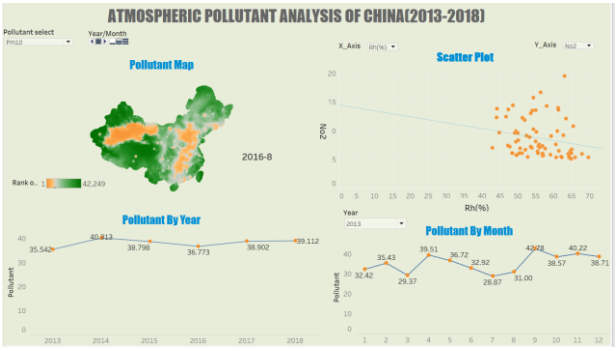


Figure 10 Dashboard V2

IV. Visualization analysis

The result is based on atmospheric pollutant on the latitude and longitude in China. And the pollution level is based on the data attribute we extract from the csv files. There are totally 6 indicators for measuring the pollution level. You can select any one of it by using the instrument penal we designed.

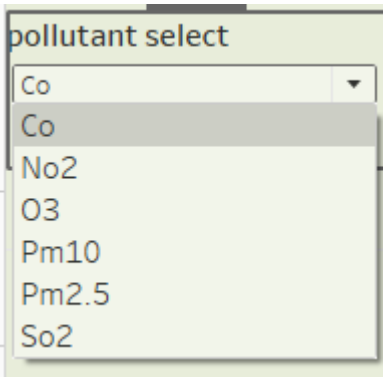


Figure 11 Selections

For CO pollution source, we can see from the below figure, color green represents the low pollution level. And red represents the high pollution level. In 2013, whole China basically have no such pollution source.

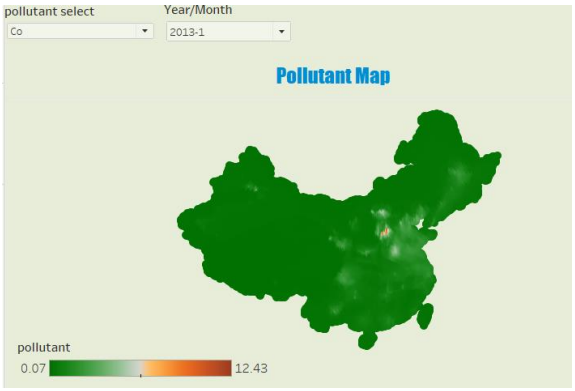


Figure 12 CO pollution in China in 2013-01

However, as time move on, the CO pollution is increasing. Until 2018, the pollution saw a significant decrease.

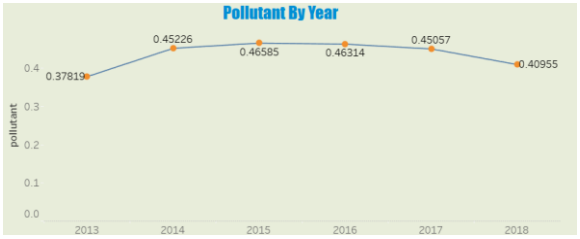


Figure 13 Line chart for CO pollution over years

And in those of the high CO pollution years, the pollution is concentrating on the central region in China.

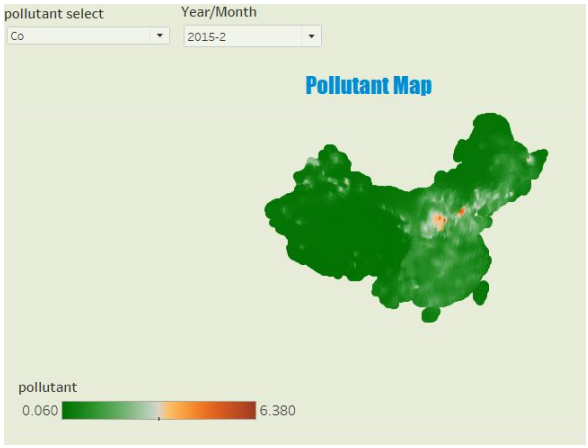
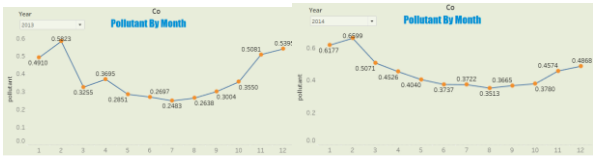


Figure 14 CO pollution in China in 2015-02

And in almost all the years, the pollution is relatively lower in summer than in winter.



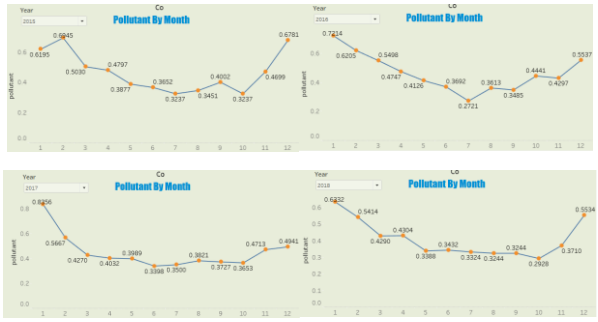


Figure 15 Line charts for CO pollution over months in each years

Then, we move to next pollutant No2, we select whole data set to show the overview. And we can see that the pollution area for No2 is located in the central east of China including our capital city Beijing according to the latitute and longtitute.

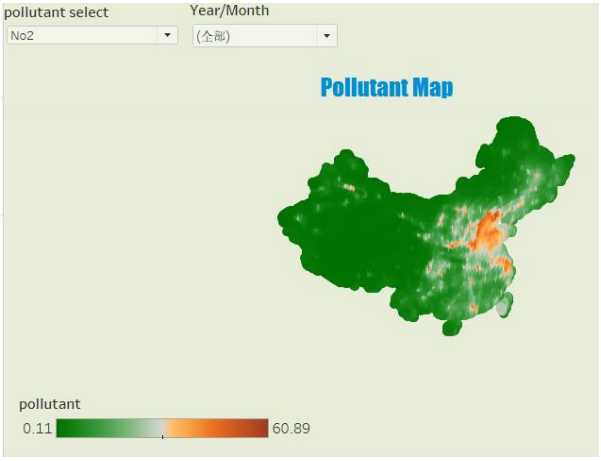


Figure 16 No2 pollution in China

And the pollution by year for No2 is continuously increasing from 2014 to 2017.

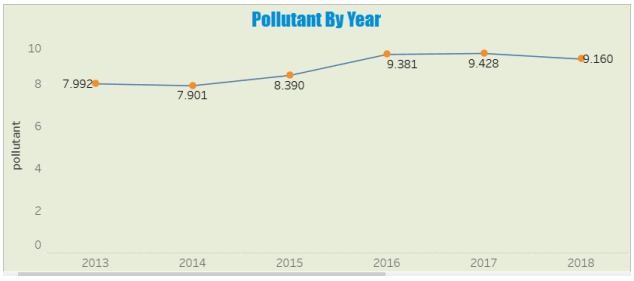


Figure 17 No2 pollution in China over years

For No2 pollutant, it is more severe in summer than winter.

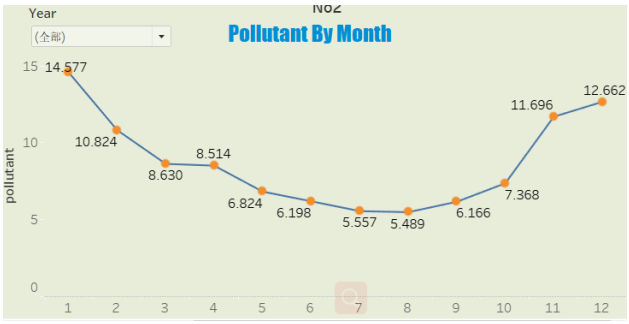


Figure 18 No2 pollution in China over months

The last gas pollution is So2, similar with Co and No2, the pollution occurs in the central region of China.

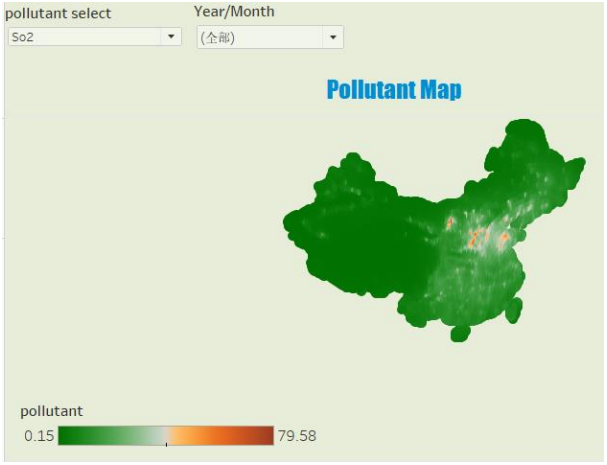


Figure 19 So2 pollution in China

It continuously decreases from 2013 to 2018.

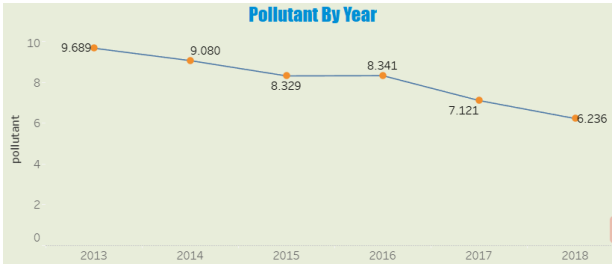


Figure 20 So2 pollution from 2013 to 2018

Same as Co and No2, the concentration of So2 is much higher in winter than in summer.

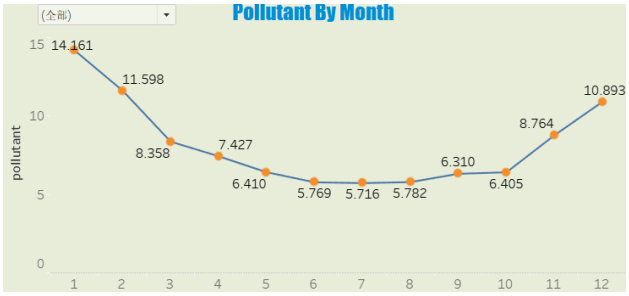


Figure 21 So2 pollution in China over months

After mearing No2, Co and So2, which are two gas pollutant that we often hear about in the News, we move to a special gas pollutant which is O3. From the geographical visualization, we found that the O3 concentration is higher in plateau in China than in plain.

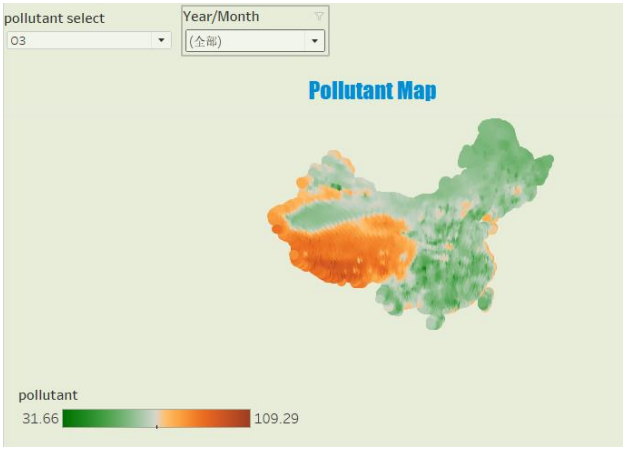


Figure 22 O3 pollution in China

And it is relatively stable during 2013 to 2018.

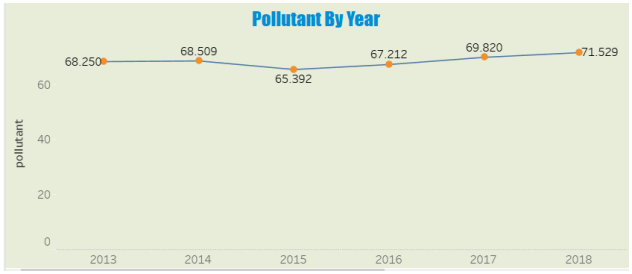


Figure 23 O3 pollution in China over years

Different from the previous three gas pollution, its concentration is higher in summer.

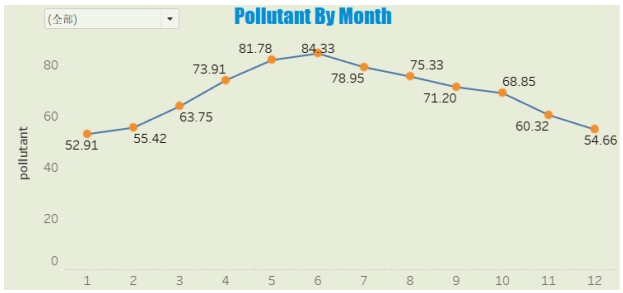


Figure 24 O3 pollution in China over months

More interestingly, O3 is negatively correlated with the other air pollutant. After searching the internet, O3 is actually not a strict measurement for air pollution. But if the concentration is too high, it will be harmful for people and plants. After some searching on the internet, we found that O3 can be good for human to purify the drinking water, washing clothes and even purify the air.

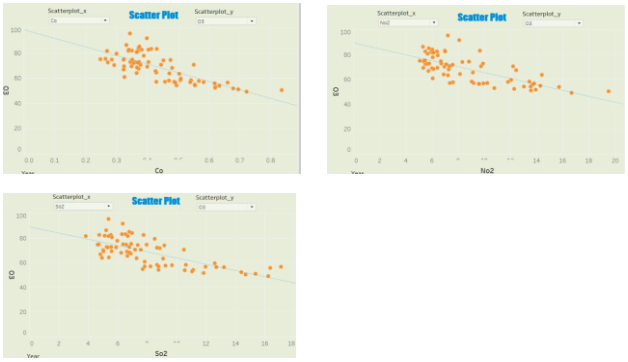
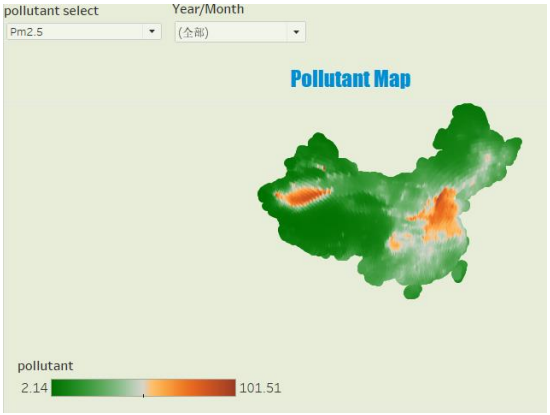


Figure 25 Scatter plot between O3 and the other gas pollution variables

Above four are all gas pollution in the environment, now let's see some particle pollution which is widely known as PM2.5 and PM10.

From the below two picture, we can see that the particle pollution is mainly distributed in the central eastern China and the dessert area in westearn China.



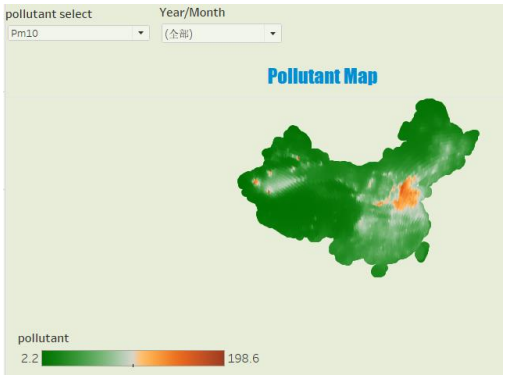


Figure 26 Particle pollution in China (PM2.5, PM10)

For both PM2.5 and PM10, the concentration firstly increase and then decrease after 2014.

Figure 27 Particle pollution in China over years (PM2.5, PM10)

And there is no clear rules for seansons. It fluctuate based on months in different years.

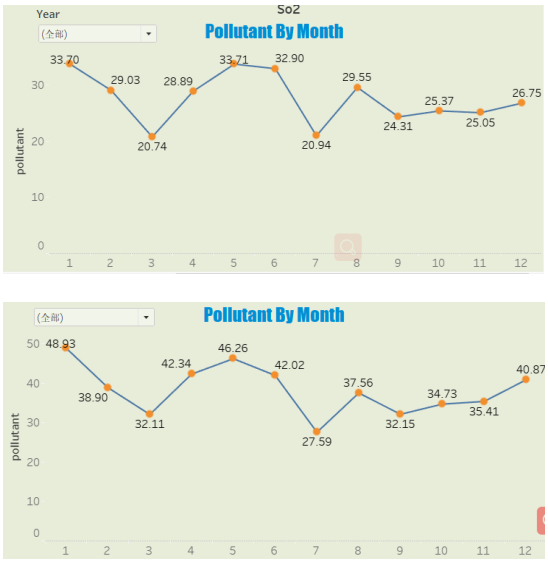


Figure 28 Particle pollution in China over months (PM2.5, PM10)

Finally, we tried to use the scatter plot to discover the correlation between the 6 pollutants and the 5 atmospheric environmental factors, which are wind speed(U&V), temperature(TEMP), air pressure(PSFC), relative humidity(RH), the outcome is shown below.

1. Temperature(TEMP) has a strong negative correlation with CO, NO2, SO2 and a strong positive correlation with O3. On the contrary, wind speed(U) and air pressure(PSFC) have a strong pisitive correlation with CO, NO2, SO2 and a strong negative correlation with O3.
2. There is no clear correlation between pm2.5, pm10 and the 5 atmospheric environmental factors.

V. Conclusion

Our whole project is divided into four parts. We start with environmental problems across China nowadays. And get the atmospheric pollutant data from the DataVis contest. After collecting the data, we did some Data analysis to see how each attribute of the data distributed and make some description of all the attributes of the data. Then, we use Tableau to do the visualization design including an animation about the degree of atmospheric pollutant during the period of the year 2013 to 2018. We can view from both the geographical perspective and time perspective. From the analysis part, we found that the air pollution can be divided into gas pollution and particle pollution. For gas pollution, O3 is a special measurement since it is opposite to the other gas pollution measurement. For particle pollution, it is hard to find a trend. Using this visualization dashboard, we are able to explore many kinds of information and correlation conveniently and easily.

References

Pengyu Chen. Visualization of real-time monitoring

datagraphic of urban environmental quality[J]. EURASIP Journal on Image and Video Processing, 2019, 2019(1) : 1-9.

Cleary Ekaterina Galkina et al. Making Air Pollution Visible: A Tool for Promoting Environmental Health Literacy.[J]. JMIR public health and surveillance, 2017, 3(2) : e16.

Du Yi et al. A Visual Analytics Approach for Station-Based Air Quality Data.[J]. Sensors (Basel, Switzerland), 2016, 17(1)

Appendix

Link of tableau file:
Link of Baidu Netdisk:
https://pan.baidu.com/s/11rWPUlH-guawoWQ2NP3VOA
Extract code：swx9