PM2.5 Levels in Beijing:

Report prepared for Feifang Hu
By Kang Fu, Dandan Feng, Xinning Chu, Ashley Hall
George Washington University
Statistical Consulting Program (SCP)
April 28, 2019

Abstract

OBJECTIVE— Our main objective of this research is to determine if PM2.5 levels in Beijing have increased or decreased and give some suggestions to improve the air quality of Beijing.

METHODS— The data, taken from the U.S. Department of State website and an AQI study, consists of the time indicators: year, month, day, and hour, and measurements of the PM2.5 in $\mu g/m^3$, and a binary variable describing weather the measurement was missing or present (QC Name). Firstly, we developed a time series model to accurately show if these measurements are increasing or decreasing over time, and to predict how these concentration levels will behave in the future. Next, EDA and t-tests were applied to search patterns of PM2.5 values in Beijing.

RESULTS— The time series of monthly average PM2.5 values has a decreasing trend, which indicates that the air quality of Beijing is getting better. In addition, some factors, such as heating and policy issued by China government, have a significant influence on PM2.5.

CONCLUSION— From results of data analysis, suggestions are proposed. For example, improving heating system, control the time of heating and reducing car dependence are necessary.

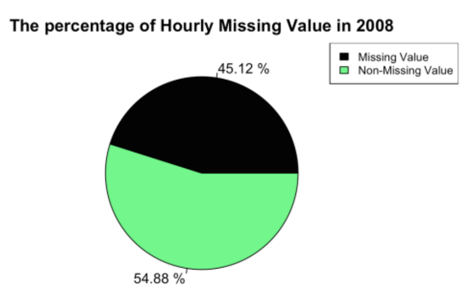
1. INTRODUCTION

PM2.5 is the most well-known measure of air quality and pollution in our cities – it is a particulate matter that enters your body through the nose and mouth. The body will eliminate a lot of the larger particles, such as PM10 (less than 10 microns in diameter) which includes dust, pollen, mold, etc., but PM2.5 (less than 2.5 microns in diameter) cannot be processed by the body. So, these smaller particles, such as combustion particles, organic compounds, metals, etc. severely damage our bodies main systems. For the respiratory system, the insoluble part accumulates at the alveolus of the lungs, causing inflammation. For the blood system, high PM2.5 concentrations cause blood toxicity, blood coagulation abnormalities, and can trigger heart disease. For the cardiovascular system, PM2.5 levels cause cardiotoxicity and severe irritation to the autonomic nervous system, which regulates the activity of the heart muscle. Lastly, for the reproductive system, PM2.5 levels lead to placental blood toxicity, which leads to direct harm to the fetus, intrauterine growth, and low birth weight in babies.

Looking holistically at the bigger, scarier, picture, air pollution truly is the silent killer. Every year, around 7 million deaths occur worldwide due to exposure to both outdoor and household air pollutants. India is home to six of the ten cities in the world with the worst air pollution, but China, and Beijing specifically, have a long way to go to steer clear of these lower ten cities and their pollutant measures. In September 2013, the State Council unveiled the "Action Plan for the Prevention and Control of Air Pollution," which aimed to increase air quality over a five-year period. Beijing, specifically, was asked to bring down PM2.5 concentrations to around 60 micrograms per cubic meter. So, in March 2014, the US Embassy in Beijing released a trove of air quality data, "fueling a big step at the government level: declaring war on air pollution. One of the reasons for that is that the health argument was very strongly presented, and the fact that the citizens were really breathing air that was totally unacceptable." This inspired Chinese officials to launch their own monitoring operations around the country. So, taking all these events we stated above into consideration, as for this project, our main objective is to examine if Beijing's PM2.5 levels have increased or decreased, thus looking at if the country's air quality is getting better or worse.

2. BACKGROUND AND DATA SET

For this project, we have two data sources: The US Department of State and an AQI Study. The US Department of State dataset consists of measurements from 2008 through June 2017 and contains the hourly estimates of PM2.5 in units of µg/mg³, with a final binary variable stating if the measurement is valid or missing. By checking the percentage of the missing values in this dataset, we found that the ratios of missing values in 2008 and 2009 are too high (shown in Figure 2.1 below), so we decided to not include these years in the study.



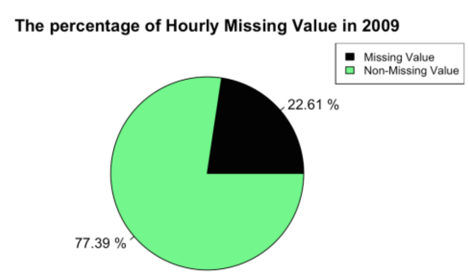
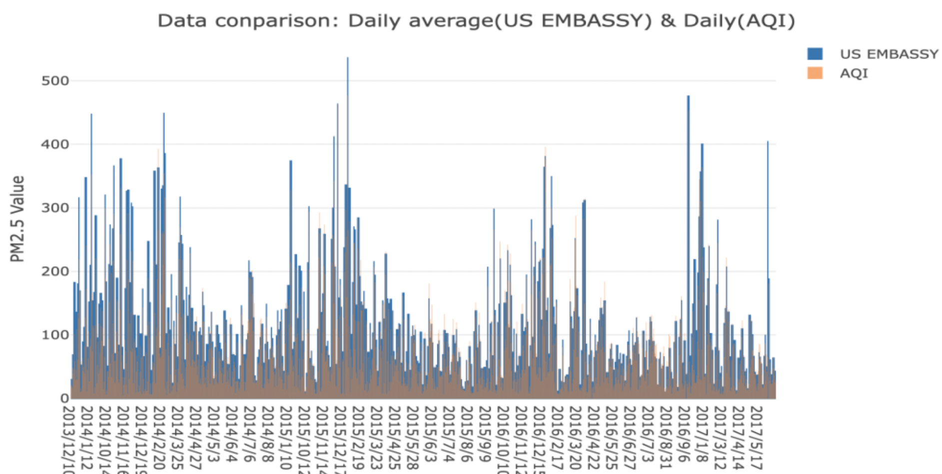


Figure 2.1: Ratios of missing values in 2008 and 2009

So, for the US Department of State dataset, we are now moving forward with just data from 2010 through June 2017. Also, using the hourly estimates of PM2.5 values in this dataset, we calculated the daily average PM2.5 value and the monthly PM2.5 value.

Then, take a look at the data from the AQI study group. The AQI Study dataset consists of measurements of the daily average and monthly average PM2.5 value from 2013 to April 2019, and there is only one missing value in the entire dataset: April 16, 2016. So, a good bit of the initial discovery phase is to the discovery phase of overlap measurements between the two datasets we mentioned above: US Department of State (now 2010 – June 2017) and AQI Study (2013-April 2019). To see this relationship between the two datasets, we will compare both visually and statistically for the overlapped period of 2013 through June 2017.

Visually, we look at comparisons for daily average and monthly maximum PM2.5 values. See Figure 2.2 below for these plots, with blue representing the US Department of State measurements and yellow representing the AQI Study.



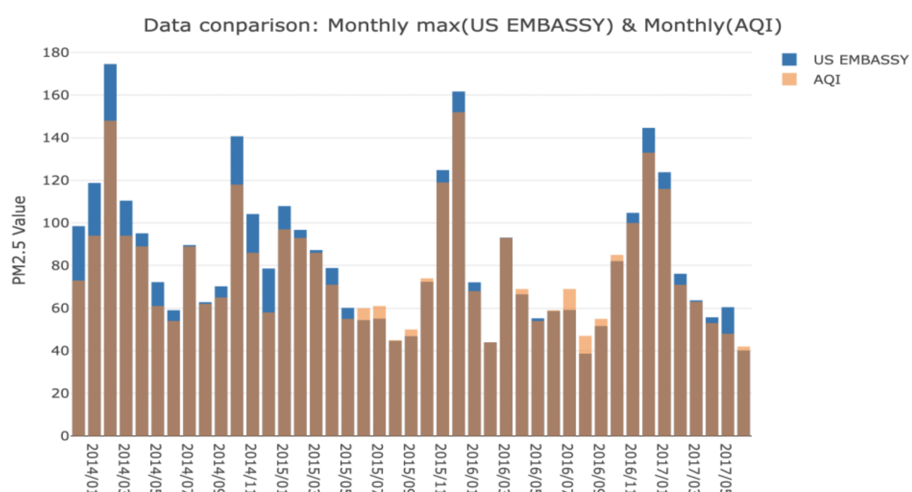


Figure 2.2: Comparisons for PM2.5 values in two datasets

From the plots, we see that the daily data in the AQI Study seems to coincide with the US Department of State data. From here, we use two statistical methods to strengthen the thought that the two datasets coincide. Firstly, Pearson correlation, which is a number between -1 and 1 that indicates the extent to which two variables are linearly related. The Pearson correlation is also known as the “product moment correlation coefficient” (PMCC) or simply “correlation”.

The resulting Pearson correlation coefficient between the US Department of State and AQI Study are shown in Table 1 below:

Table 1: Pearson correlation coefficient between two datasets

Year	Pearson Correlation
2014	0.9636894
2015	0.9841159
2016	0.9836109
2017	0.9034997

From what are shown in Table 1, we can now conclude that the two datasets are highly correlated. Secondly, the two-sample t-test, which is used to determine if two population means are equal. The null hypothesis states that there are no significant differences in population means, while the alternative hypothesis states that there are. The resulting p-values for this test per year are shown in Table 2 below:

Table 2: The resulting p-values for the t-test per year

Year	P-value
2014	0.1792
2015	0.6578
2016	0.9911
2017	0.5026

From what are shown in Table 2, we can now conclude that are no significant differences in population means for the two data sets. So, based on these visual and statistical conclusions, we can now link the two datasets together to get a new dataset that contains the data from 2010 through April 2019.

3. DATA ANALYSIS AND RESULTS

3.1 Exploratory data analysis

In this part, our goal is to discover the possible sorts of trends occurring in the data. We simply conducted two-sample t-tests to explore the possible variation trend of the PM2.5 value according to the years. So, we performed 4 tests: 2014 population mean vs. 2015 population mean, 2015 vs. 2016, 2016 vs. 2017, and 2017 vs 2018, and derived the p-values and estimated means for the years been compared shown in Table 3 below:

Table 3: P-values and estimated means for the years been compared		
Year	P-Value	Estimated mean: (x, y)
2014(x) vs 2015(y)	0.009186(**)	(97.83859, 82.54337)
2015(x) vs 2016(y)	0.06692	(82.54337, 72.89106)
2016(x) vs 2017(y)	0.003919(**)	(72.89106, 57.37808)
2017(x) vs 2018(y)	0.02834(*)	(57.37808,49.44658)

From this table, the estimated means for the years been compared seem to show a decreasing trend in measurements, maybe implying an improvement in air quality.

3.2 Time Series Analysis

To address the main objectives of the study, we conducted a trend analysis of PM2.5 values in the following ways: making plots (time plot, ACF/PACF plot), identifying the characteristics of ACF and PACF, trend test, building a seasonal ARIMA model, diagnostics on residuals and forecasting.

In order to derive some patterns of the air quality, we started by plotting pm2.5 values indexing by time based both on hourly and monthly average data, which can be seen as time series. The Figure 3.1 is the time plot of hourly PM2.5 values. In addition, how the monthly average pm2.5 values changed over time is shown in the Figure 3.2, and lines with different colors represent different data sources respectively.

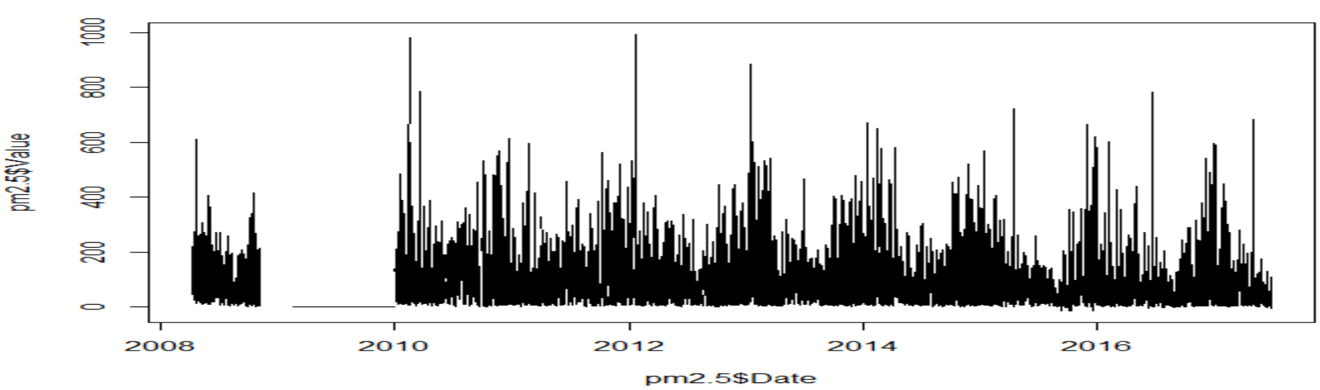


Figure 3.1 Time Plot of The Hourly Data

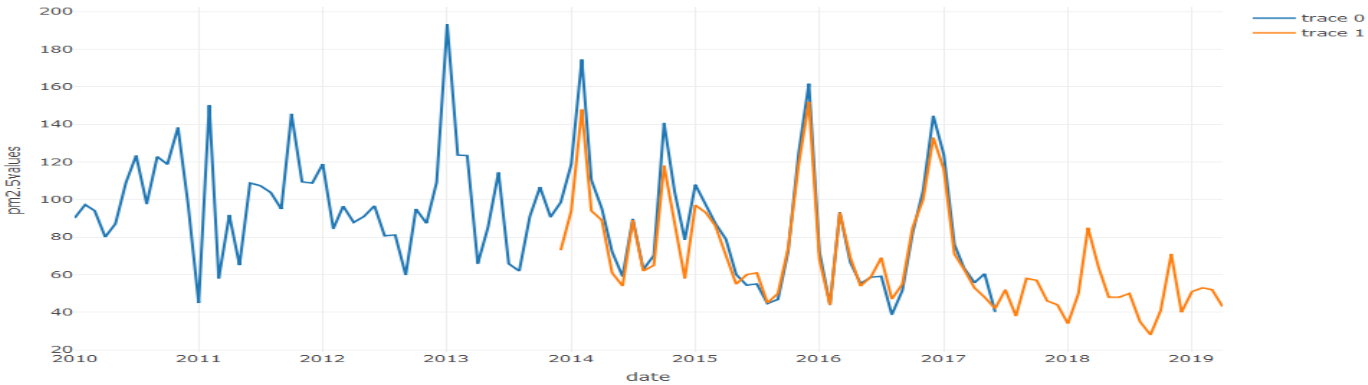


Figure 3.2 Time Plot of The Monthly Data

Then we plot the ACF and PACF for the monthly average PM2.5 values. According to the Figure 3.3, the sample PACF appears to tail off at lags 12k(k=0,1,2,...), whereas it's hard to inspect the behaviors of the sample ACF. So a seasonal ARIMA could be suggested. After that, decomposition function in R was applied to decompose the original data into a trend, a seasonal factor and a random error, just as the Figure 3.4 shows. In other words, this time series is sum of three factors. If we delete the seasonality and keep the other two terms, a new time series would be created, which is shown in the Figure 3.5.

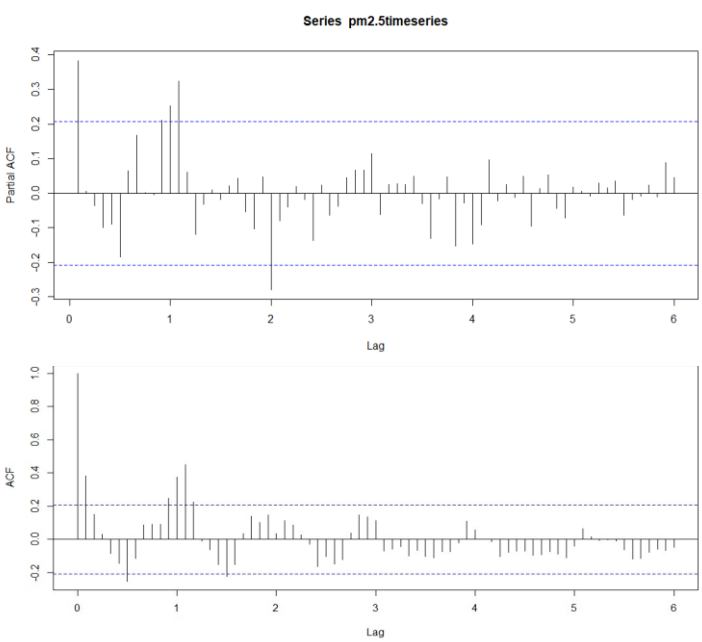


Figure 3.3: ACF and PACF Values

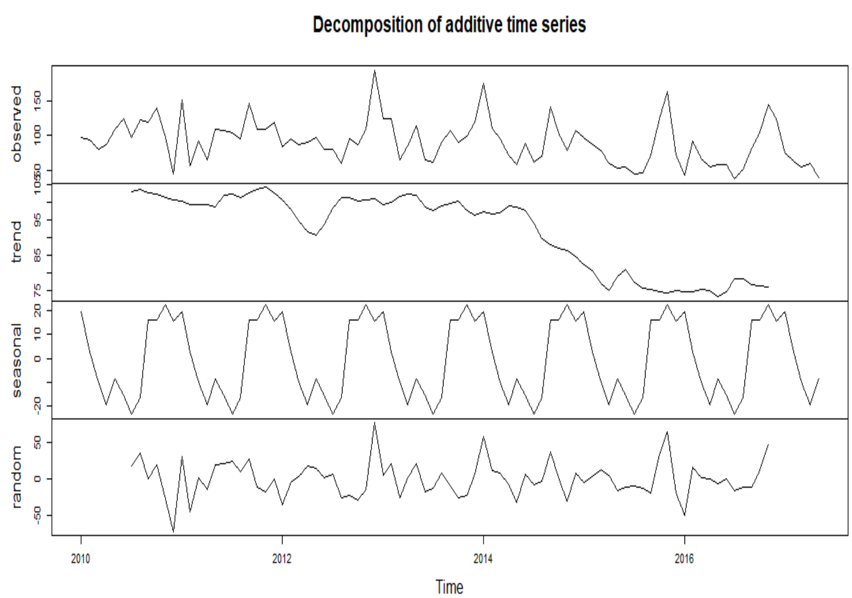


Figure 3.4: Decomposition of the Monthly Average PM2.5

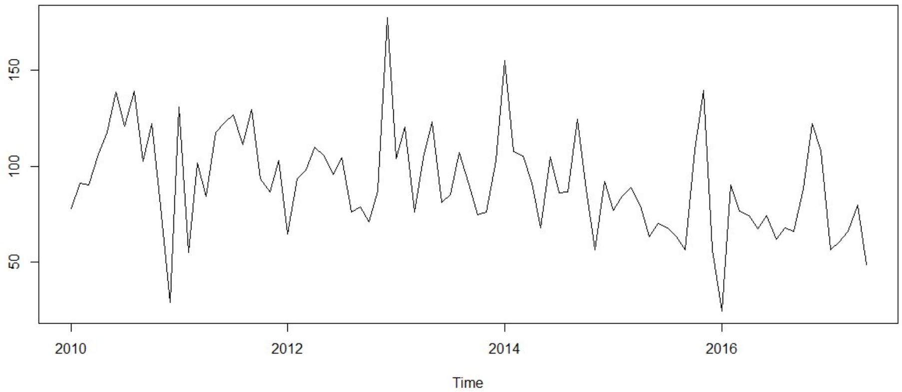


Figure 3.5: Time plot of PM 2.5 concentration without seasonal factor

According to Figure 3.4, there seems to be a decreasing trend of the time series. Thus the next step is to implement statistical methods to determine the existence of a trend. Here Mann-Kendall test was used. It's a nonparametric test used to identify a trend in a series, even if there is a seasonal component in the series.   So we performed Mann-Kendall test for the original time series. Besides, this test also was applied to the time series without seasonal factor to improve the accuracy of our study. Results are shown as following.

Table 4: Results of Mann-Kendall Test				
Variable	Z value	p-value	S	tau
PM2.5	-3.71	0.000203	-1067	-0.27
PM2.5 without seasonal factor	-4.19	2.799e-05	-1203	-0.30

From the table above, for PM2.5 the p-value is smaller than 0.05, thus the null hypothesis that it doesn't have a monotonic trend is rejected. And the Z value is less than Z(1-α), where Z(1-α) is the 100(1−α)th percentile of the standard normal distribution. This suggests a downward monotonic trend. We can also draw the same conclusion from the results of test for non-seasonal time series.

Parameters Estimation:

Model: (0,1,1)(0,0,1)[12]

```
Coefficients:
          ma1    sma1  constant
      -1.0000  0.4949   -0.3593
s.e.   0.0309  0.1154    0.1435

$ttable
         Estimate     SE  t.value p.value
ma1       -1.0000 0.0309 -32.3951  0.0000
sma1       0.4949 0.1154   4.2893  0.0000
constant  -0.3593 0.1435  -2.5031  0.0142

$AIC
[1] 7.612872

$AICc
[1] 7.640694

$BIC
[1] 6.696758
```

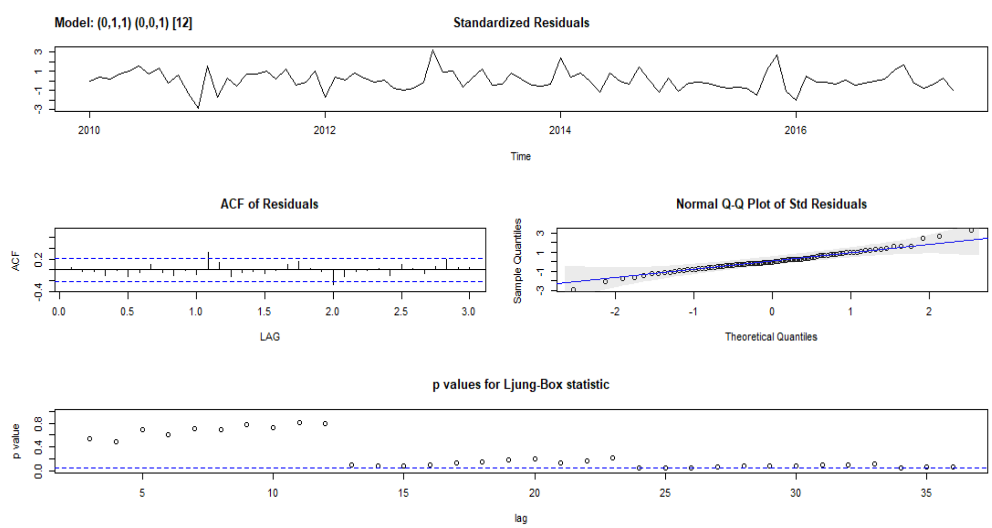


Figure 3.6: Parameters Estimation and model diagnosis

After stating this conclusion based on our trend testing procedures, we fit a SARIMA model. The residual diagnostics are shown in Figure 3.6, and except for one or two outliers, the model seems to fit well. Using it we could forecast Beijing's PM2.5 measurements out slightly (not 10/15 years) to test accuracy of our SARIMA model. **Since time series can fluctuate greatly, it is not best-practice to project years in advance, but better to input more data to refine model and project short-term future measurements.** So, see below for these short-term forecasts. We think it's important to note the variance between the forecasted and true values for our built SARIMA model. The makes it difficult to track and predict natural fluctuations in the time series.

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95	True Value
Jul 2017	76.65289	41.61452	111.69126	23.066327	130.2395	118
Aug 2017	65.71653	30.62840	100.80466	12.053872	119.3792	85
Sep 2017	70.36549	35.22767	105.50330	16.626837	124.1041	98
Oct 2017	87.56754	52.38011	122.75497	33.753007	141.3821	84
Nov 2017	82.45709	47.22011	117.69407	28.566785	136.3474	75
Dec 2017	89.87369	54.58724	125.16015	35.907717	143.8397	81
Jan 2018	106.74014	71.40440	142.07588	52.698787	160.7815	66
Feb 2018	83.95034	48.56526	119.33542	29.833529	138.0671	81
Mar 2018	67.19598	31.76163	102.63033	13.003816	121.3881	119
Apr 2018	68.44031	32.95676	103.92386	14.172907	122.7077	105
May 2018	73.11653	37.58384	108.64921	18.773978	127.4591	100

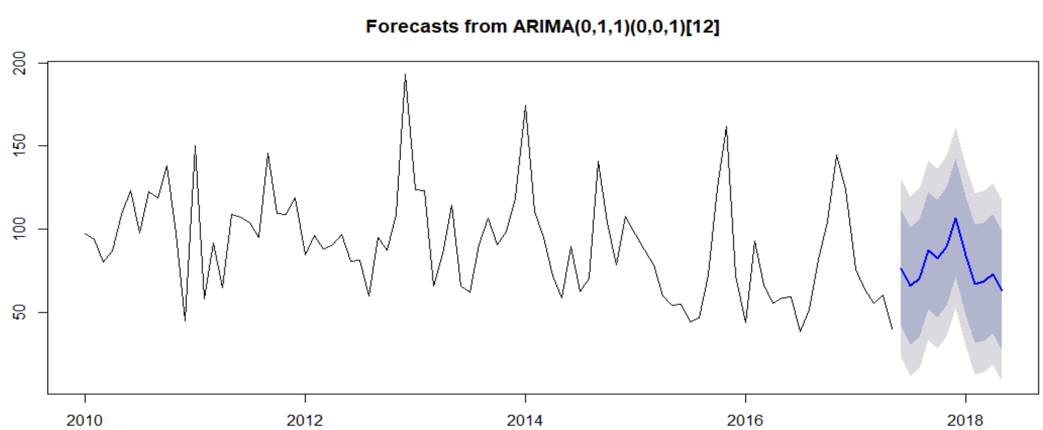


Figure 3.7: Forecasts using ARIMA model

3.4 Seasonal patterns of PM2.5 concentration

To address policies and procedures in Beijing that could potentially be implemented to halt the current stationary PM2.5 level trend, we first look at seasonal patterns. Does **heating in the winter** months influence the PM2.5 measurements? We note that every year winter heating is regulated to begin November 15th and end March 15th. So, we took daily average and daily maximum values of PM2.5 for 2017, for example, and plotted box-plots to show the distribution and difference in levels from the start and end of the months of November and March (See Appendix). There does not seem to be significant differences in PM2.5 values from the implementation of heating systems in the winter months.

Next, we look at the APEC Blue, which is the rare blue sky in Beijing during APEC CHINA 2014 due to a emission reduction campaign directed by the Chinese government. So, are these policies issued by the government effective? There were two stages of strict measures on controlling air pollution: 11/03/2014 – 11/05/2014 and 11/06/2014 – 11/12/2014. So, we first take a look at the period of 11/03 – 11/12 and plot the box-plot distributions for years 2010 – 2014 (below to the left). Second, we take a look at the period 11/06 – 11/12 and plot the box-plot distributions for years 2010 – 2014 (below to the right). As done in previous tests, we then want to see if PM2.5 values increased or decreasing during this APEC Summit. So, we perform two-sample t-tests on the following years: 2014 vs. 2013, 2014 vs. 2012, 2014 vs. 2011, and 2014 vs. 2010. The resulting p-values are as follows, respectively: 2.559e-05, 0.03346, 0.0001492, 6.642e-07. So, all t-tests would reject the null hypothesis and accept that there are significant differences in the mean populations between 2014 and 2013 – 2010. It simply concludes that there are differences in mean populations of PM2.5 for that particular APEC period, which suggests us that measures taken by government are effective.

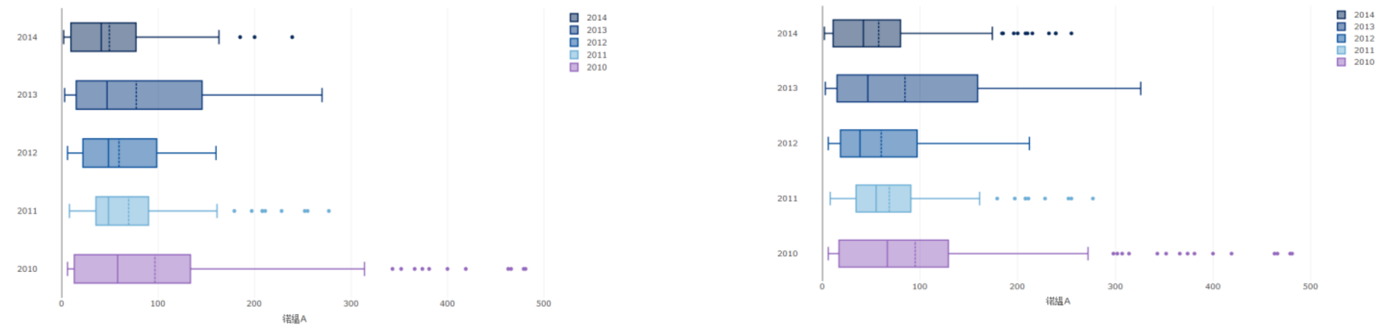


Figure 3.8: Comparison between APEC and other periods

3.5 Diurnal patterns of PM2.5 concentration

Next, we will look at diurnal patterns of PM2.5 concentrations. Taking the hourly data from the US Department of State, we look at three things: a severe days distribution analysis, hourly average in different seasons, and a weekday vs. weekend analysis. To address the severe days distribution analysis, we calculate a cumulative number of severe days based on daily maximum values. See the table below.

Table 5： Severe days calculated by daily max

Severe Days calculated by Daily max					
	Satisfactory	Polluted	Severe	NA	Sum
2010	38	73	208	45	364
2011	39	75	180	71	365
2012	43	67	192	64	366
2013	32	85	213	35	365
2014	40	85	208	32	365
2015	71	105	162	27	365
2016	85	114	143	24	366
2017	51	52	60	18	181

Satisfactory	pm2.5<60
Polluted	60≤pm2.5<120
Severe	pm2.5≥120

This can help us answer one of our objectives: has Beijing hit the goal of bringing PM2.5 concentrations down to around 60 micrograms per cubic meters. As seen in the key, 60 micrograms per cubic meters would fall into the satisfactory category, and even though the amount of days in this category is increasing from 2010 to 2017, it is not by a significant amount and is also not great in relation to the polluted and severe day counts. The pie charts below also show this, when comparing the distribution of satisfactory to polluted/severe from 2010 to 2017. The polluted/severe still take up too much of the population of days in 2017.

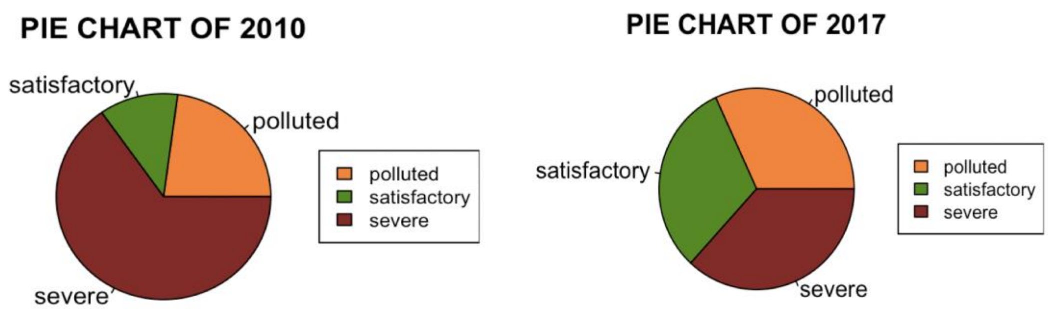
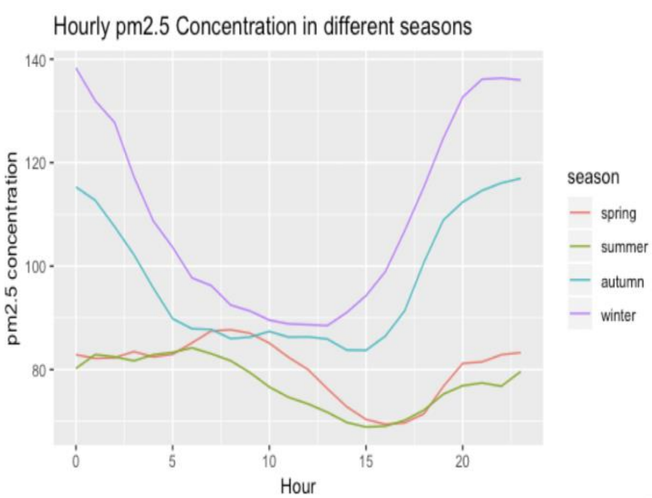
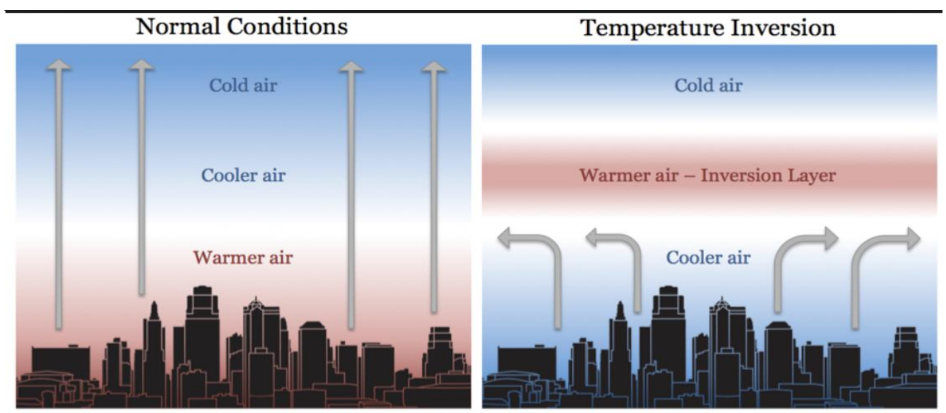


Figure 3.9: Pie charts of distribution of days in 2010 and 2017

To address the hourly average in different seasons, we qualify Spring as March – May, Summer as June – August, Autumn as September – November, and Winter as December – February. Taking these grouped seasons from 2010 – 2017, we calculate the hourly average and plot. See below for this plot. As noticed, the daily pattern seems to present systematic seasonal variations, there may be a **mobile-source influence** (the peak rises at 18:00 – 20:00 PM, thus indicating the evening rush hour peak; in the summer and autumn, the peak rises at 6:00 – 10:00, thus indicating the morning rush hour peak), and PM2.5 measurements still stay high in the evening. The fact that PM2.5 measurements are still high in the evening is due to **temperature inversion**, in which warm air acts as a capped layer, effectively shutting down convection and trapping smog over the city, as seen in the image to the right.





Arrows show air flow in normal conditions on the left and during temperature inversion on the right. In normal conditions, warm air rises and normal convective patterns persist. During temperature inversion, the warm air acts as a cap, effectively shutting down convection and trapping smog over the city.

Figure 3.10: Hourly pm2.5 concentration in different seasons

To address the weekday vs. weekend analysis, we group Saturday and Sunday as the weekend, calculate the hourly average, and do the same for weekday. We simply plotted this, and noticed that overall the hourly PM2.5 concentrations in weekend was higher than that of the weekdays. This is probably due to **the driving restrictions in Beijing**, where 20% of the cars have to stay off of the roads during the weekdays. Since the vehicle possession level amounts to close to 5.35 million units, this means there will be 1.07 million more cars on the road on the weekends rather than the weekdays (Yao, L., Lu, N., Yue, X., Du, J., & Yang, C. 2015).

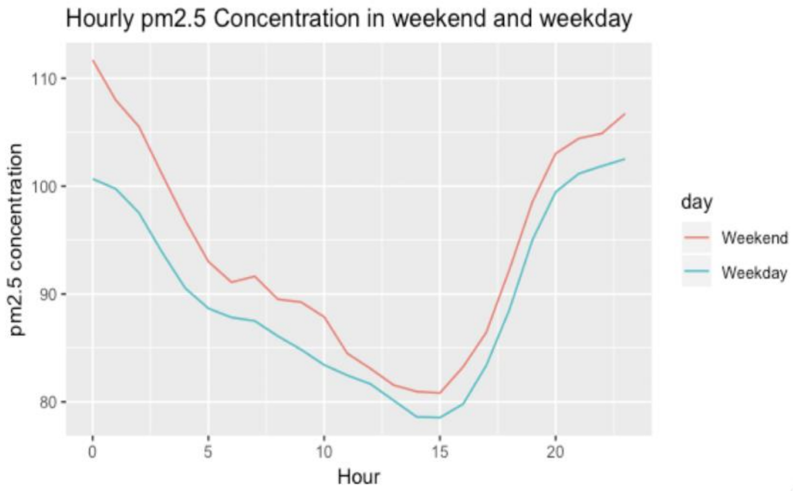


Figure 3.11 : Hourly pm2.5 concentration in weekend and weekday

4. CONCLUSION

There are revolving and complicated factors that contribute to pollution in our cities: vehicles, coal-burning, airborne dust, mobile sources, industrial production, etc. However, the PM2.5 values in Beijing is decreasing due to efforts of people and government.

Based on our results, suggestions are put forward to reduce the air pollution of Beijing. Initially, it is far more efficient to heat into a system that can warm an entire city that to heat buildings individually with boilers, and heating in winter should be timely adjusted according to actual temperature changes.

Secondly, persevere in following the fundamental guidelines and policies during APEC. For example, it's essential to reduce car dependence, ban barbequing and straw burning. In government administrations, there should be harsher supervision and enforcement of penalization.

Lastly, encourage an economic transition toward clean fuels to lessen the need of coal and convert waste products into fuel by applying new technologies.

5. REFERENCES

Yao, L., Lu, N., Yue, X., Du, J., & Yang, C. (2015). Comparison of Hourly PM2.5 Observations Between Urban and Suburban Areas in Beijing, China. *International journal of environmental research and public health*, *12*(10), 12264–12276. doi:10.3390/ijerph121012264

Robert H. Shumway, David S. Stoffer. Time Series Analysis and Its Applications with R Examples (Fourth Edition)

Khaled H.Hamed. (2008) Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology, Volume 349, Issues 3–4, 1 February 2008, Pages 350-363*

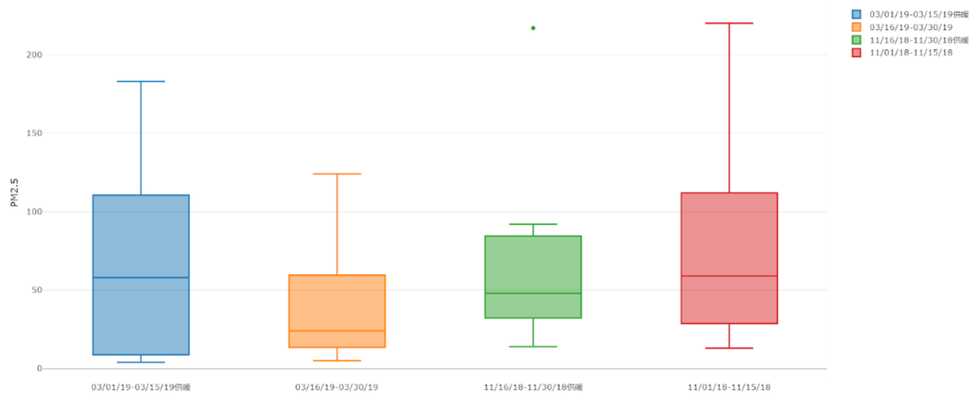
https://vsp.pnnl.gov/help/Vsample/Design_Trend_Mann_Kendall.htm

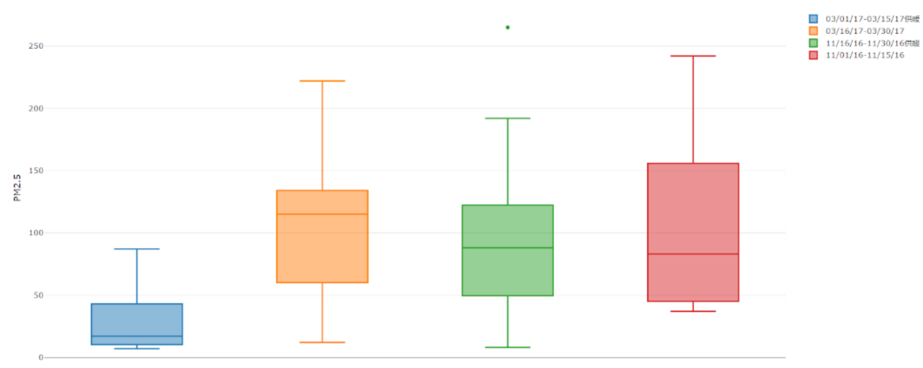
https://plot.ly/r/

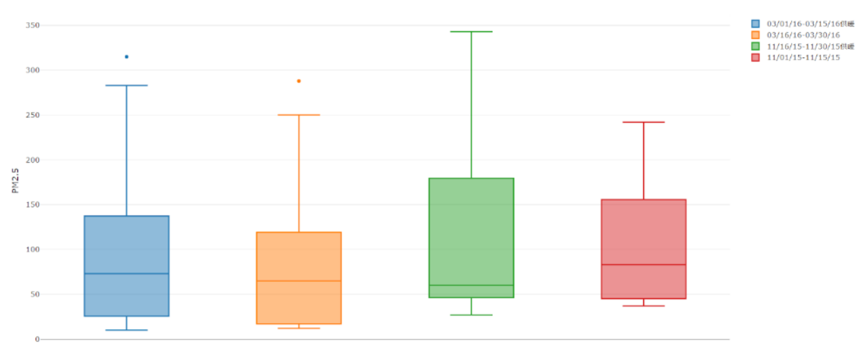
APPENDIX``

1. Other Outputs

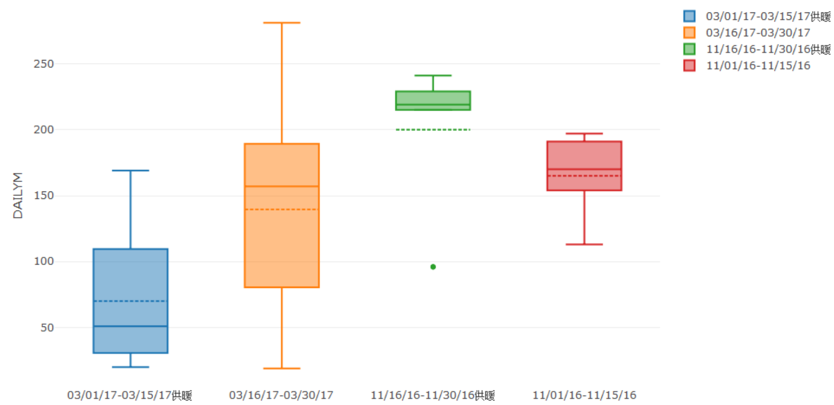
Comparison for daily average pm2.5 values

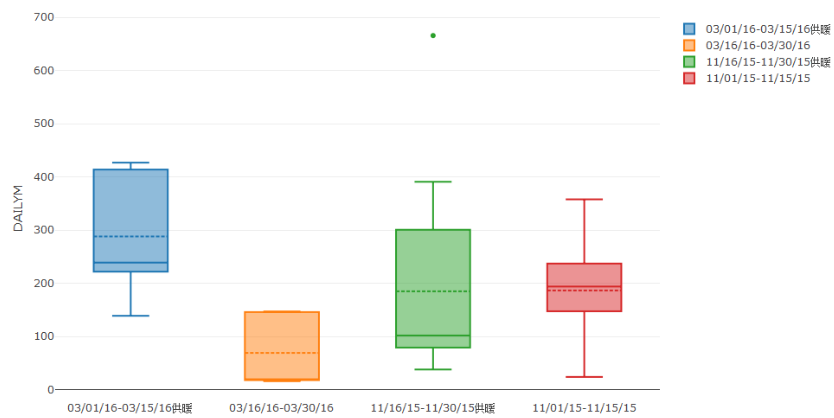


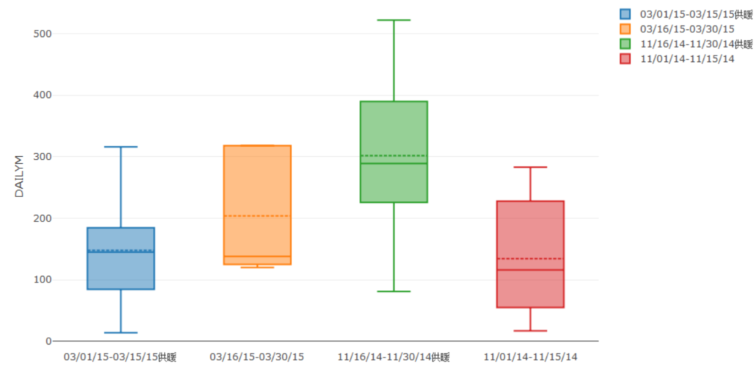




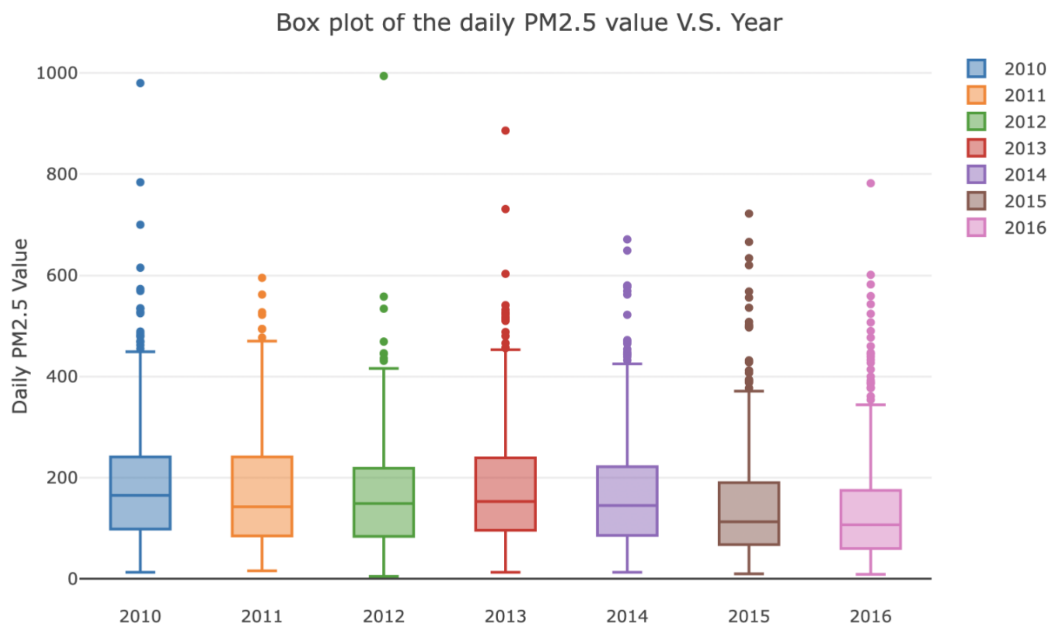
Comparison for daily maximum pm2.5 values



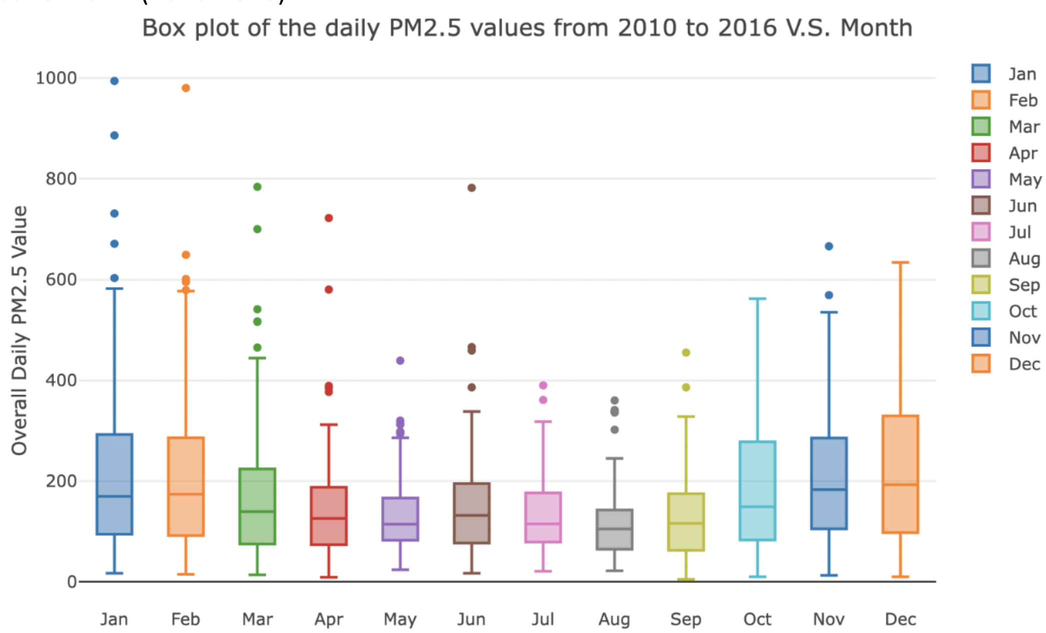




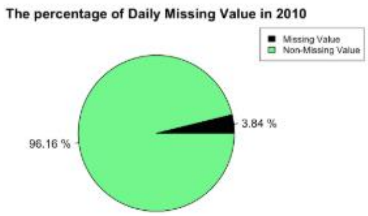
Daily pm2.5 values vs year

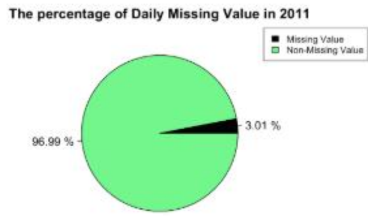


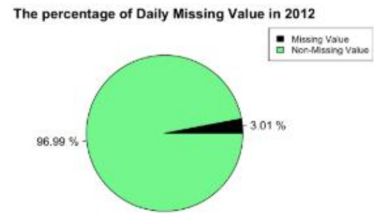
Daily pm2.5 values vs month(2010-2016)

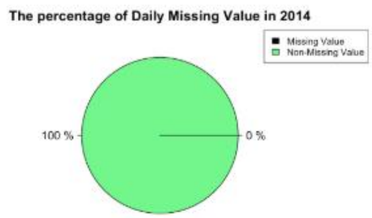


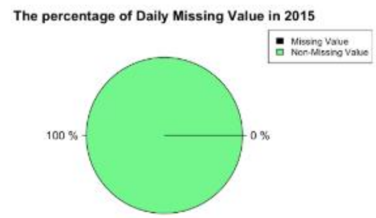
Missing Values in the daily dataset:



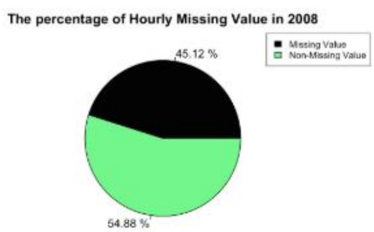


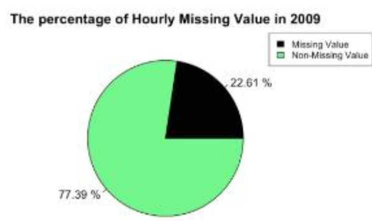


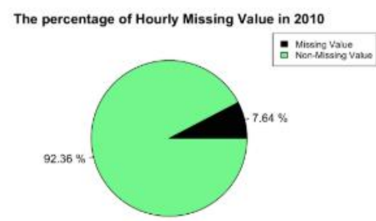


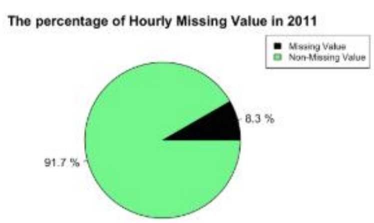


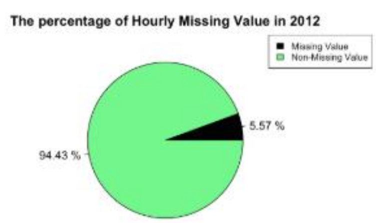
Missing Values in the hourly dataset:

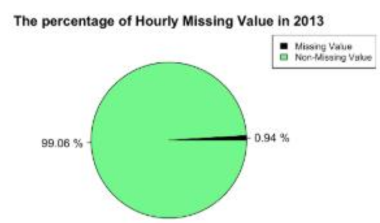












2. R Codes using in the project

```
setwd("C:/Users/derri/Onedrive/r")
library(readxl)
Beijing_2008_HourlyPM2_5<- as.data.frame(read_excel("Beijing_2008_HourlyPM2.5.xls",na="-999"))
Beijing_2009_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2009_HourlyPM2.5.xls",na="-999"))
Beijing_2010_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2010_HourlyPM2.5.xls",na="-999"))
Beijing_2011_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2011_HourlyPM2.5.xls",na="-999"))
Beijing_2012_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2012_HourlyPM2.5.xls",na="-999"))
Beijing_2013_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2013_HourlyPM2.5.xls",na="-999"))
Beijing_2014_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2014_HourlyPM2.5.xls",na="-999"))
Beijing_2015_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2015_HourlyPM2.5.xls",na="-999"))
Beijing_2016_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2016_HourlyPM2.5.xls",na="-999"))
Beijing_2017_HourlyPM2_5 <- as.data.frame(read_excel("Beijing_2017_HourlyPM2.5.xls",na="-999"))
pm2.5<-
rbind(Beijing_2008_HourlyPM2_5,Beijing_2009_HourlyPM2_5,Beijing_2010_HourlyPM2_5,Beijing_2011_HourlyPM2_5,
Beijing_2012_HourlyPM2_5,Beijing_2013_HourlyPM2_5,Beijing_2014_HourlyPM2_5,Beijing_2015_HourlyPM2_5,Beijin
g_2016_HourlyPM2_5,Beijing_2017_HourlyPM2_5)
attach(pm2.5)
```

```
names(pm2.5)
pm2.5$Date<-as.Date(pm2.5$Date, format=c('y-m-d','h:m:s'))
plot(pm2.5$Value ~ pm2.5$Date,axt = "n", type = "l")
data <- read.table("monthly2.txt",header=TRUE)
pm2.5ts <- ts(data,start=c(2010,1))
plot(pm2.5ts)
library(forecast)
library(astsa)
library(tseries)
adf.test(pm2.5ts)
install.packages("aTSA")
library(aTSA)
acf2(pm2.5ts)
library(zoo)
library(xts)
library(trend)
pm2.5timeseries <- ts(data,frequency=12,start=c(2010,1))
plot(pm2.5timeseries)
pm2.5decom <- decompose(pm2.5timeseries)
plot(pm2.5decom)
pm2.5noseason <- pm2.5timeseries-pm2.5decom$seasonal
plot(pm2.5noseason)
mk.test(pm2.5timeseries)
mk.test(pm2.5noseason)
adf.test(pm2.5noseason)
pp.test(pm2.5noseason)
stationary.test(pm2.5noseason, method = "pp") # same as pp.test(x)
stationary.test(pm2.5noseason, method = "kpss") # same as kpss.test(x)
fit1 <- auto.arima(pm2.5timeseries,seasonal = TRUE, approximation = FALSE, stepwise = FALSE)
fit2 <- forecast(fit1,h=12)
plot(fit2)
plot_ly(data=new,x=~date)%>%
  add_lines(y=~pm2.5values,name = "Measured by US Embassy")%>%
        add_lines(y=~pm2.5c, name = "Measured by China")

fit1 <- auto.arima(pm2.5values,seasonal = TRUE, approximation = FALSE, stepwise = FALSE)
fit1
fit2 <- forecast(fit1,h=12)
plot(fit2)
newdata <- read.table("monthly3.txt",header=TRUE)
pm2.5ts <- ts(newdata,frequency=12,start=c(2010,1))
pm2.5decom1 <- decompose(pm2.5ts)
plot(pm2.5decom1)
pm2.5noseason1 <- pm2.5ts-pm2.5decom1$seasonal
plot(pm2.5noseason)
plot(pm2.5decom1$trend)
adf.test(pm2.5noseason1)
pp.test(pm2.5noseason1)
attach(newdata)
fit3 <- auto.arima(pm2.5values,seasonal = TRUE, approximation = FALSE, stepwise = FALSE)
fit3
```

```
fit4 <- forecast(fit4,h=12)
plot(fit4)
data <- read.csv("NEW13-17.csv", header = TRUE)
library(ggplot2)
library(plotly)

data1 <- data[c(90:94),] #2014/03/01~2014/03/15
data1 <- data.frame(data1)
data2 <- data[c(95:99),] #2014/03/16~2014/03/30
data2 <- data.frame(data2)
pm2.5b <- data2$DAILYM

data3 <- data[c(:364),] #2013/11/16-2013/11/30
data3 <- data.frame(data3)
pm2.5c <- data3$DAILYM
data4 <- data[c(335:349),] #2013/11/01-2013/11/15
data4 <- data.frame(data4)
pm2.5d <- data4$DAILYM

pm2.5 <- cbind(data1,pm2.5b,pm2.5c,pm2.5d)

plot_ly(data=pm2.5) %>%
  add_boxplot(y=~DAILYM, name = "03/01/15-03/15/15 供暖", boxmean = TRUE) %>%
  add_boxplot(y=~pm2.5b, name = "03/16/15-03/30/15", boxmean = TRUE)%>%
  add_boxplot(y=~pm2.5c, name = "11/16/14-11/30/14 供暖", boxmean = TRUE) %>%
  add_boxplot(y=~pm2.5d, name = "11/01/14-11/15/14", boxmean = TRUE)

setwd("C:/Users/derri/Onedrive/r")

data <- read.table("monthly2.txt",header=TRUE)
x <- ts(data,frequency=12,start=c(2010,1))
newdata <- read.table("new.txt",header = TRUE)
y <- ts(newdata,frequency=12,start=c(2013,12))

plot(x)
plot(y)

require(graphics)
ts.plot(x, y,
      gpars=list(xlab="date", ylab="pm2.5values", lty=c(1:3)))

library(ggplot2)
attach(data)
pm2.5old <- pm2.5values[47,90]
pm2.5new <- newdata$pm2.5c[1,43]

time1=seq(from=as.Date("2010-01-01"),to = as.Date("2017-06-01"),by='month')
data1=data.frame(date=time1,pm2.51=data)

time2=seq(from=as.Date("2013-12-01"),to = as.Date("2019-04-01"),by='month')
data2=data.frame(date=time2,pm2.52=newdata)
```

```
new=merge(data1,data2,by='date',all=TRUE)
library(plotly)

plot_ly(data=new,x=~date)%>%
  add_lines(y=~pm2.5values,
  add_lines(y=~pm2.5c)

pm2.5 <- read.csv("AQIDATA.csv", header = TRUE)
data1 <- pm2.5[c(1915:1929),] #2019/03/01~2019/03/15
data1 <- data.frame(data1)
data2 <- pm2.5[c(1930:1944),] #2019/03/16~2019/03/30
data2 <- data.frame(data2)
pm2.5b <- data2$PM2.5

data3 <- pm2.5[c(1810:1824),] #2018/11/16-2018/11/30
data3 <- data.frame(data3)
pm2.5c <- data3$PM2.5

data4 <- pm2.5[c(1795:1809),] #2018/11/01-2018/11/15
data4 <- data.frame(data4)
pm2.5d <- data4$PM2.5

data <- cbind(data1,pm2.5b,pm2.5c,pm2.5d)

plot_ly(data=data) %>%
  add_boxplot(y=~PM2.5, name = "2019/03/01-2019/03/15 供暖") %>%
  add_boxplot(y=~pm2.5b, name = "2019/03/16-2019/03/30 不供暖")%>%
  add_boxplot(y=~pm2.5c, name = "2018/11/16-2018/11/30 供暖") %>%
  add_boxplot(y=~pm2.5d, name = "2018/11/01-2018/11/15 不供暖")


data5 <- pm2.5[c(1550:1564),] #2018/03/01~2018/03/15
data5 <- data.frame(data5)
data6 <- pm2.5[c(1565:1579),] #2018/03/16~2018/03/30
data6 <- data.frame(data6)
pm2.5e <- data6$PM2.5
data7 <- pm2.5[c(1445:1459),] #2017/11/16-2017/11/30
data7 <- data.frame(data7)
pm2.5f <- data7$PM2.5
data8 <- pm2.5[c(1430:1444),] #2017/11/01-2017/11/15
data8 <- data.frame(data8)
pm2.5g <- data8$PM2.5
datan <- cbind(data5,pm2.5e,pm2.5f,pm2.5g)
plot_ly(data=datan) %>%
  add_boxplot(y=~PM2.5, name = "03/01/18-03/15/18 供暖") %>%
  add_boxplot(y=~pm2.5e, name = "03/16/18-03/30/18")%>%
  add_boxplot(y=~pm2.5f, name = "11/16/17-11/30/17 供暖") %>%
  add_boxplot(y=~pm2.5g, name = "11/01/17-11/15/17")

data9 <- pm2.5[c(1185:1199),] #2017/03/01~2017/03/15
```

```
data9 <- data.frame(data9)
data10 <- pm2.5[c(1200:1214),] #2017/03/16~2017/03/30
data10 <- data.frame(data10)
pm2.5h <- data10$PM2.5
data11 <- pm2.5[c(1080:1094),] #2016/11/16-2016/11/30
data11 <- data.frame(data11)
pm2.5i <- data11$PM2.5
data12 <- pm2.5[c(1065:1079),] #2016/11/01-2016/11/15
data12 <- data.frame(data12)
pm2.5j <- data12$PM2.5
datam <- cbind(data9,pm2.5h,pm2.5i,pm2.5j)
plot_ly(data=datam) %>%
  add_boxplot(y=~PM2.5, name = "03/01/17-03/15/17 供暖") %>%
  add_boxplot(y=~pm2.5h, name = "03/16/17-03/30/17")%>%
  add_boxplot(y=~pm2.5i, name = "11/16/16-11/30/16 供暖") %>%
  add_boxplot(y=~pm2.5j, name = "11/01/16-11/15/16")

data13 <- pm2.5[c(821:835),] #2016/03/01~2016/03/15
data13 <- data.frame(data13)
data14 <- pm2.5[c(836:850),] #2016/03/16~2016/03/30
data14 <- data.frame(data14)
pm2.5k <- data14$PM2.5
data15 <- pm2.5[c(715:729),] #2015/11/16-2015/11/30
data15 <- data.frame(data15)
pm2.5l <- data15$PM2.5
data16 <- pm2.5[c(700:714),] #2015/11/01-2015/11/15
data16 <- data.frame(data16)
pm2.5m <- data12$PM2.5
datap <- cbind(data13,pm2.5k,pm2.5l,pm2.5m)
plot_ly(data=datap) %>%
  add_boxplot(y=~PM2.5, name = "03/01/16-03/15/16 供暖") %>%
  add_boxplot(y=~pm2.5k, name = "03/16/16-03/30/16")%>%
  add_boxplot(y=~pm2.5l, name = "11/16/15-11/30/15 供暖") %>%
  add_boxplot(y=~pm2.5m, name = "11/01/15-11/15/15")

apec <- read.csv("apec.csv", header = TRUE)
plot_ly(data=apec) %>%
  add_boxplot(x=~锘組 A, name = "2014", boxmean = TRUE, marker = list(color = 'rgb(7,40,89)'),
        line = list(color = 'rgb(7,40,89)')) %>%
  add_boxplot(x=~BB, name = "2013", boxmean = TRUE,  marker = list(color = 'rgb(9,56,125)'),
        line = list(color = 'rgb(9,56,125)')) %>%
  add_boxplot(x=~CC, name = "2012", boxmean = TRUE,  marker = list(color = 'rgb(8,81,156)'),
        line = list(color = 'rgb(8,81,156)')) %>%
  add_boxplot(x=~DD, name = "2011", boxmean = TRUE,  marker = list(color = 'rgb(107,174,214)'),
        line = list(color = 'rgb(107,174,214)')) %>%
  add_boxplot(x=~EE, name = "2010", boxmean = TRUE)

t.test(apec$锘組 A, apec$BB)
t.test(apec$锘組 A, apec$CC)
t.test(apec$锘組 A, apec$DD)
t.test(apec$锘組 A, apec$EE)
```

```
apec3 <- read.csv("apec3.csv", header = TRUE)
plot_ly(data=apec3) %>%
  add_boxplot(x=~锘緼 A, name = "2014", boxmean = TRUE, marker = list(color = 'rgb(7,40,89)'),
            line = list(color = 'rgb(7,40,89)')) %>%
  add_boxplot(x=~BB, name = "2013", boxmean = TRUE,  marker = list(color = 'rgb(9,56,125)'),
            line = list(color = 'rgb(9,56,125)')) %>%
  add_boxplot(x=~CC, name = "2012", boxmean = TRUE,  marker = list(color = 'rgb(8,81,156)'),
            line = list(color = 'rgb(8,81,156)')) %>%
  add_boxplot(x=~DD, name = "2011", boxmean = TRUE,  marker = list(color = 'rgb(107,174,214)'),
            line = list(color = 'rgb(107,174,214)')) %>%
  add_boxplot(x=~EE, name = "2010", boxmean = TRUE)

t.test(apec3$锘緼 A, apec3$BB)
t.test(apec3$锘緼 A, apec3$CC)
t.test(apec3$锘緼 A, apec3$DD)
t.test(apec3$锘緼 A, apec3$EE)
```