

Stat 605 Final Project Report

Yukun Fang, Mengkun Chen, Runze You, Mengqi Li

November 2020

1 Introduction

"In Sofia, air pollution norms were exceeded 70 times in the heating period from October 2017 to March 2018", citizens' initiative AirBG.info says. During the time, the air pollution has become a serious problem for Sofia. This problem attracted the researchers all around the world.

Air Quality Index has been used to determine the level of air pollution across different regions worldwide. As part of it, the level of particulate matter (PM) is measured as well. This is the term used for a mixture of solid particles and liquid droplets found in the air.

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

For this dataset, we try to figure out whether there are some trends in PM2.5 and PM10 and whether there are some relationships between climate variables and air-quality measurements. We draw time series plots, conduct t-test, make short videos to answer these questions. We find that both meteorological measurements and air-quality measurements (PM2.5 and PM10) influenced by season and the air pollution problems are serious during the winter. Furthermore, there exist relationships between temperature, humidity and PM2.5, PM10.

2 Analysis

2.1 Data description

Meteorological measurements and air quality measurements are recorded between July 2017 and June 2019. There are two kinds of datasets which contain the air quality data and meteorological data separately. We make **videos** to show the monthly change in PM10 and PM2.5 by longitude and latitude. Here are two geographical distribution maps in our videos.

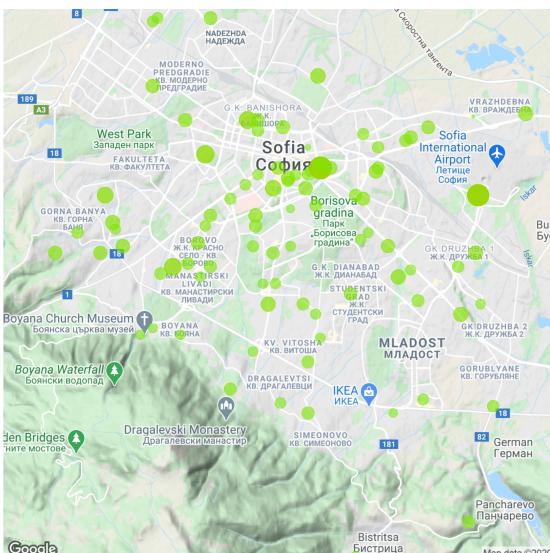


Figure 1: PM2.5 in June 2017

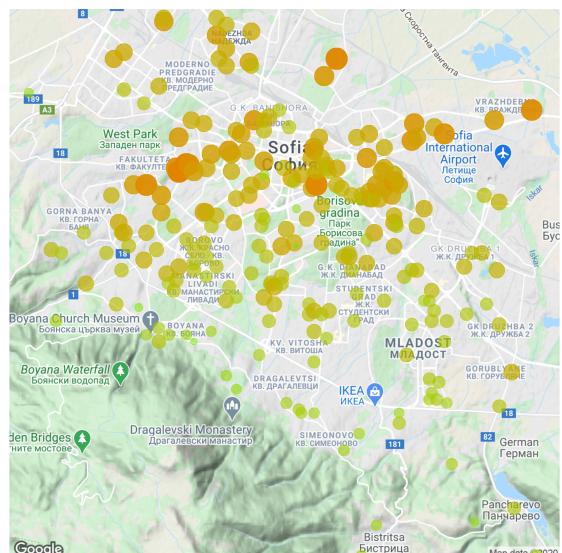


Figure 2: PM2.5 in January 2018

Animation shows that pollution is more serious in winter. From July 2017 to June 2019, areas with more residential communities or with large populations are more polluted than other areas. We noticed that the vicinity of West Park, which locates in the west of Sofia, is quite polluted in the recorded time span. We will

focus on this area to explain our findings.

- Variable Explanation

sensor_id: sensor identification. Sensors record air quality measurements and meteorological measurements around every five minutes.

lat, lon: latitude and longitude coordinates, recording geographical information of each sensor. They are not 1-to-1 matches due to precision of the data.

P1: PM2.5 concentration

P2: PM10 concentration

Pressure, Temperature, Humidity: meteorological measurements reported by sensors.

- Air Quality Measurements

Table 1: Air Quality Data Overview

	sensor_id	location	lat	lon	timestamp	P1	P2
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

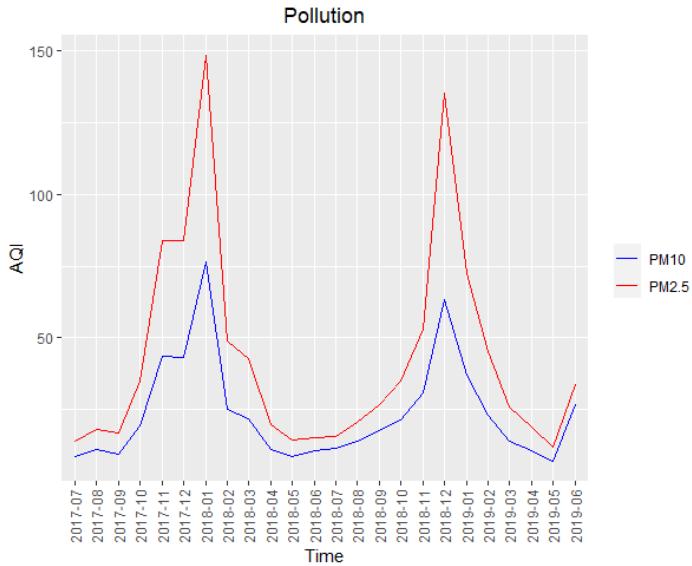


Figure 3: Time series plot for PM2.5 and PM10

Figure 3 indicates that both air quality measurements have seasonal patterns. They go to the highest level at around January each year and drop to the lowest level in May. Concentration of PM10 is generally lower than that of PM2.5.

- Meteorological Data

Table 2: Meteorological Data Overview

	sensor_id	location	lat	lon	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

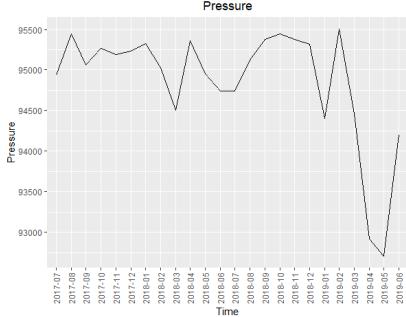


Figure 4: Ts plot for pressure

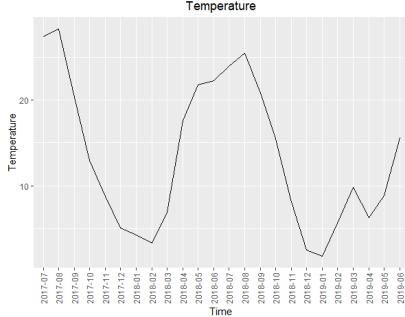


Figure 5: Ts plot for temperature

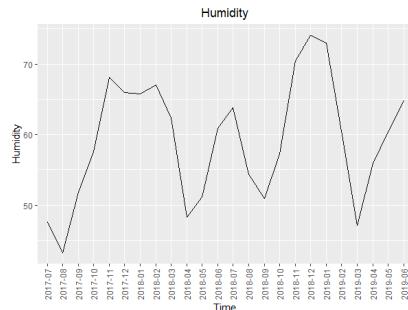


Figure 6: Ts plot for humidity

Pressure in the recorded time span does not show any seasonal pattern. However, humidity shows similar pattern with two air quality measurements. Humidity reaches its peak at around December every year. This can probably be explained by the climate and topographical characteristics around. Since Sofia locates in northern hemisphere, the temperature is high in summer and low in winter.

2.2 Statistical analysis

- Box plots for air quality measurements

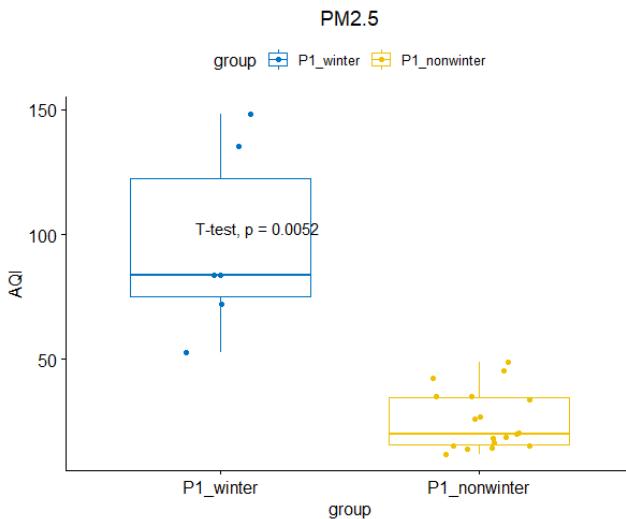


Figure 7: PM2.5

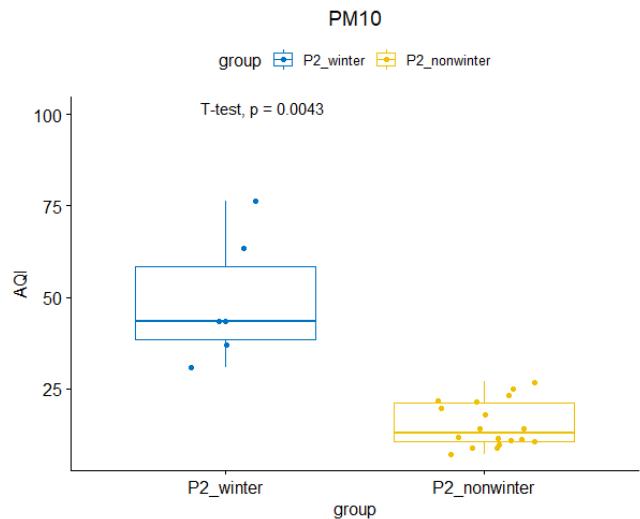


Figure 8: PM10

According to the box plots, we find that both PM2.5 and PM10 concentrations are higher in winter months, which are November, December and January.

- t-test results

Table 3: Test for difference between winter months and non-winter months

	winter mean	non-winter mean	t-statistics	df	p-value
PM2.5	96.0452	25.5388	4.5526	5.3341	0.0052
PM10	49.0558	15.2183	4.6905	5.4434	0.0043

From the test results, we can conclude that the pollution is heavier during the winter with great confidence.

- Correlation between air quality measurements and meteorological measurements

We also draw the correlation plot to see the relationship between air-quality measurements and meteorological measurements.

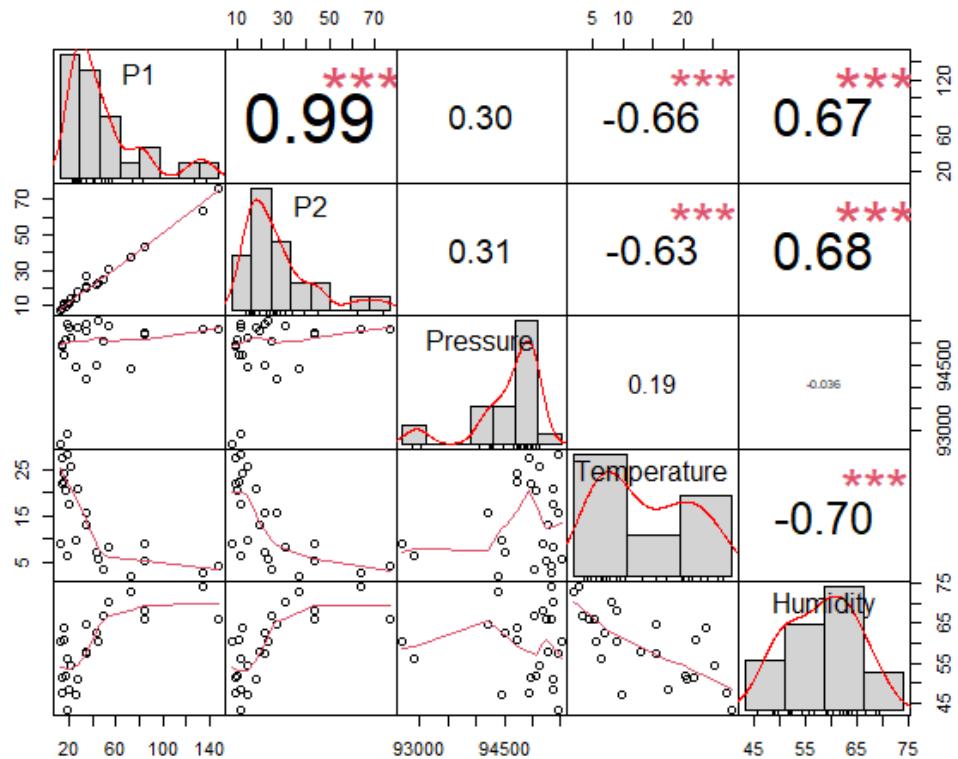


Figure 9: Correlation plot

Clearly PM2.5 and PM10 are highly correlated with temperature and humidity.

2.3 Difficulties

It is difficult for us to show daily data for two years. Actually we would make efforts to overcome that analyze every time in the data. Drawing pictures and running code in CHTC are quite time consuming. Plus, the datasets contain many bad records like -30 celsius temperature in June and it is tedious for us to identify.

3 Conclusion

Air pollution is a serious problem in Sofia during winter. It is correlated with temperature and humidity. Such correlation provides us a tool to predict air quality measurements and apply precautionary plans. What we need to do in the future is to research the terrain of Sofia to make it more reasonable.