Stat 605 Final Project Report

Yukun Fang, Mengkun Chen, Runze You, Mengqi Li

STAT 605 Group 10

November 2020

Outline

Introduction

- In Sofia, national capital of Bulgaria, air pollution norms were exceeded 70 times in the heating period from October 2017 to March 2018.

- Sofia air quality data set describes daily five air-quality measurements-PM2.5, PM10, and climate statistics pressure, temperature and humidity from July 2017 to June 2019.

Pollution Data Description

- A glance at the pollution data sets.

	sensor_id	location	latitude	longitude	timestamp	PM2.5	PM10
1	2888	1453	42.71	23.28	2017-08-01T00:00:01	7.52	6.48
2	3641	1837	42.69	23.36	2017-08-01T00:00:02	10.70	6.10
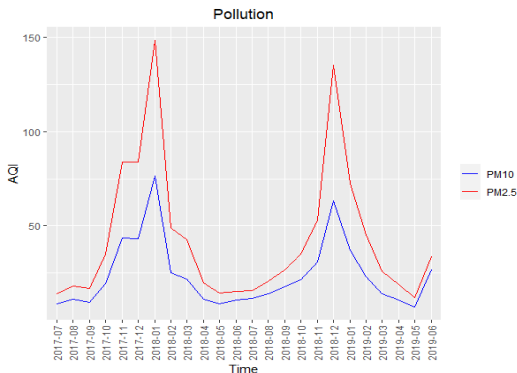
Pollution Data Description

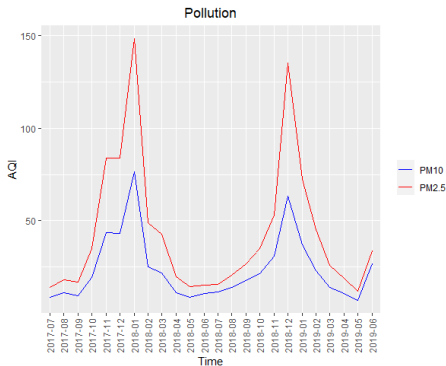Here are two videos of the change of PM10 and PM2.5.

Figure: A PyMOL movie object embedded in PDF, using mpg format.

Climate Data Description

- A glance at the pollution data sets.

|	sensor_id	location	latitude	longitude	timestamp	pressure	temperature	humidity
1	2266	1140	42.74	23.27	2017-07-01T00:00:07	95270.27	23.46	62.48
2	2292	1154	42.66	23.27	2017-07-01T00:00:08	94355.83	23.06	59.46
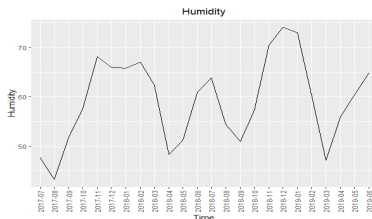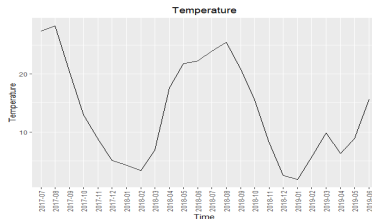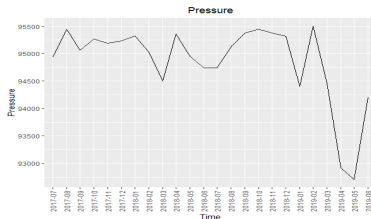
Statistical analysis

- According to the plots above, we find that the pollution plot is a seasonal plot which contains two peaks in the winter month. So, from following box plots and p-value from t-test, winter is a time which the pollution is so heavy.
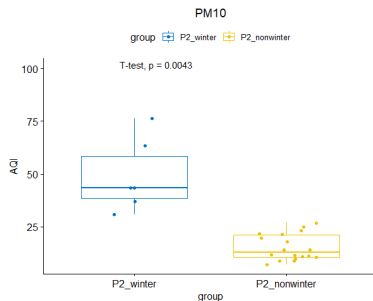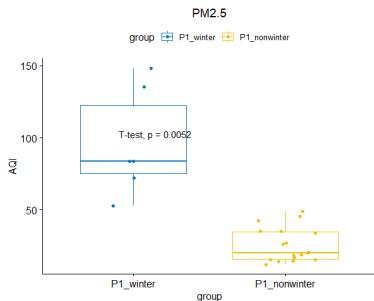
Statistical analysis

- More precisely, the following correlation plot reveal clear relationships between PM2.5/PM10 and temperature/humidity.
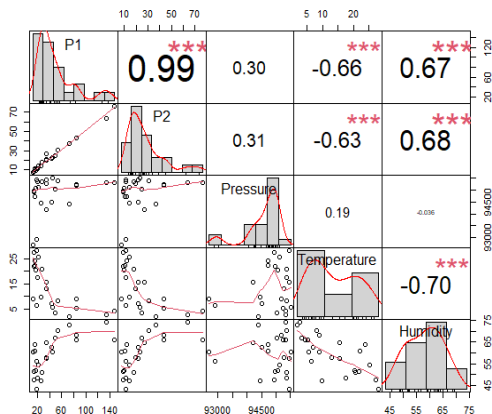
Difficulties

It is difficult for us to show every day of two years. Actually we would
make efforts to overcome that analyze every time in the data. Drawing
pictures and run the code in CHTC cost a lot of time.

Conclusion

- Sofia has a serious pollution problem which is especially in winter.
- It has relationship with temperature and humidity which git us a great signal to predict it.
- What we need to do in the future is to research the terrain of Sofia to make it more reasonable