PM2.5 SHANGHAI

ZHU WANG (王祝)
WEI ZHANG (张薇)

PM2.5 SHANGHAI

ZHU WANG (王祝)
WEI ZHANG (张薇)

WHY PM 2.5 MATTERS?

An Introduction

SHANGHAI

- China's east coast
- Located in the Yangtze River Delta
- A global financial center and transport hub

AIR POLLUTION AND GOVERNMENT REACTION

- Not as severe as in many other Chinese cities, but still substantial by world standards

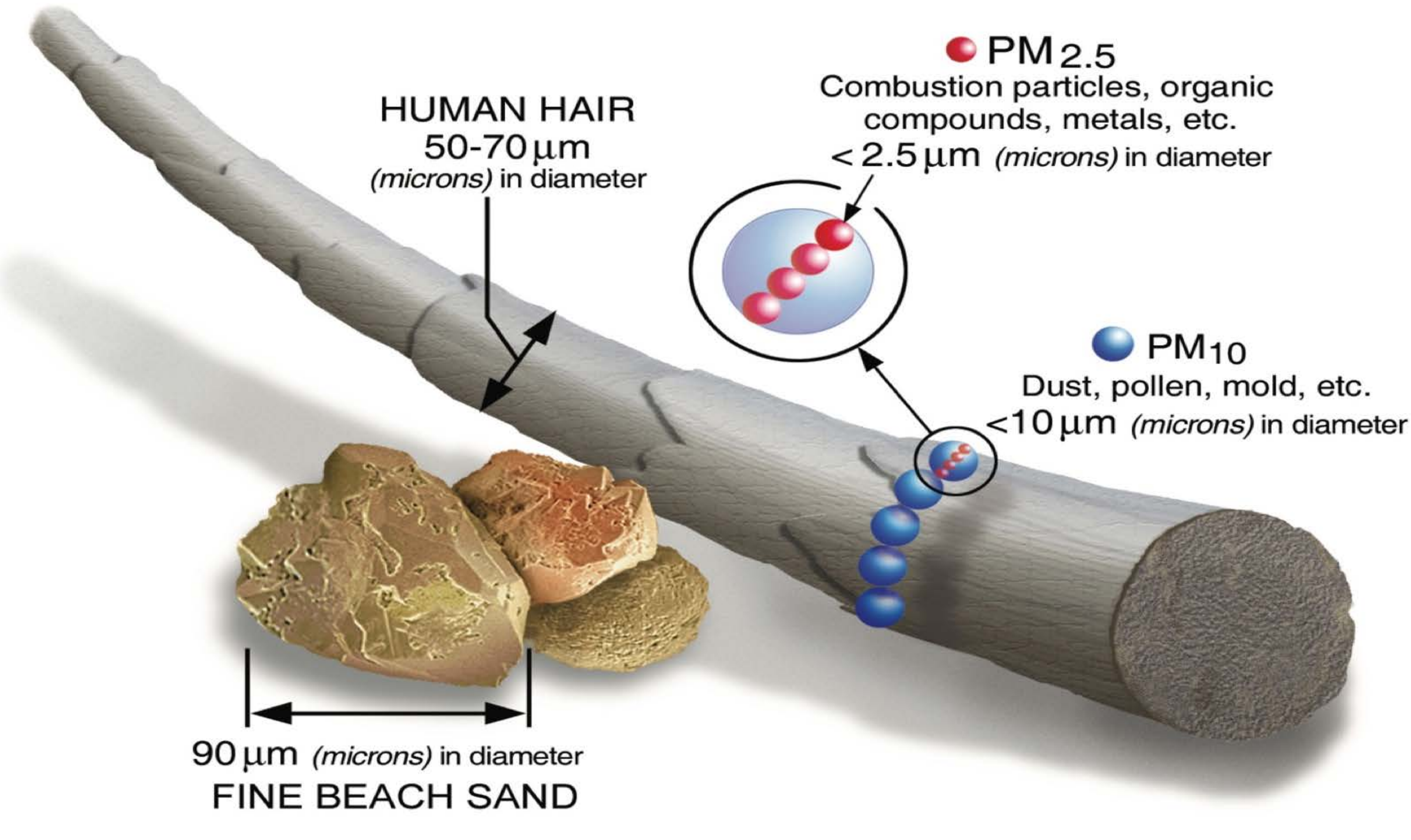
- On 23 January 2014, the mayor of Shanghai municipality announced that three main measures would be taken to manage the air pollution in Shanghai, along with surrounding Anhui, Jiangsu and Zhejiang provinces.

- The measures involved delivery of the 2013 air cleaning program, linkage mechanism with the three surrounding provinces and improvement of the ability of early warning of emergency situation.

WHAT IS PM

- particulate matter (also called particle pollution)
- a mixture of solid particles and liquid droplets found in the air
- Particle pollution includes:
  - PM10 : inhalable particles, with diameters that are generally 10 micrometers and smaller;
  - PM2.5 : fine inhalable particles, with diameters that are generally 2.5 micrometers and smaller.

WHAT IS PM



PM2.5
Combustion particles, organic compounds, metals, etc.
< 2.5 µm (microns) in diameter
HUMAN HAIR
50-70 µm
(microns) in diameter
PM10
Dust, pollen, mold, etc.
<10 µm (microns) in diameter
90 µm (microns) in diameter
FINE BEACH SAND

SOURCE OF PM

- These particles come in many sizes and shapes and can be made up of hundreds of different chemicals.
- From natural:
  - volcanoes, dust storms, forest and grassland fires, living vegetation and sea spray
- From human activities:
  - burning of fossil fuels in vehicles, stubble burning, power plants, and various industrial processes

HARMFUL EFFECTS OF PM

- Health Effect
- Small particles less than 10 micrometers in diameter can get deep into your lungs, and some may even get into your bloodstream.
- Exposure to such particles can affect both your lungs and your heart.
- premature death in people with heart or lung disease
- nonfatal heart attacks
- irregular heartbeat
- aggravated asthma
- decreased lung function

HARMFUL EFFECTS OF PM

- Environmental Effect
- Particles can be carried over long distances by wind and then settle on ground or water.
- making lakes and streams acidic
- changing the nutrient balance in coastal waters and large river basins
- damaging sensitive forests and farm crops
- Affecting the diversity of ecosystems
- contributing to acid rain effects.

STANDARD

- In 2013, the ESCAPE study involving 312,944 people in nine European countries revealed that there was no safe level of particulates and that for every increase of 10 $\mu g/m$^3 in PM10, the lung cancer rate rose 22%. For PM2.5 there was a 36% increase in lung cancer per 10 $\mu g/m$^3.
- In a 2014 meta-analysis of 18 studies globally including the ESCAPE data, for every increase of 10 $\mu g/m$^3 in PM2.5, the lung cancer rate rose 9%.

AVERAGE PM 2.5 FOR 24 HOURS

0~35µg/m³	Good
35~75µg/m³	Moderate
75~115µg/m³	Unhealthy for sensitive groups
115~150µg/m³	Unhealthy
150~250µg/m³	Very Unhealthy
>250µg/m³	Hazardous

OUTLINE OF OUR PROJECT

- **Background Introduction**
- **Questions Addressed**
  - What are the relevant factors for PM2.5? How are they effecting PM2.5?
  - Is PM2.5 getting better or worse in the past several years?
  - Suggestions?
- **Data Source**
  - Where are the data from
  - Data Types

OUTLINE OF OUR PROJECT

- **Data Exploration Analysis**
    - Missing data
    - Trends in recent years
    - More exploration
- **Statistical Analysis**
    - Trends analysis
    - Fitting Models
    - Forecast
- **Discussion and Suggestion**

DATA SOURCE

WHERE WE GET THE DATA

- **Pollutions data:**
  - The US Embassy
  - Shanghai Environmental Monitoring Center
  - Shanghai Municipal Bureau of Ecology and Environment

- **Others**:
  - National Geomatics Center of China
  - Tutiempo.net

DATA TYPES

- 2012-2017 Hourly data from US Embassy

Site	Parameter	Date..LST.	Year	Month	Day	Hour	Value	Unit
Shanghai	PM2.5	1/1/2012 0:00	2012	1	1	0	112	_/m
Shanghai	PM2.5	1/1/2012 1:00	2012	1	1	1	113	_/m
Shanghai	PM2.5	1/1/2012 2:00	2012	1	1	2	115	_/m
Shanghai	PM2.5	1/1/2012 3:00	2012	1	1	3	144	_/m
Shanghai	PM2.5	1/1/2012 4:00	2012	1	1	4	152	_/m
Shanghai	PM2.5	1/1/2012 5:00	2012	1	1	5	138	_/m

DATA TYPES

- 2016-2019 Daily data from Shanghai Environmental Monitoring Center

id	date	pm2.5	pm10	o3	so2	no2	co	aqi	quality
1118	2016-01-01	79	58	41	17	75	23	79	良
1117	2016-01-02	133	87	32	26	104	35	133	轻度污染
1116	2016-01-03	207	114	17	28	124	45	207	重度污染
1115	2016-01-04	165	83	45	27	102	35	165	中度污染
1114	2016-01-05	36	NA	47	13	35	15	47	优
1113	2016-01-06	64	55	37	18	44	18	64	良

DATA TYPES

- 2018 Hourly Data of different blocks from Shanghai Municipal Bureau of Ecology and Environment

time	site	aqi	level	pm25	pm10	co	no2	ozone1hour	ozone8hour
12/1/2018 1:00	Shiwuchang	67	2	35	83	0.5	58	52	52
12/1/2018 1:00	Hongkou	60	2	26	70	0.5	43	64	64
12/1/2018 1:00	Xuhui	59	2	29	68	0.4	55	48	48
12/1/2018 1:00	Yangpu	68	2	34	86	0.5	62	38	38
12/1/2018 1:00	Qingpu	77	2	33	103	0.4	50	34	34
12/1/2018 1:00	Jingan	63	2	22	75	0.3	45	76	76

DATA TYPES

- 2016-2019 Shanghai weather data

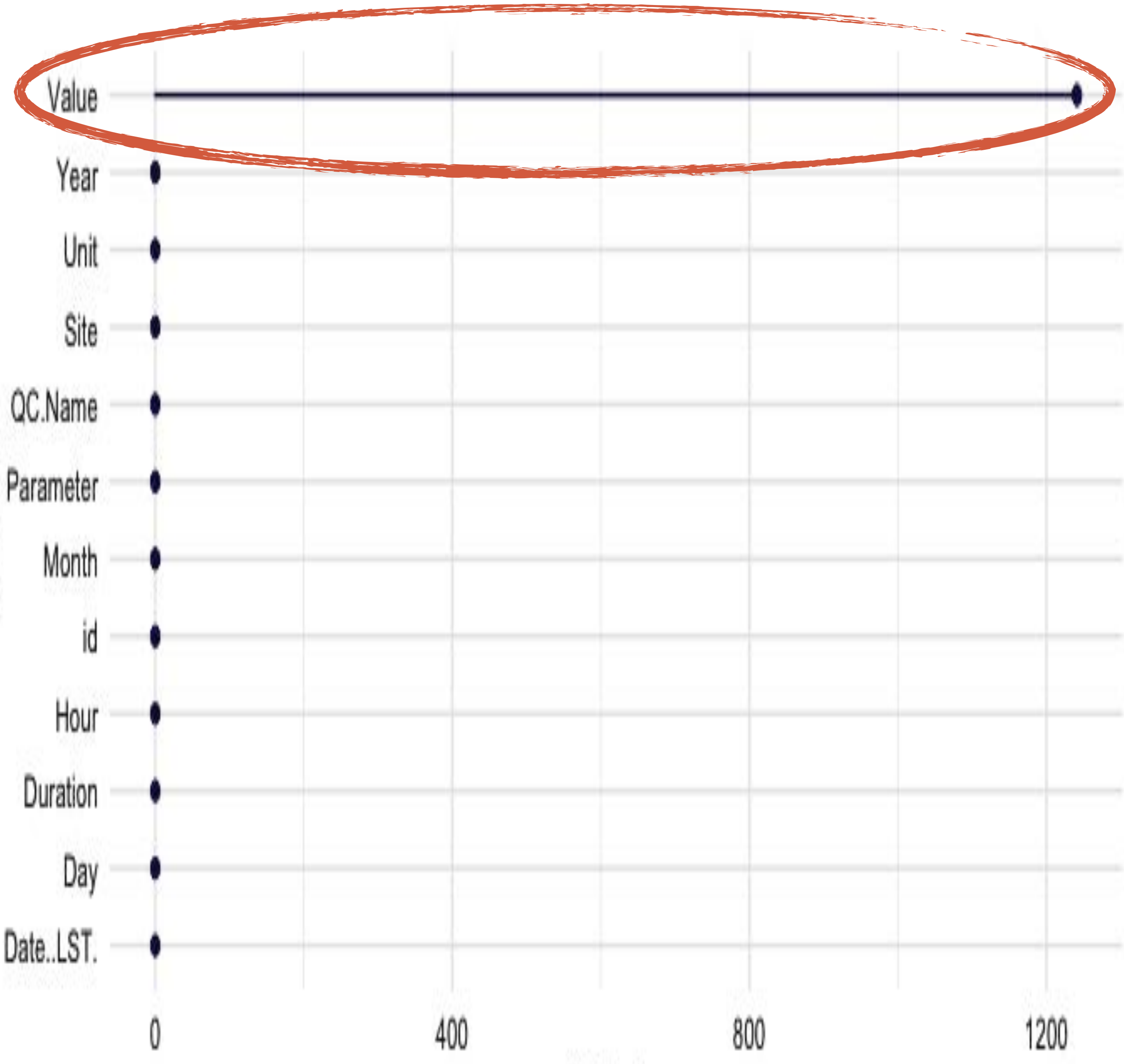
Y	M	D	T	TM	Tm	SLP	STP	H	PP	VV	V	VM	VG	FG
2016	1	1	9.1	12.5	-0.9	1029	1028.1	67	0	7.6	6.7	10.7	-	0
2016	1	2	11.1	16.9	7.3	1022.6	1021.8	62	0	5	5.4	10.7	-	0
2016	1	3	12.1	15.7	7.3	1021.3	1020.5	78	1.78	3.1	3.1	3.5	-	0
2016	1	4	11.3	13.5	9	1021.4	1020.6	89	9.14	2.7	11.7	18	-	0
2016	1	5	8.6	9.2	7.7	1025.2	1024.4	81	1.27	14.8	13	18	-	0
2016	1	6	7.4	9.2	4	1028.4	1027.5	74	0	8.5	9.1	18	-	0

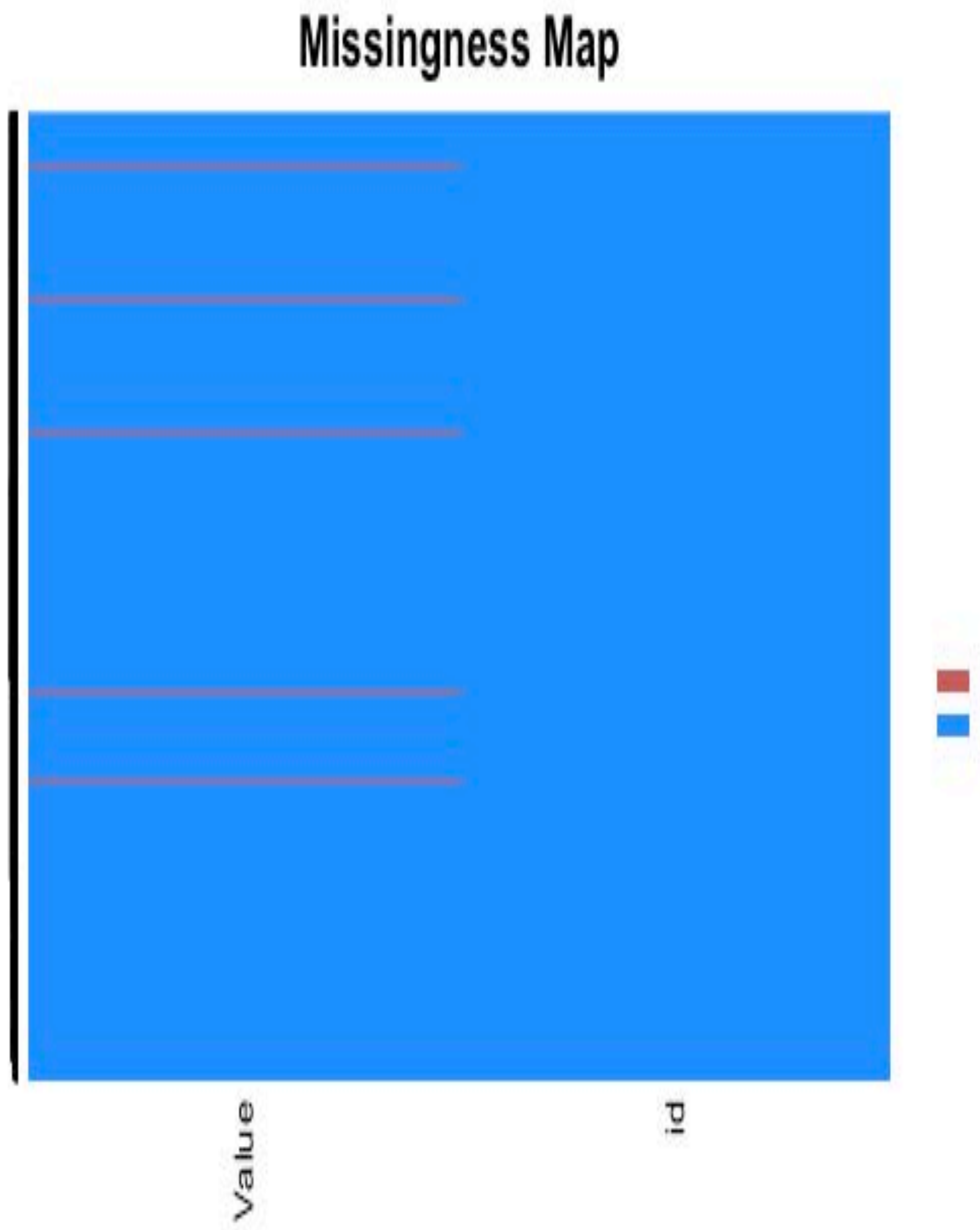
- *H: Humidity*
- *T: Temperature*
- *PP: Rainfall*
- *V: Wind Speed*
- *VV: Visibility*

EXPLORING DATA

MISSING DATA

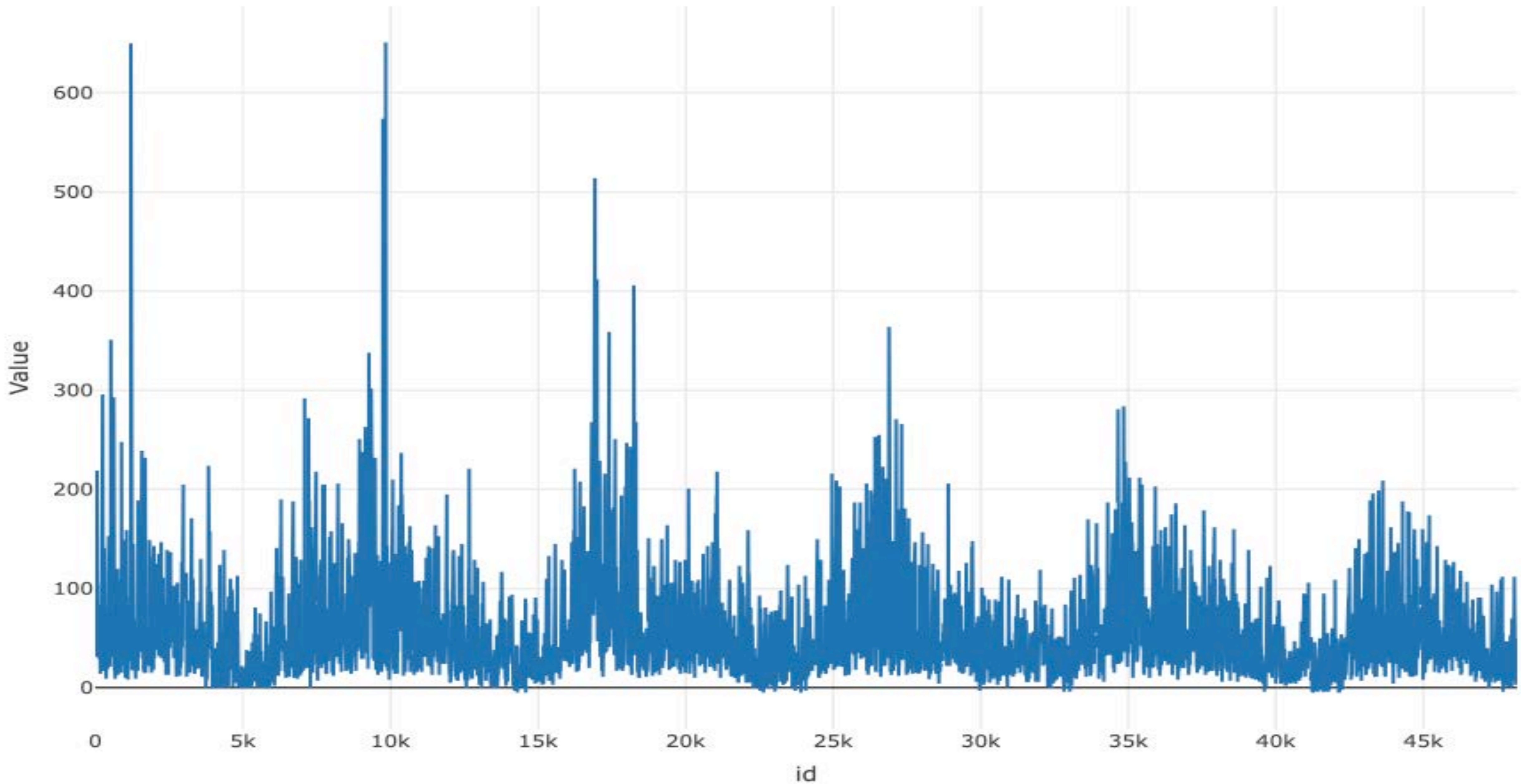
- 2012-2017 Hourly data from US Embassy





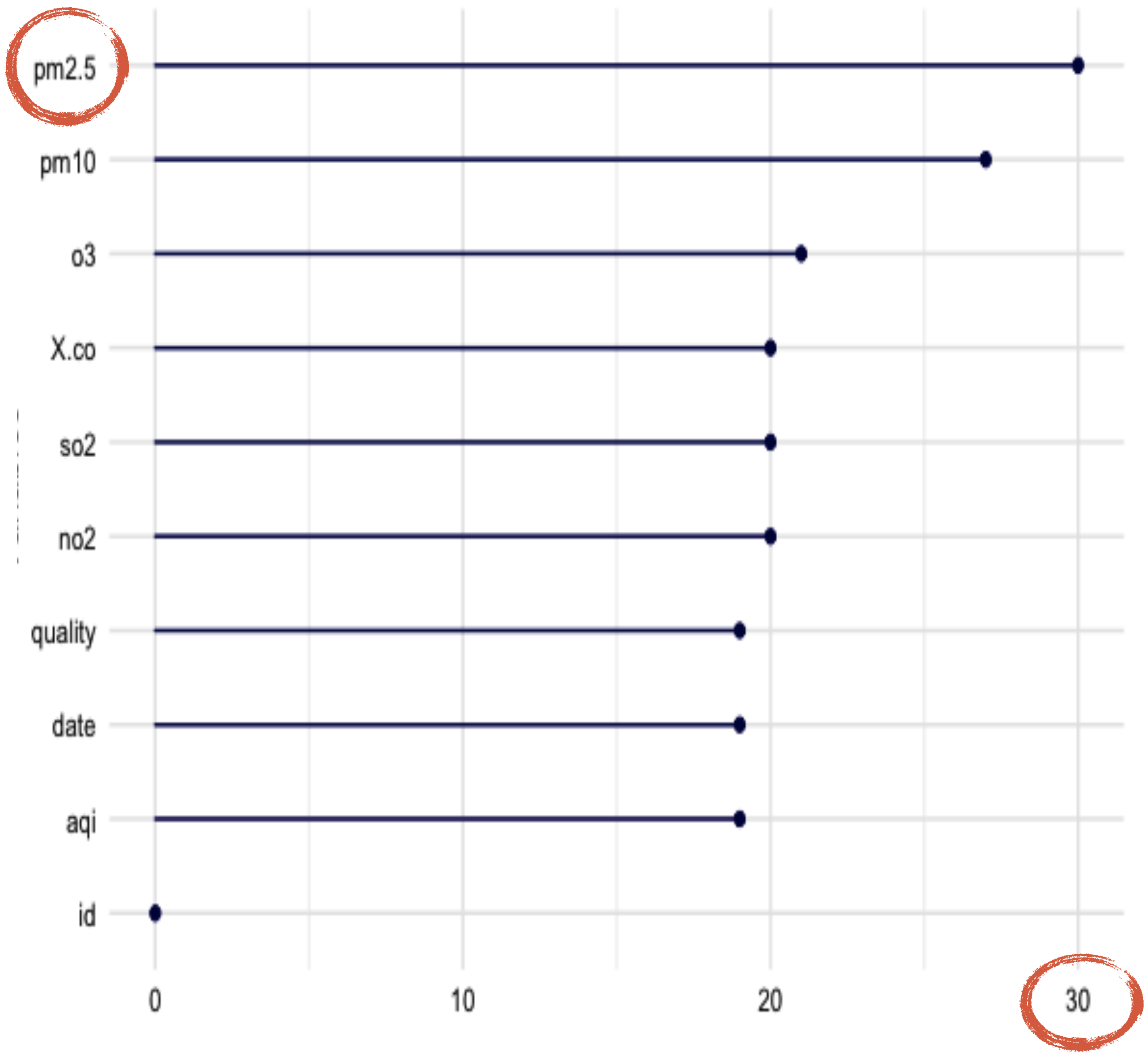
MISSING DATA

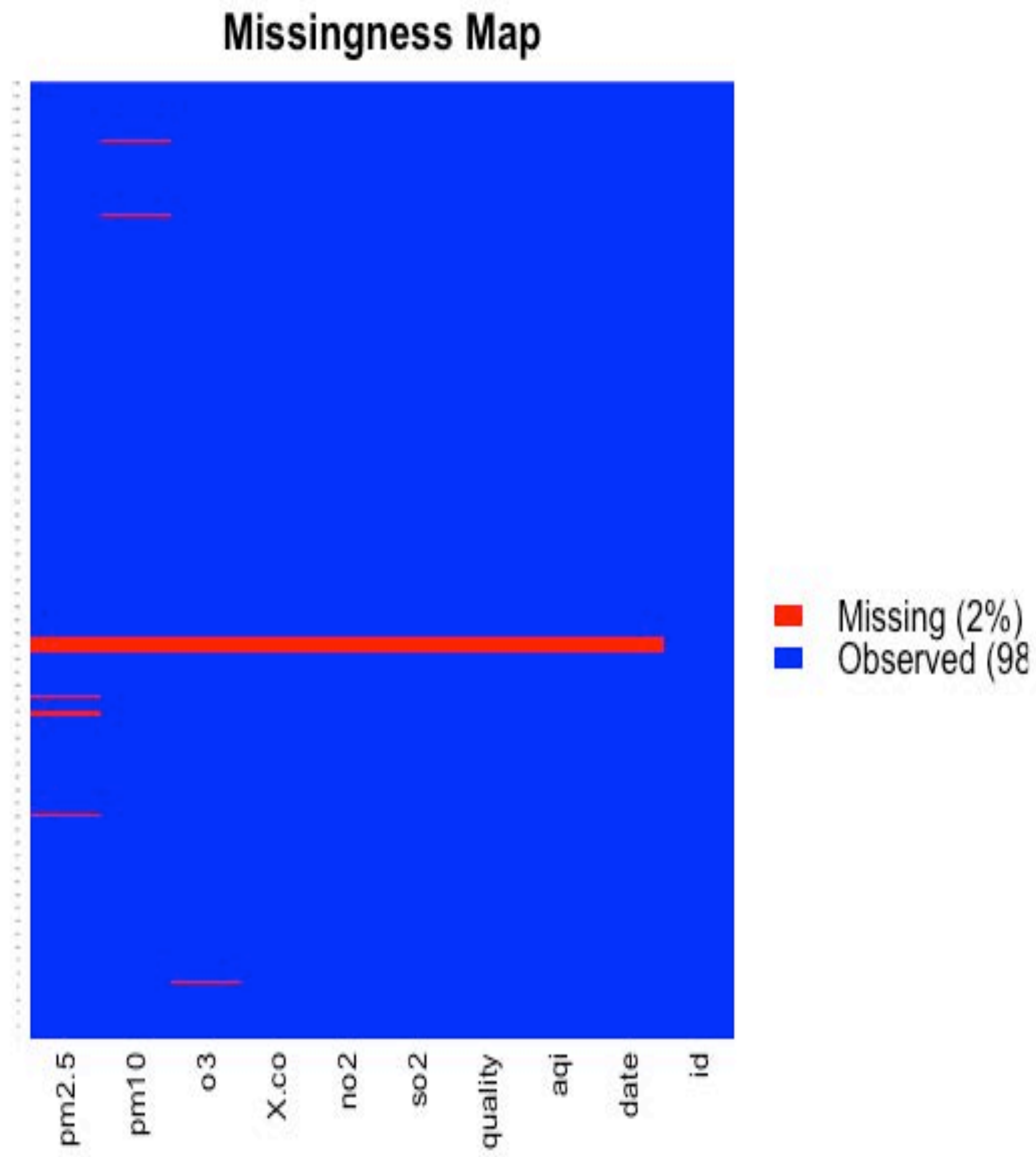
- 2012-2017 Hourly data from US Embassy



MISSING DATA

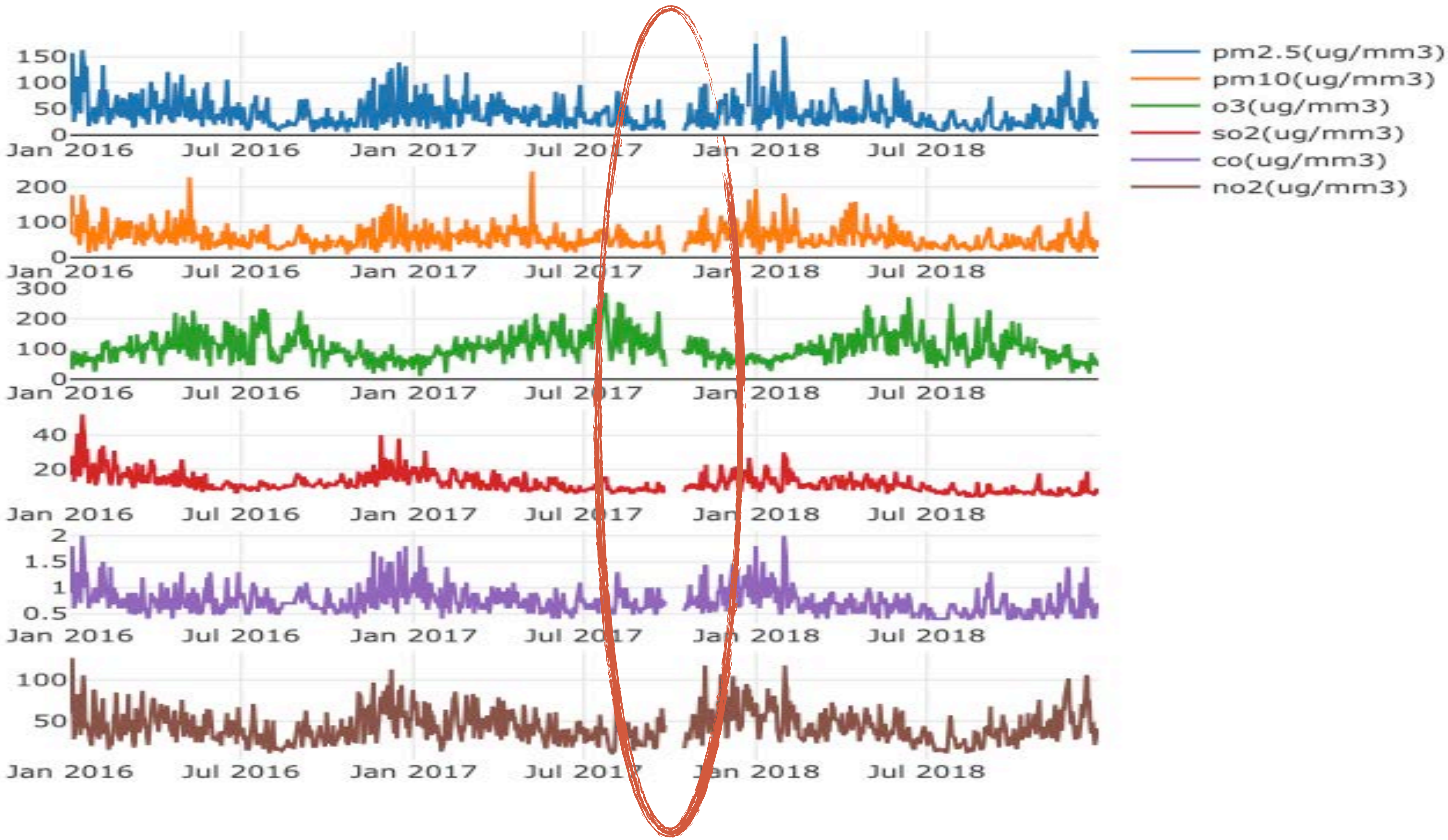
- 2016-2019 Daily data from Shanghai environmental monitoring center





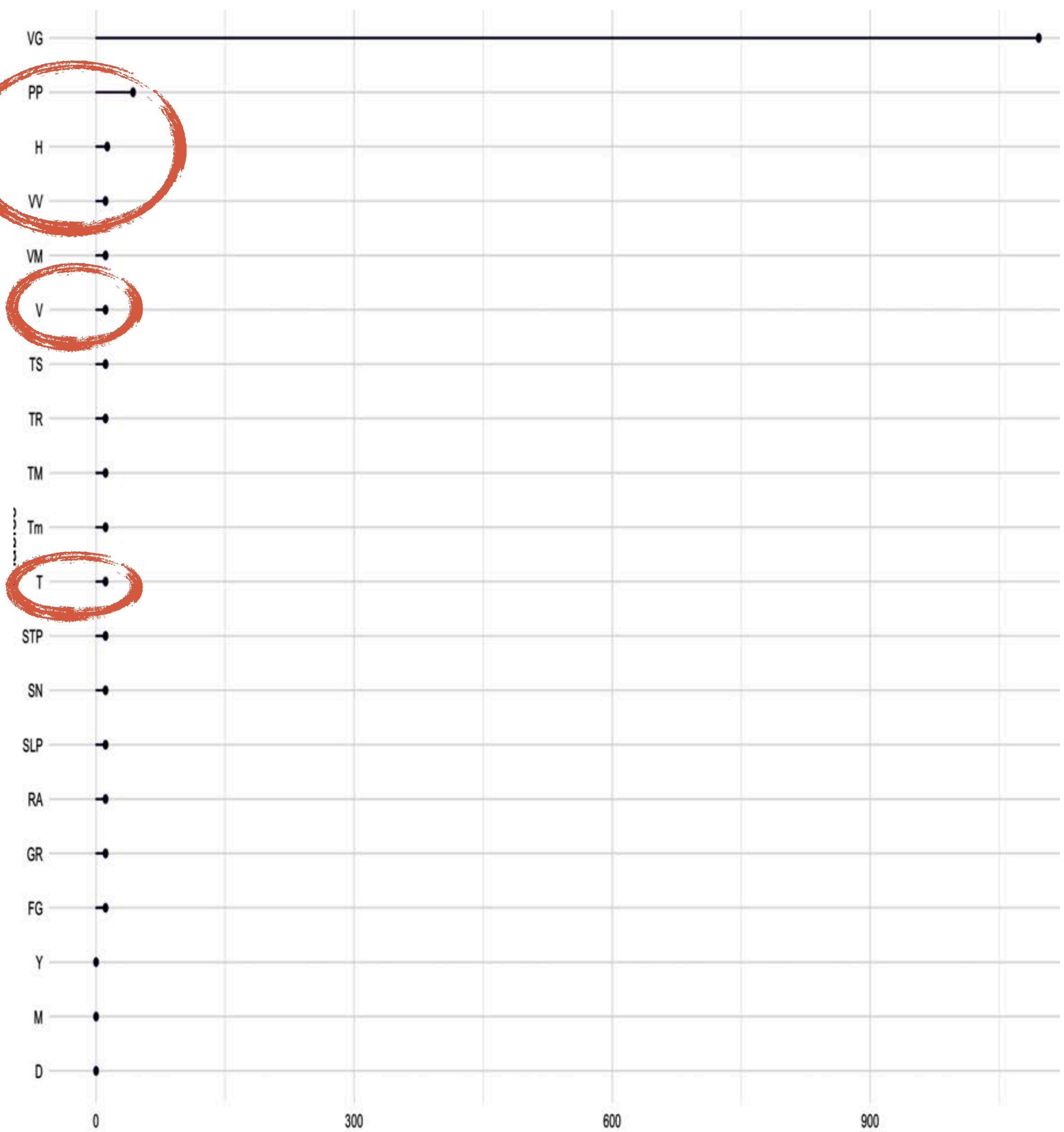
MISSING DATA

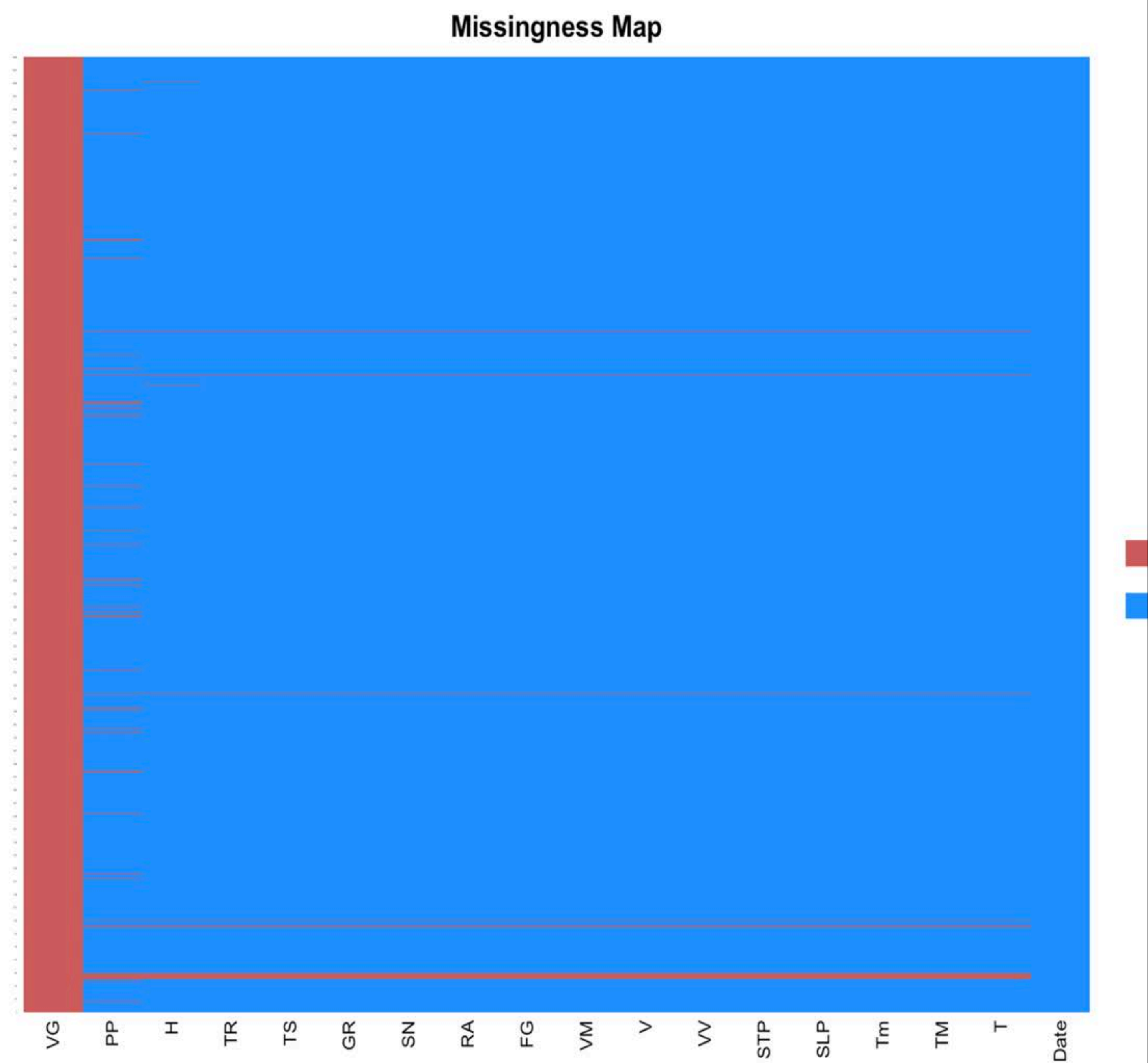
- 2016-2019 Daily data from Shanghai environmental monitoring center



MISSING DATA

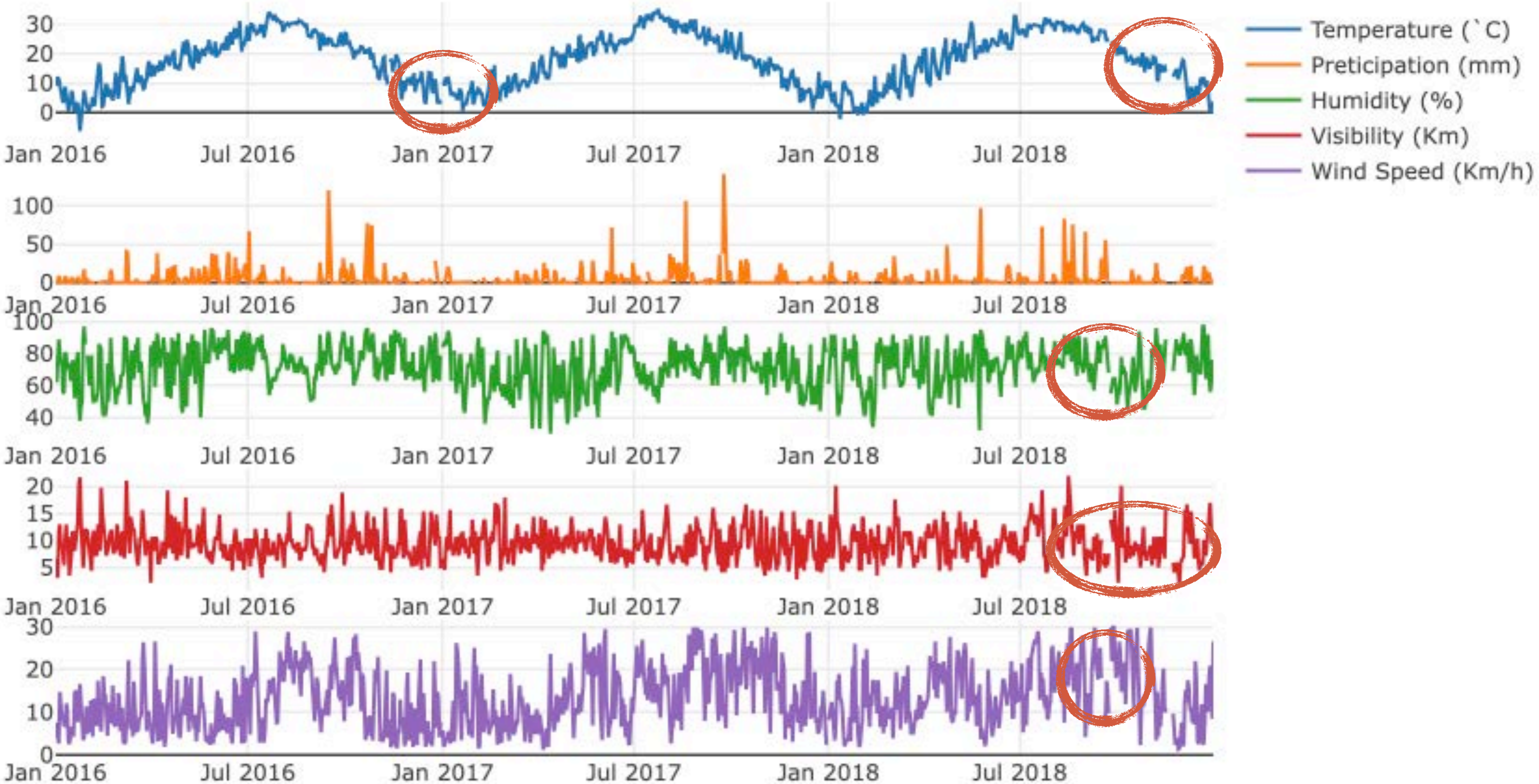
➤ 2016-2019 Shanghai weather





MISSING DATA

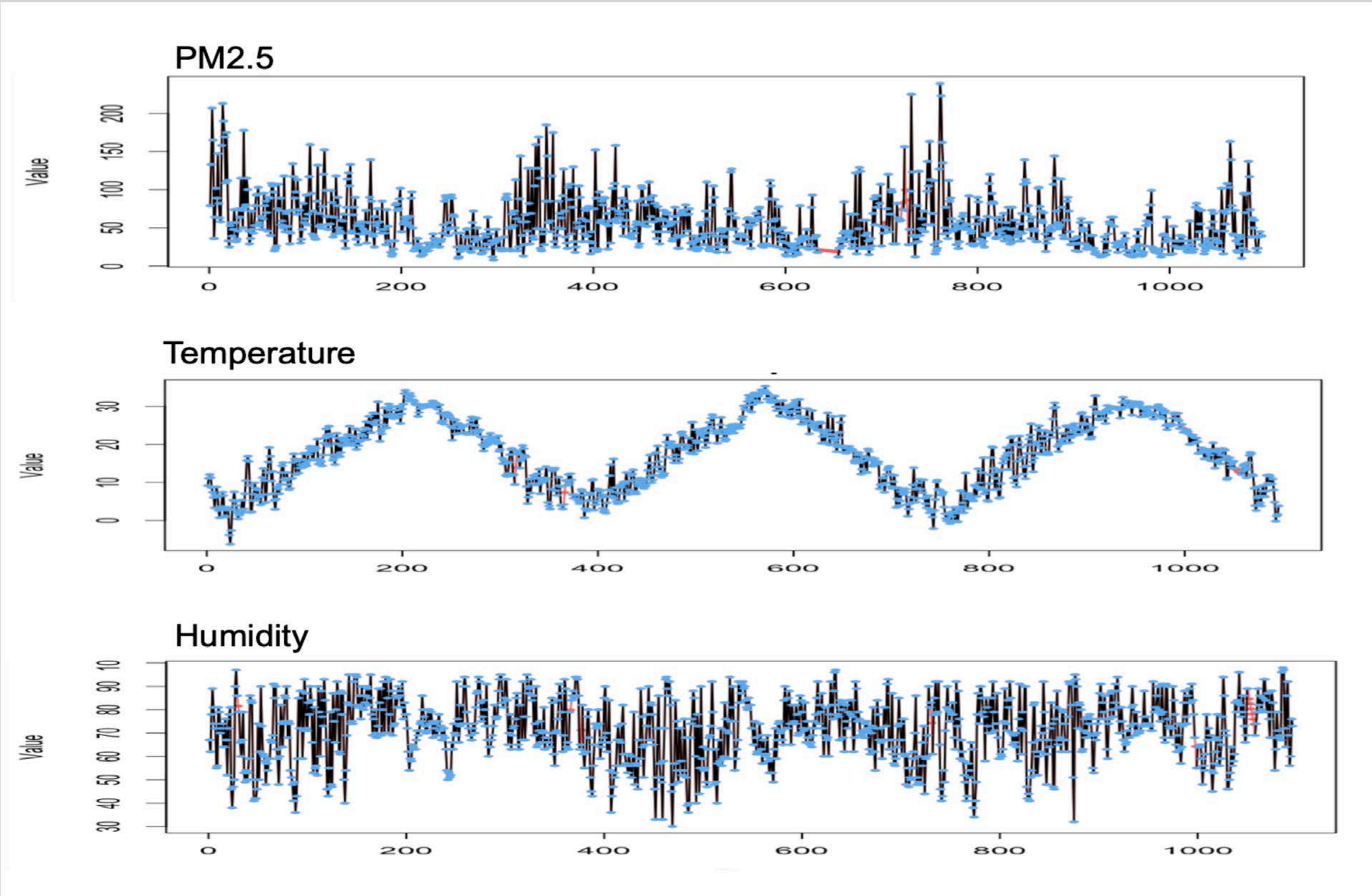
- 2016-2019 Shanghai weather



MISSING DATA

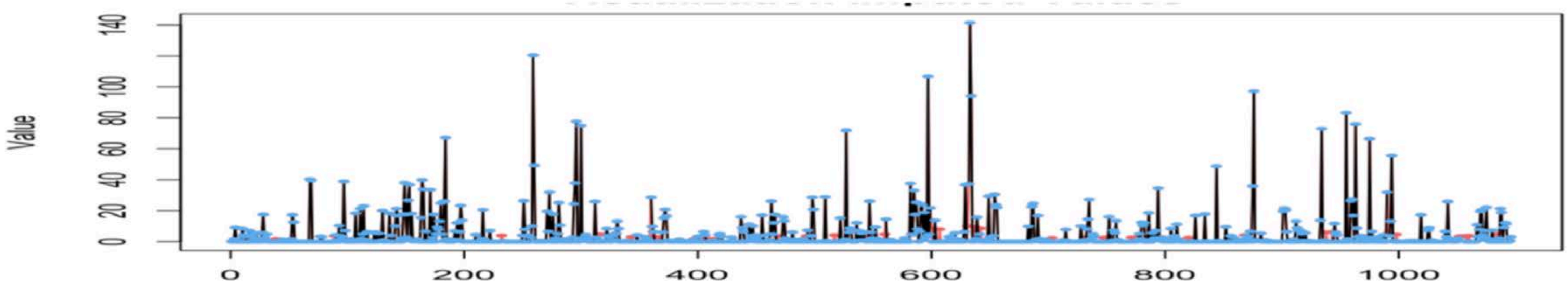
- Imputing Method:
    - EM-Algorithm:
        - E-Step: A Kalman filter
        - M-Step: Use the filtered estimates within maximum-likelihood calculations to obtain updated parameter estimates

MISSING DATA

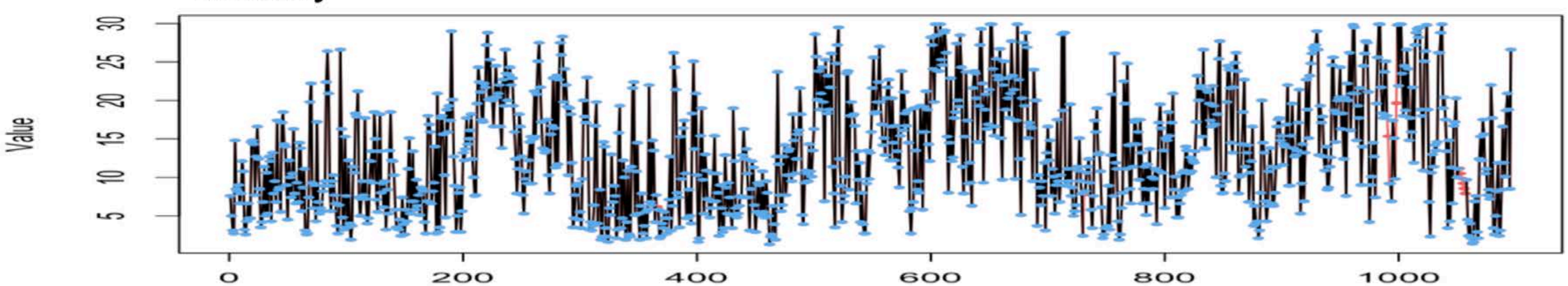


MISSING DATA

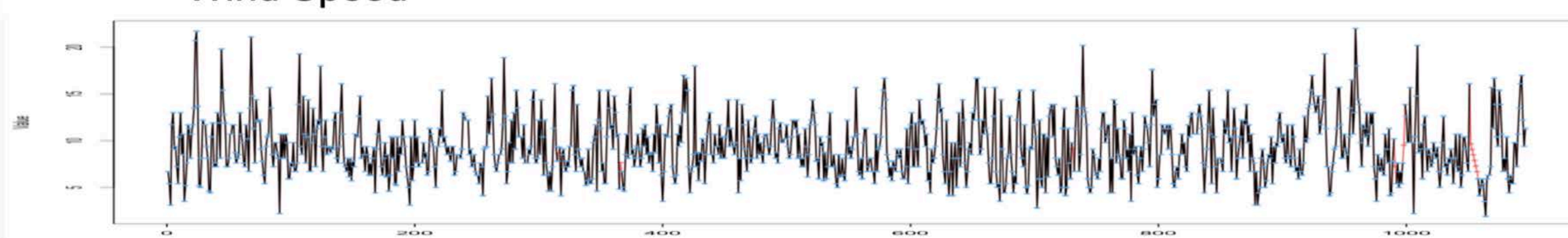
Rainfall



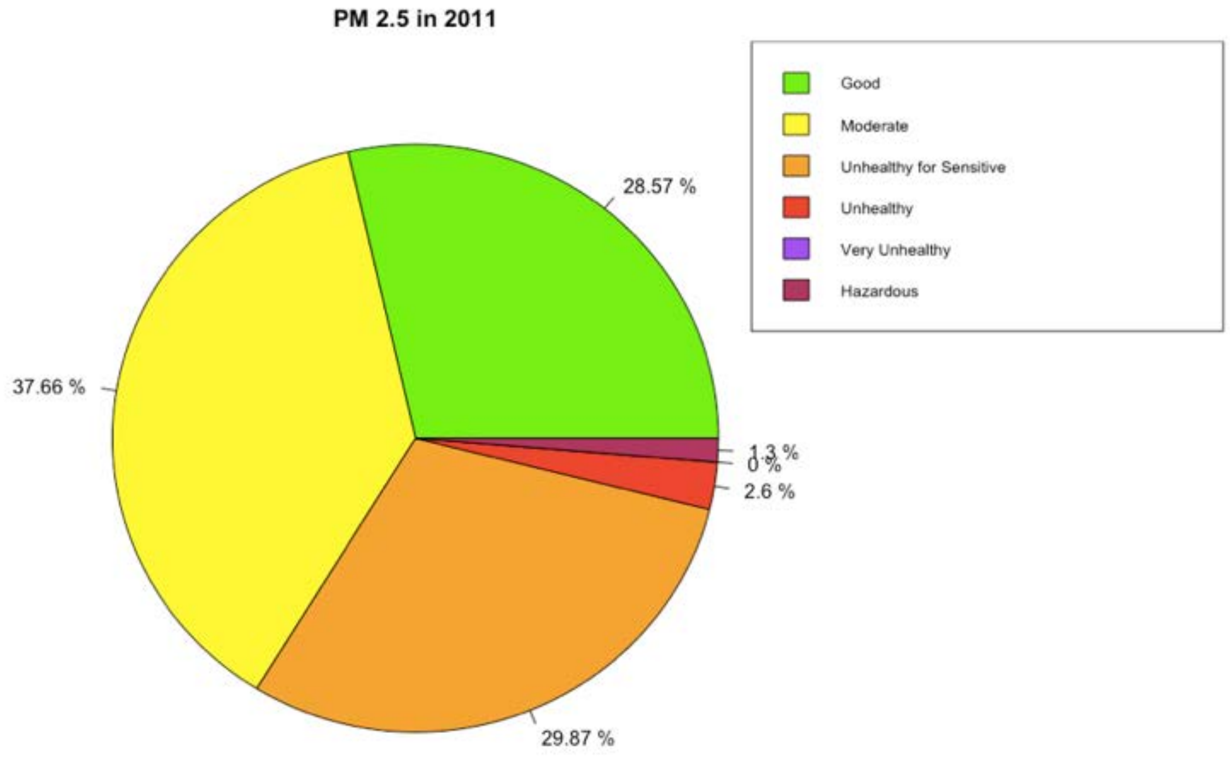
Visibility

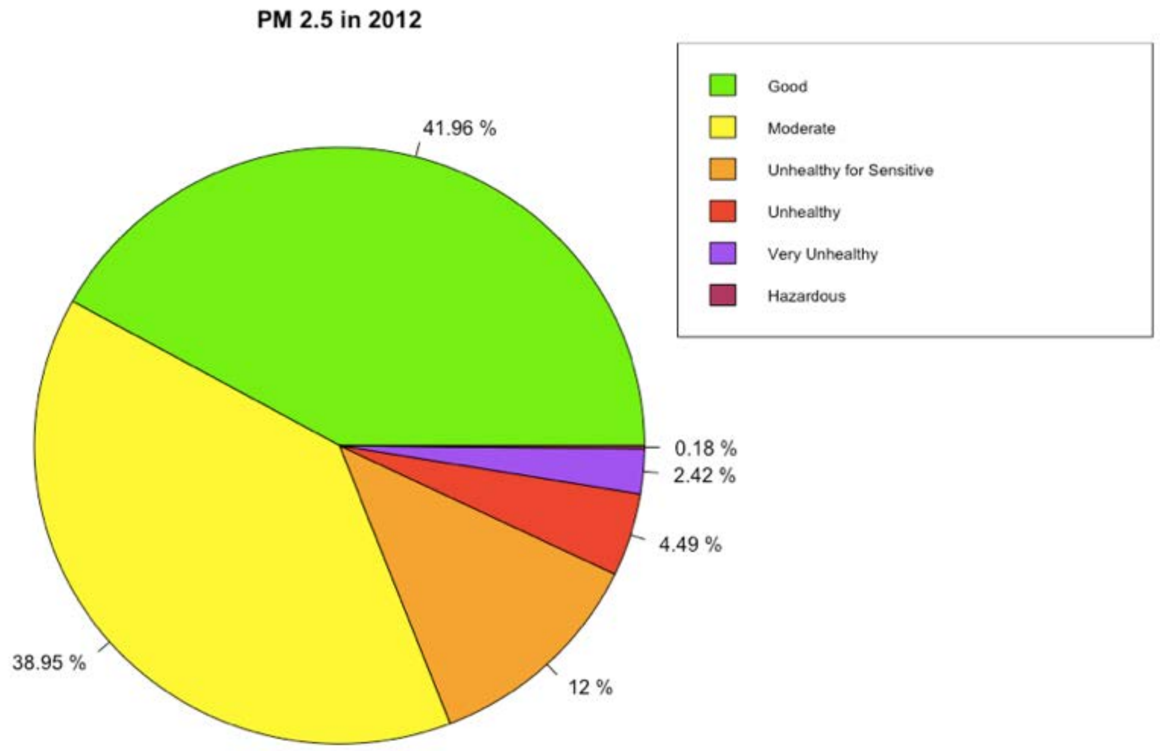


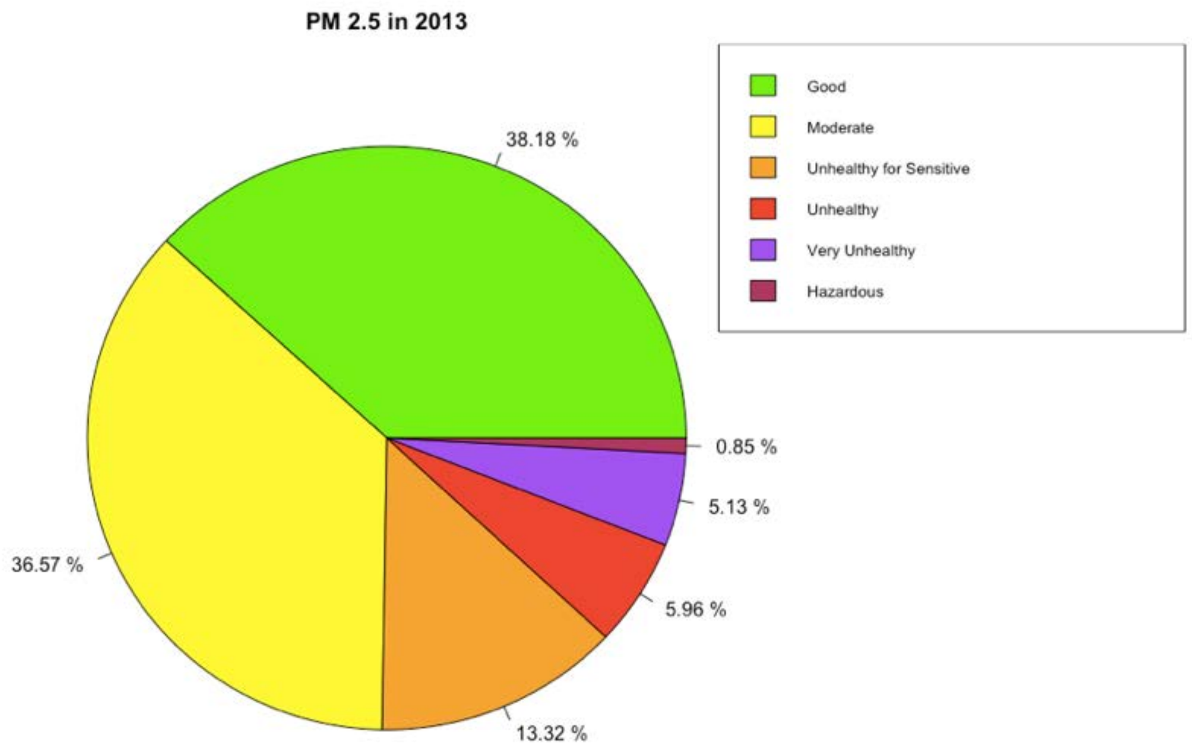
Wind Speed

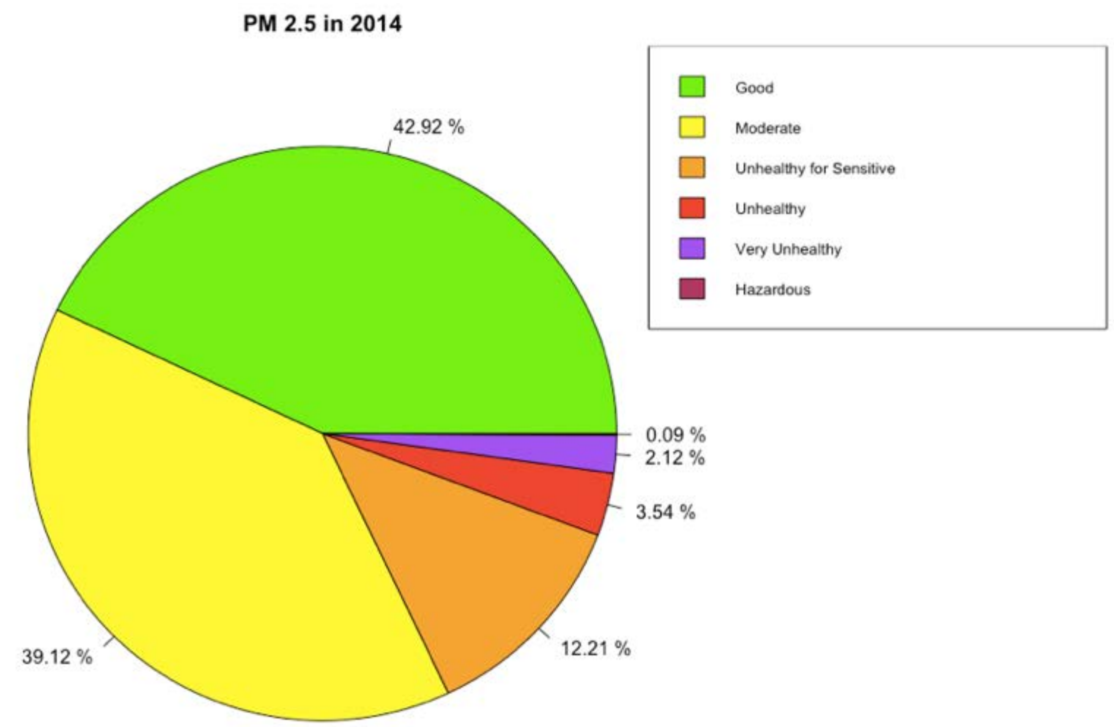


TRENDS IN RECENT YEARS



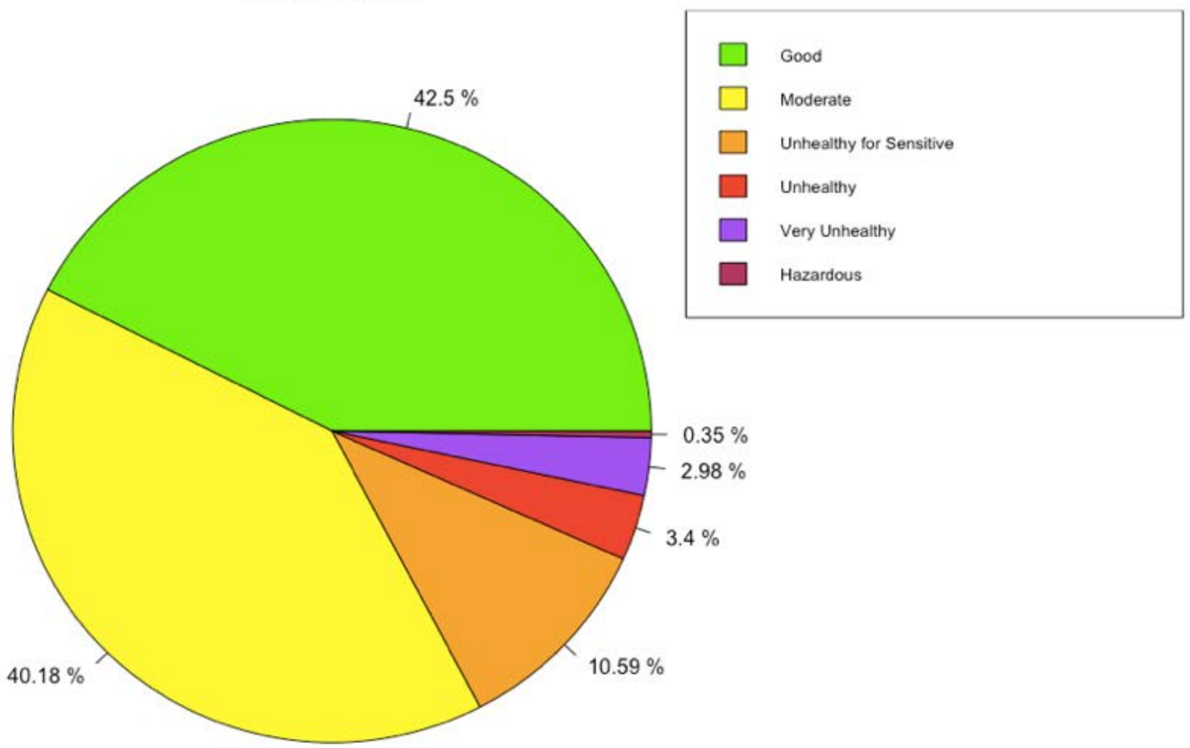




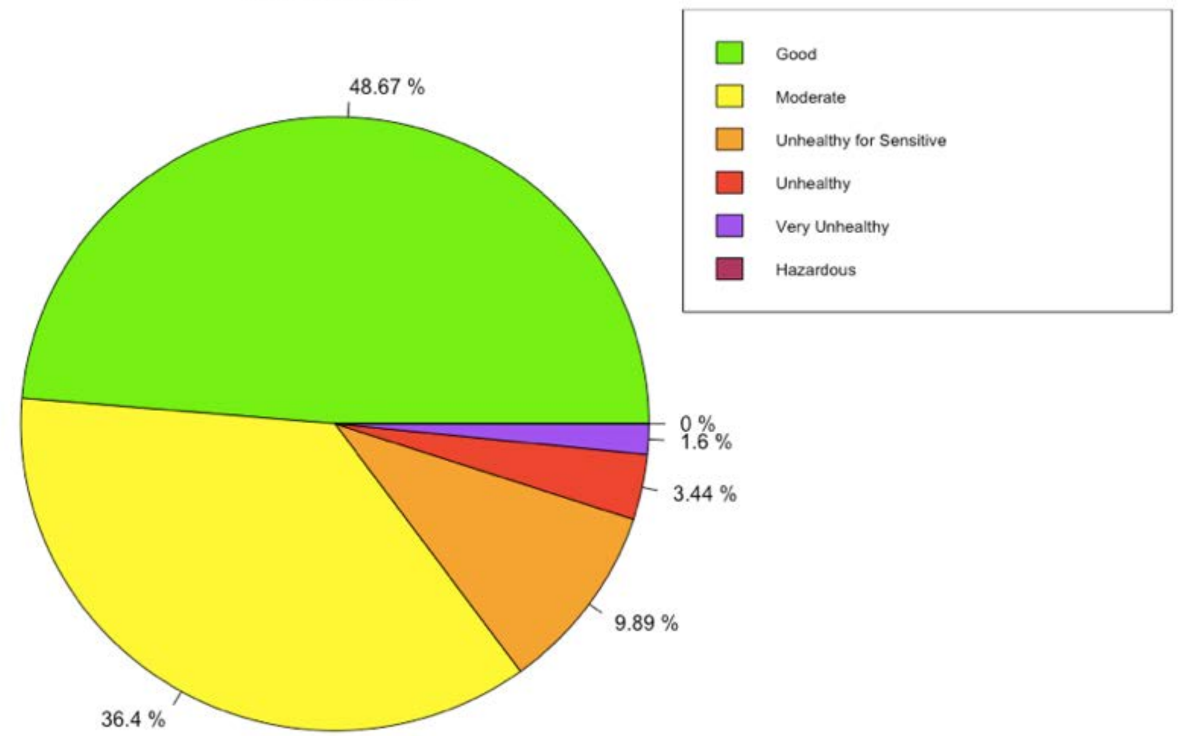


TRENDS IN RECENT YEARS

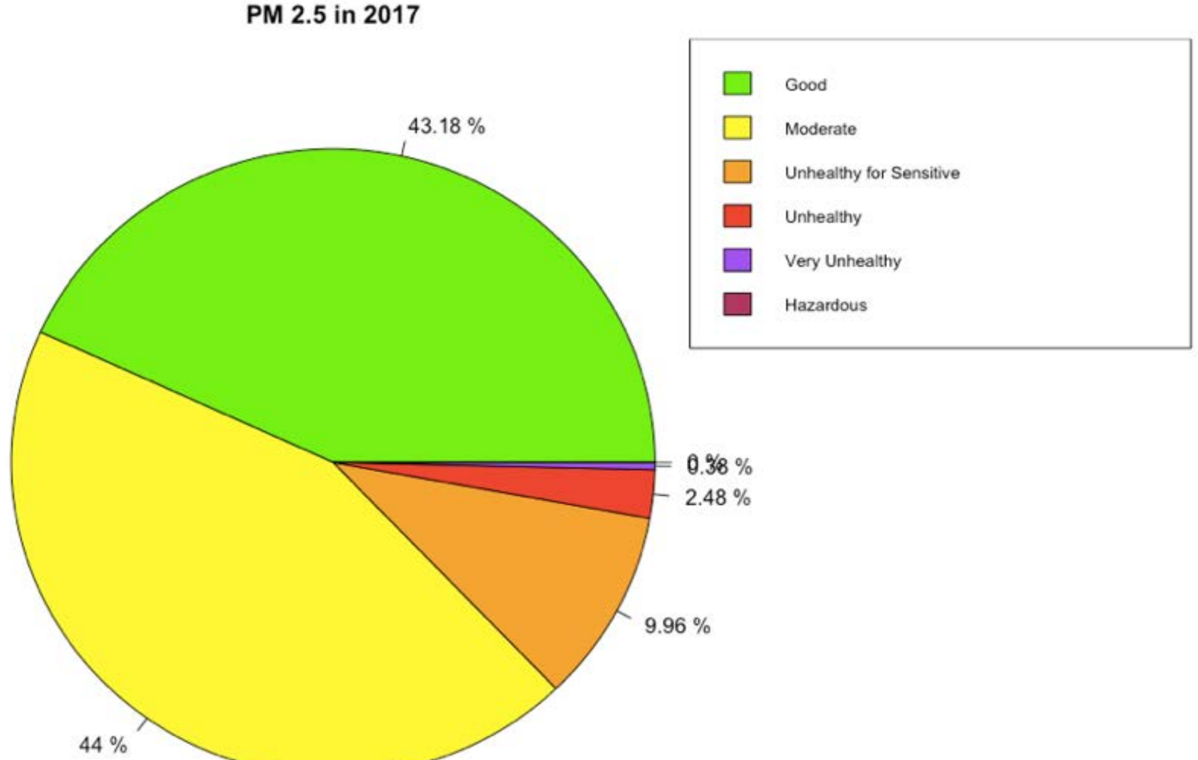
PM 2.5 in 2015



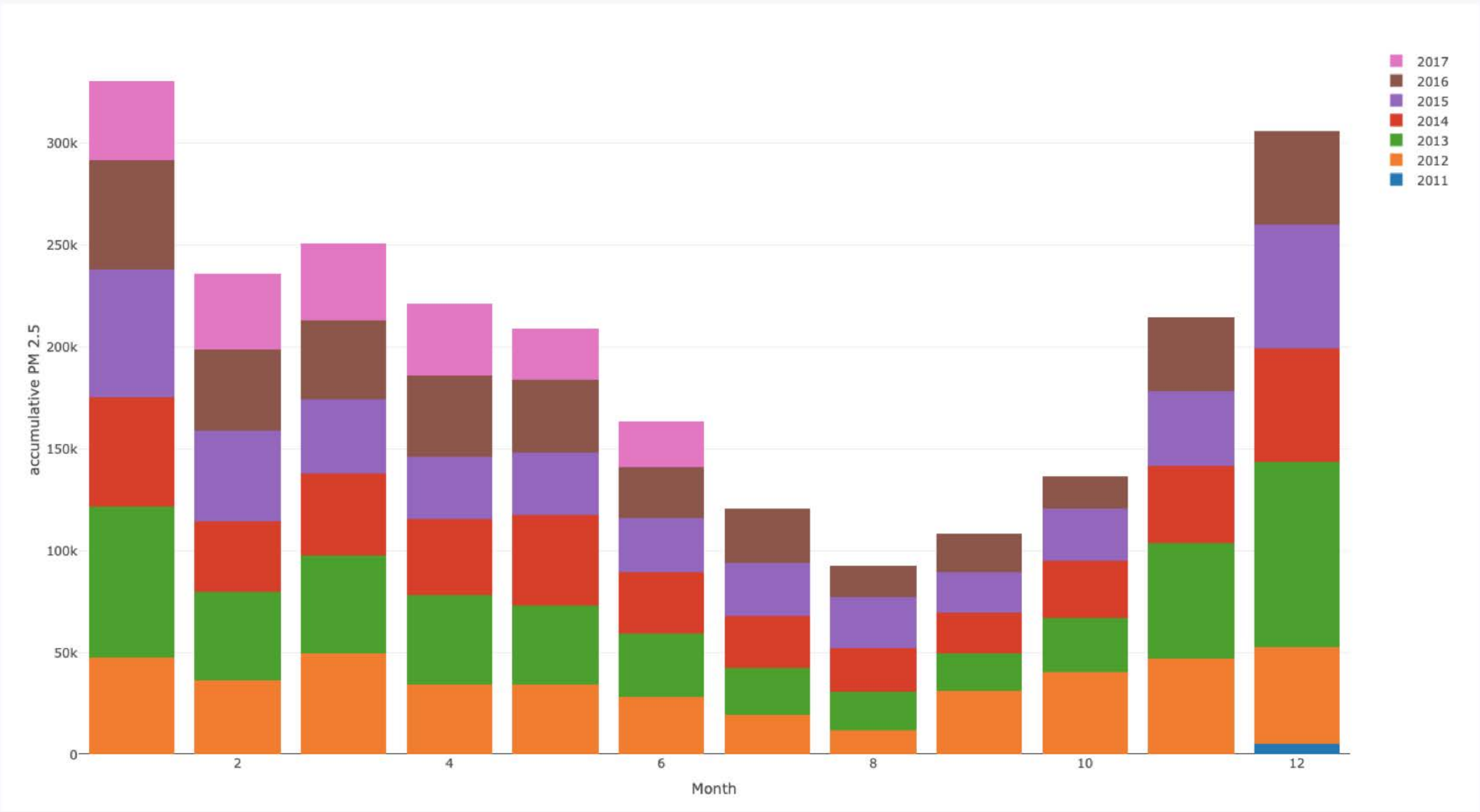
PM 2.5 in 2016



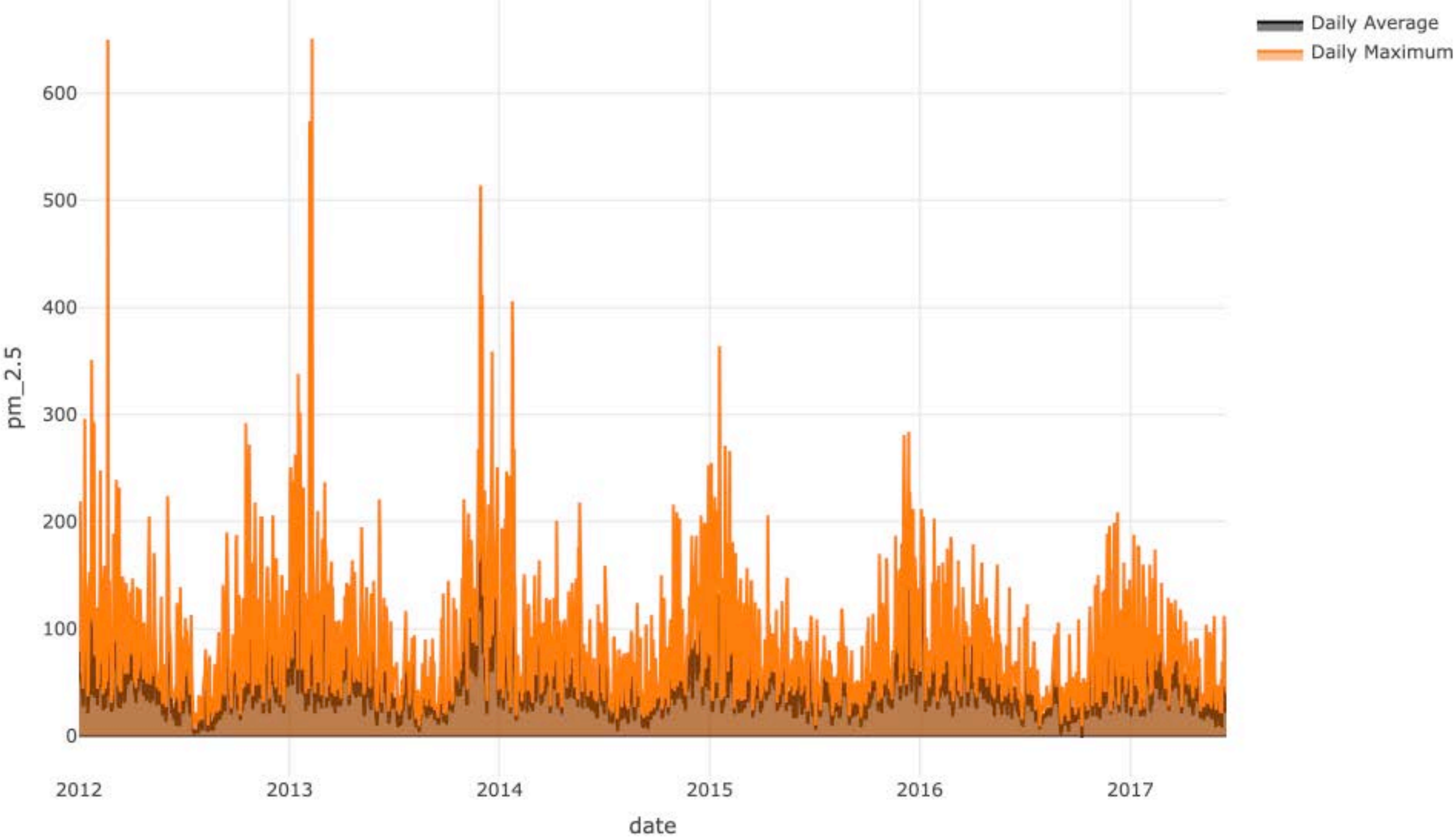
PM 2.5 in 2017



TRENDS IN RECENT YEARS

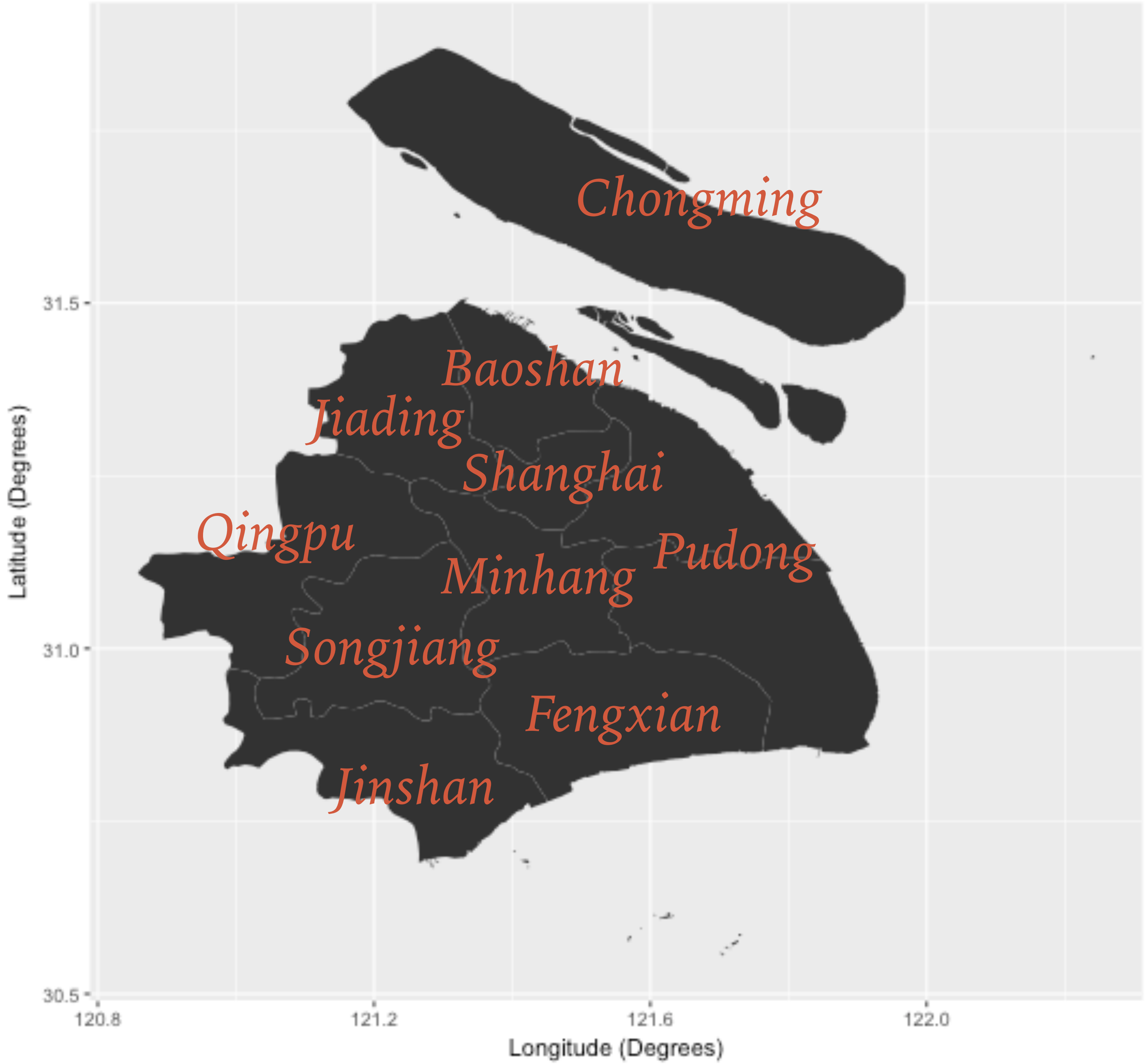


TRENDS IN RECENT YEARS



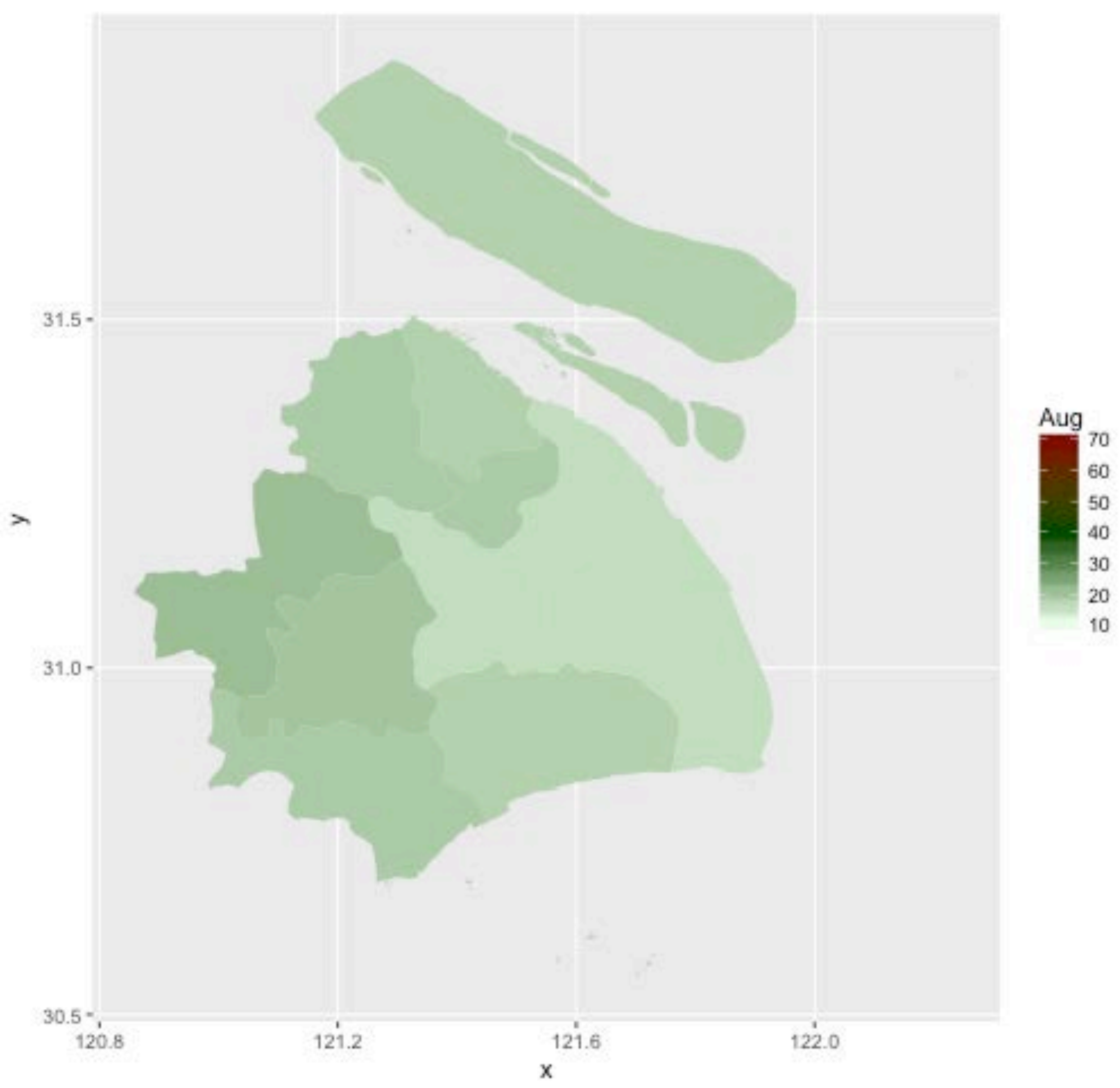
MORE EXPLORATION

- Shanghai map from GIS

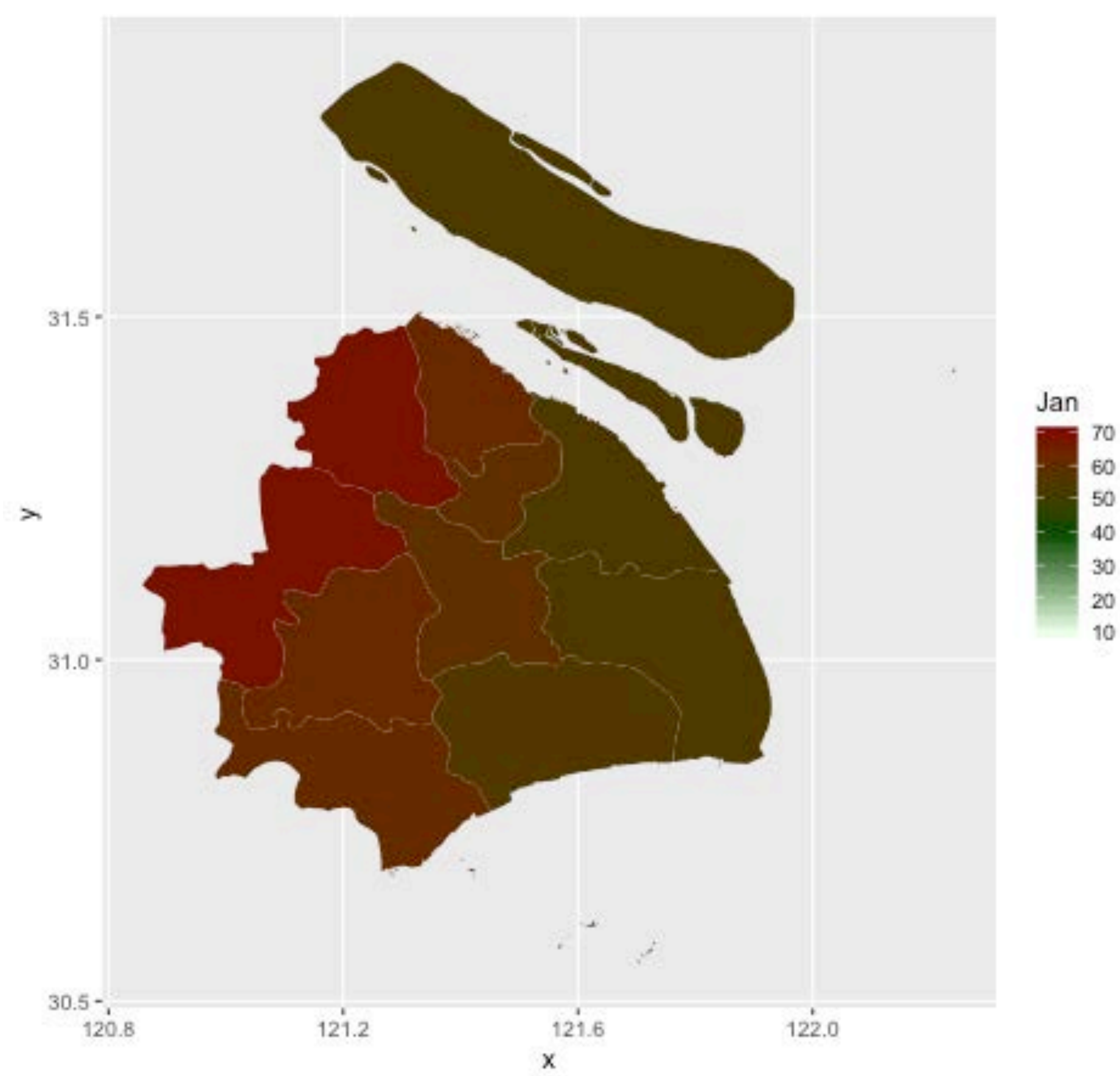


MORE EXPLORATION

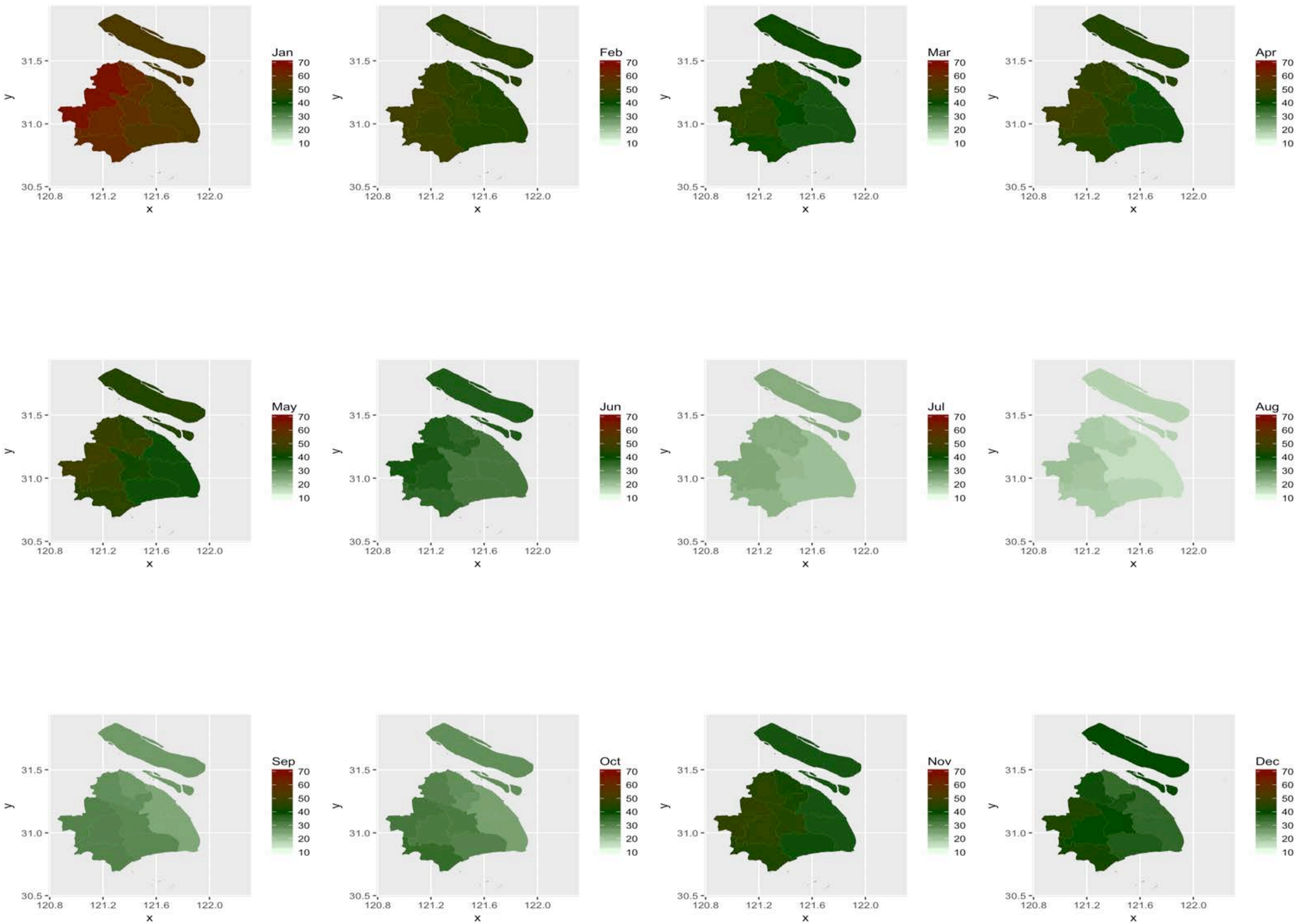
SUMMER



WINTER

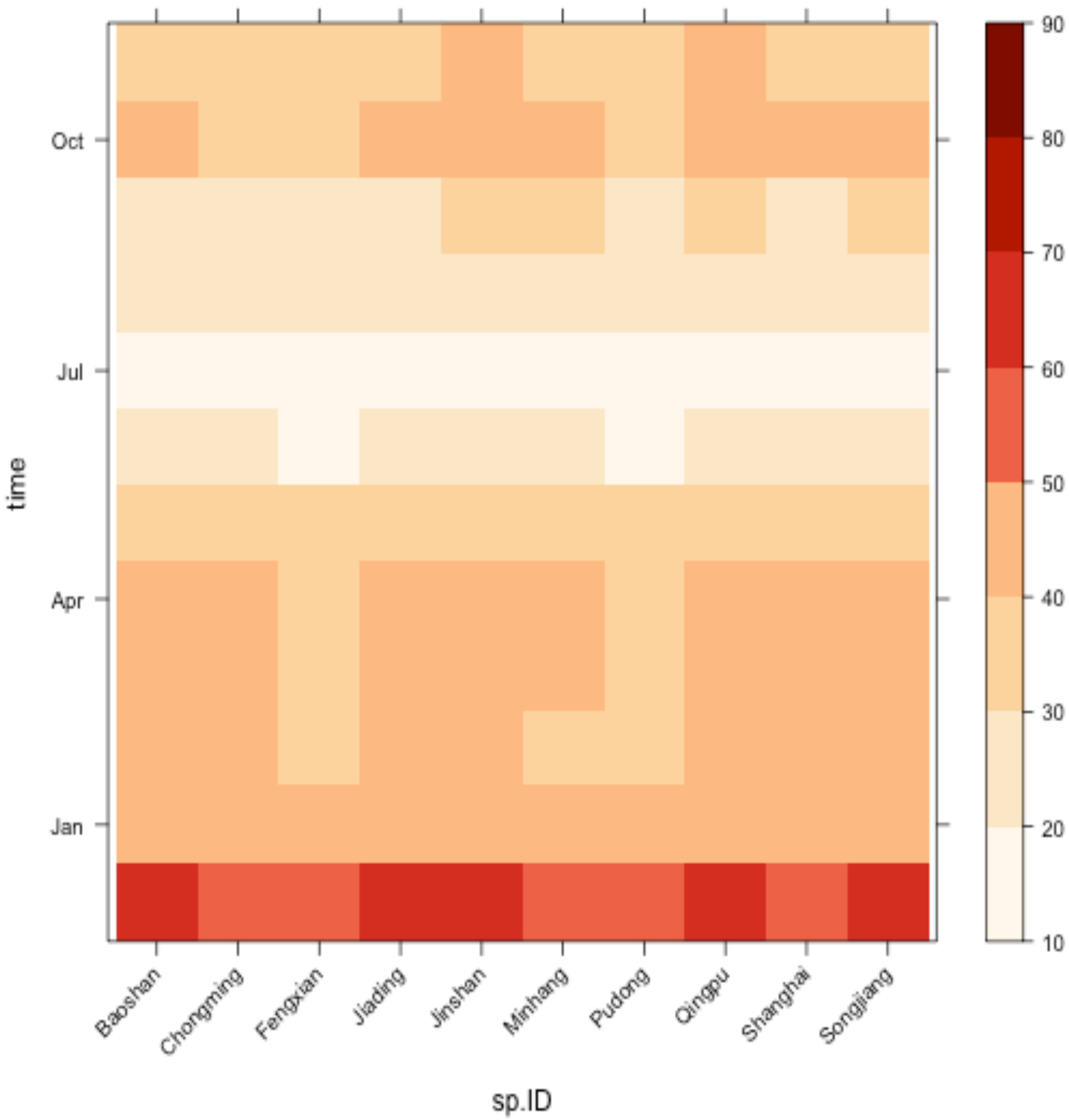


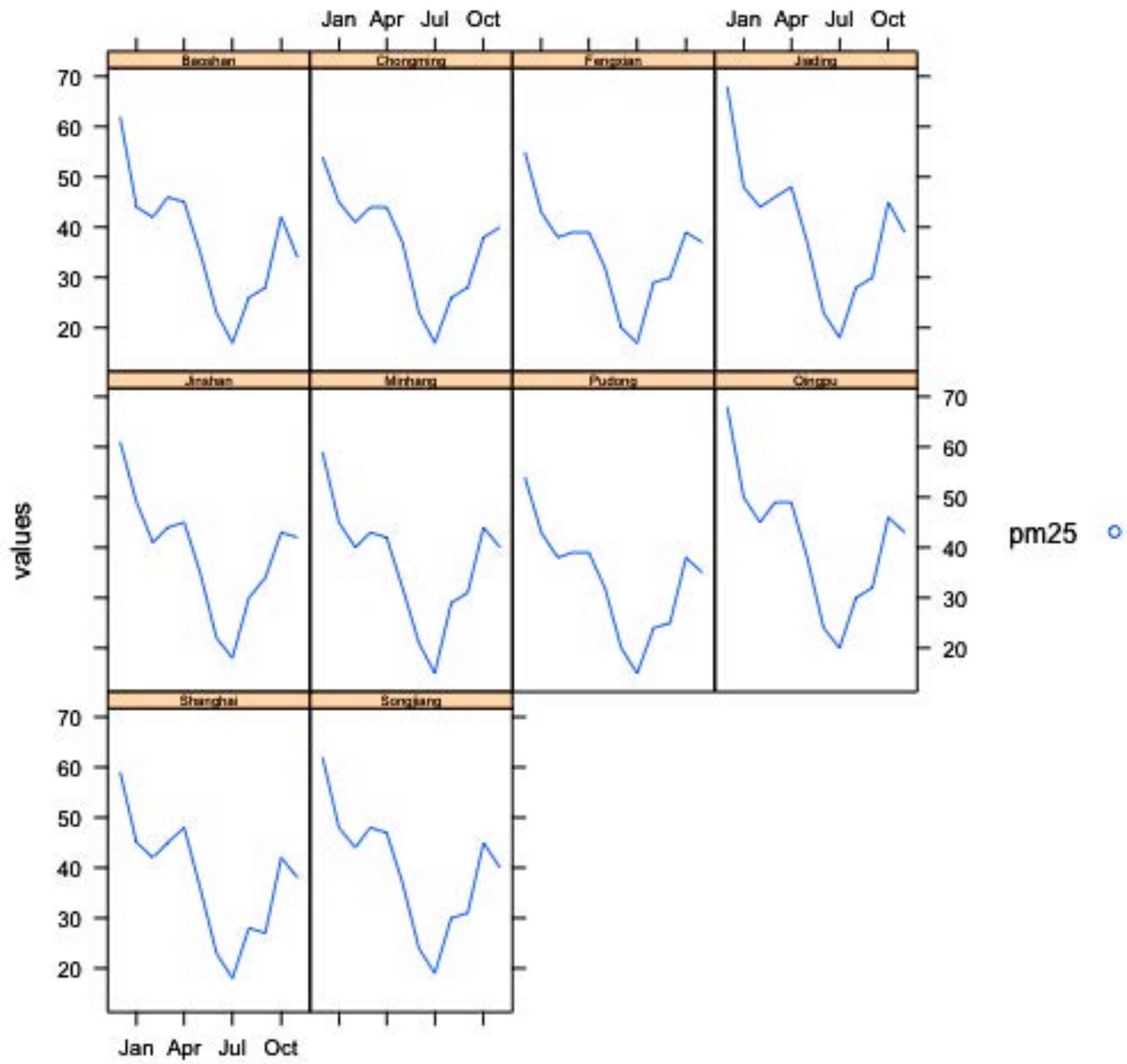
MORE EXPLORATION



MORE EXPLORATION

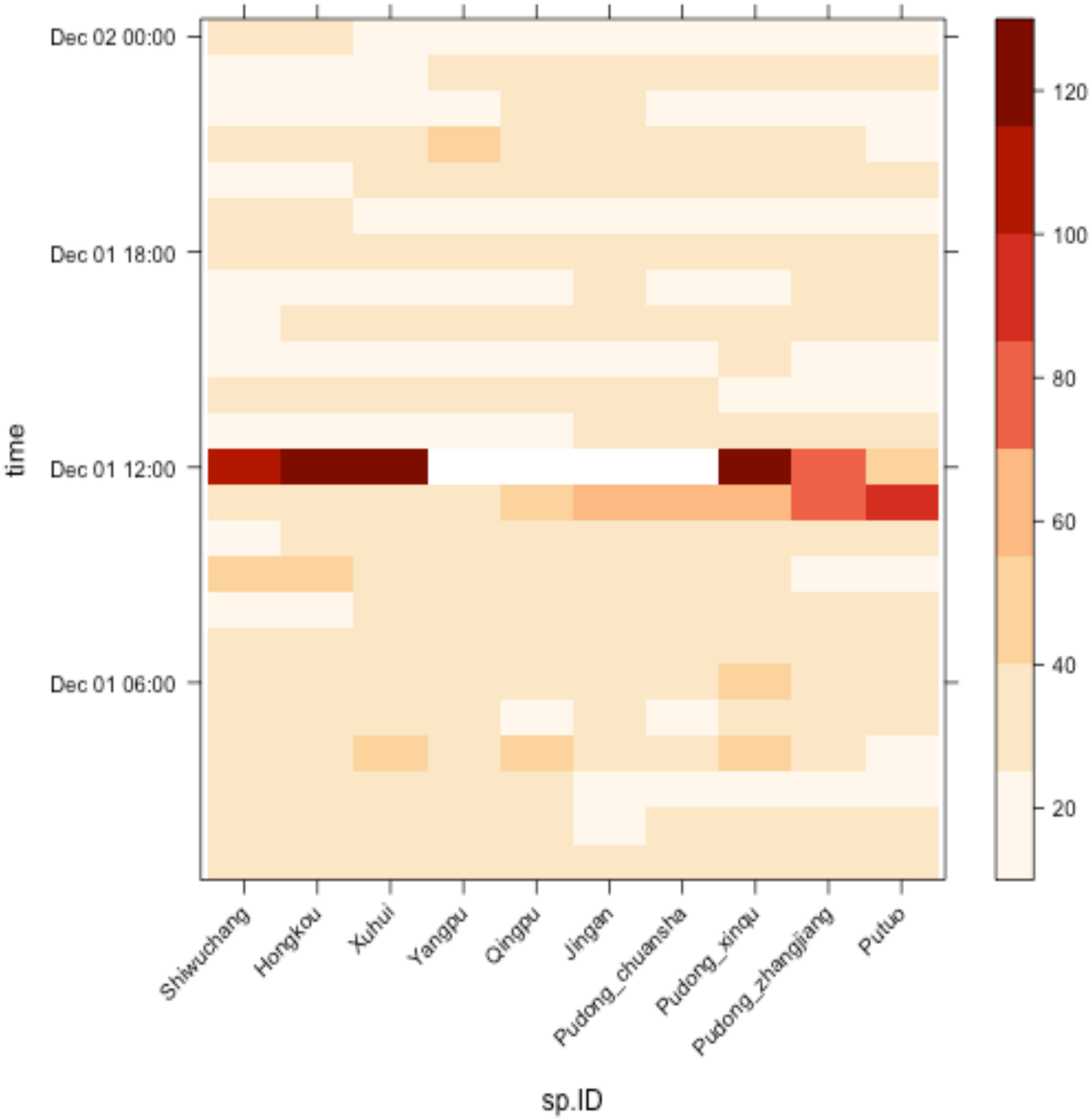
➤ Monthly Plots in 2018

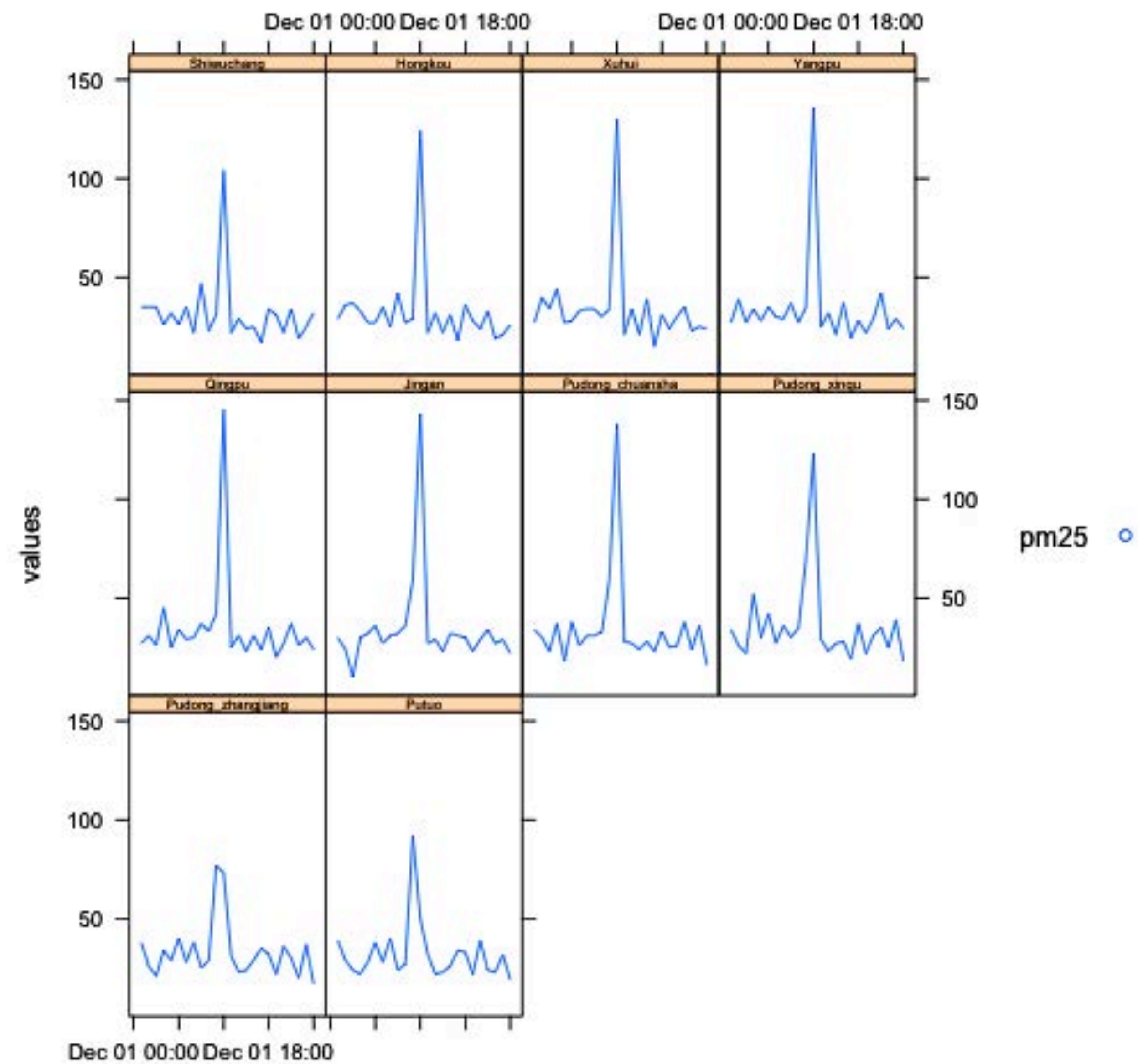




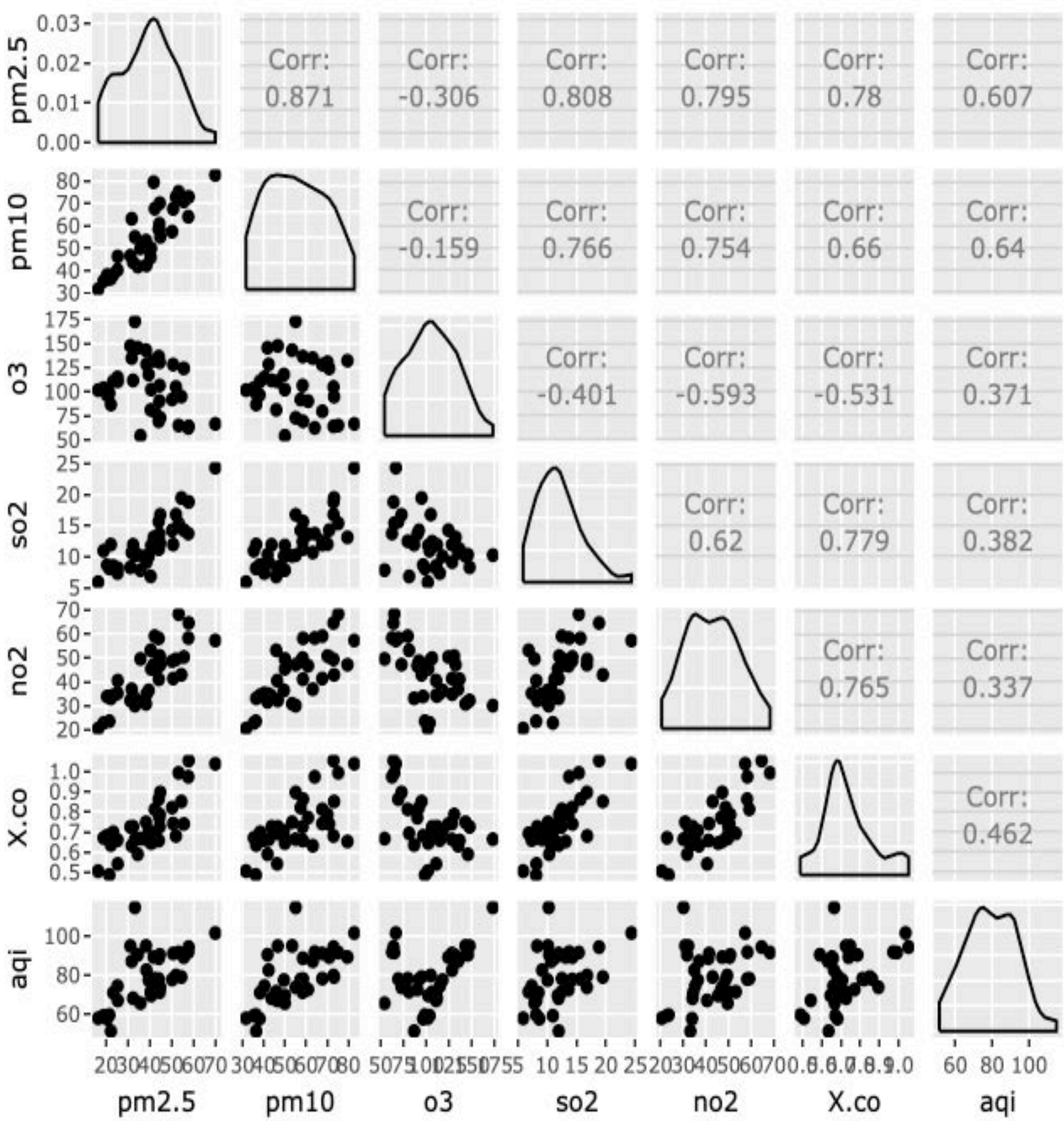
MORE EXPLORATION

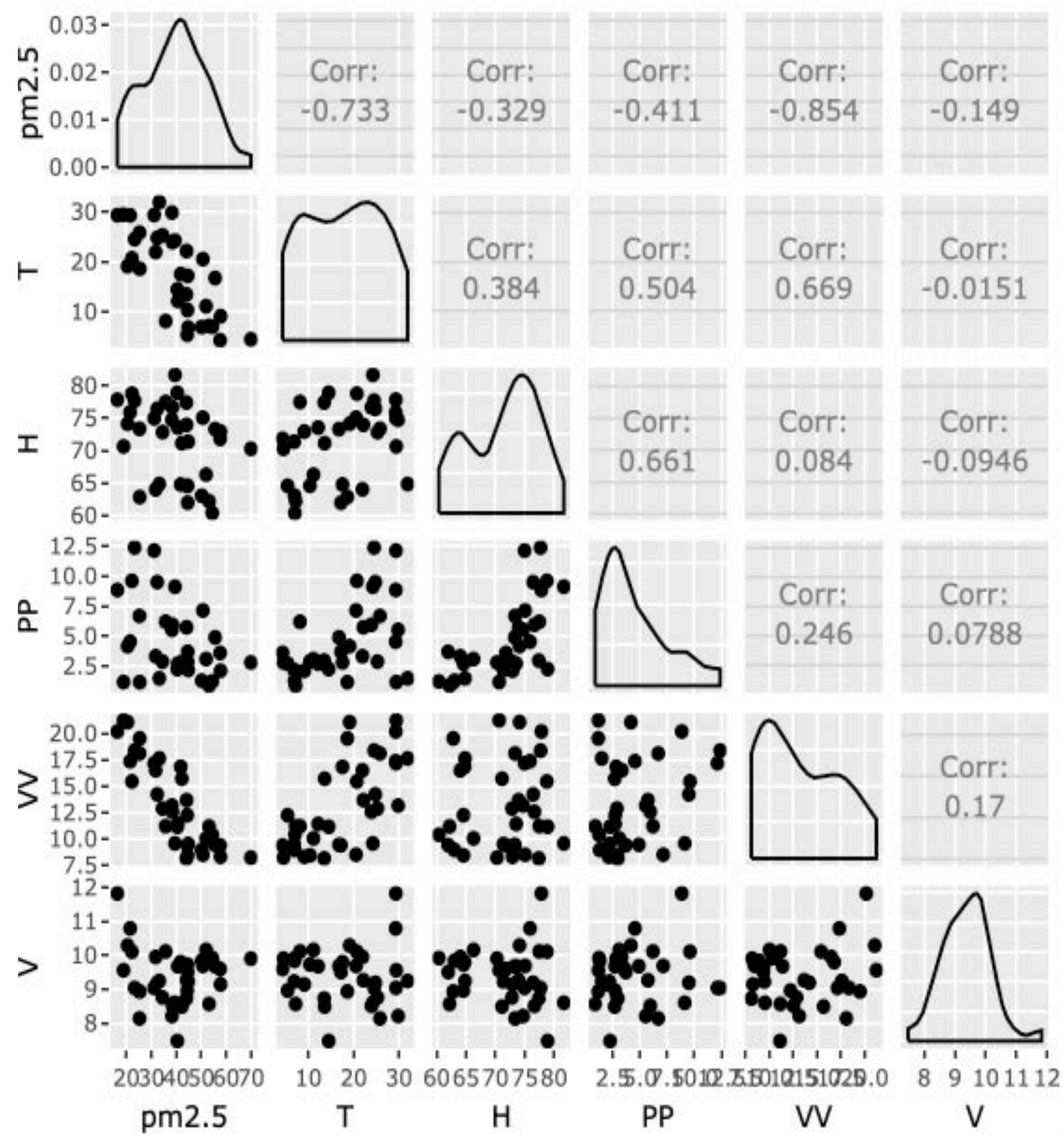
- Dec. 01, 2018





CORRELATION





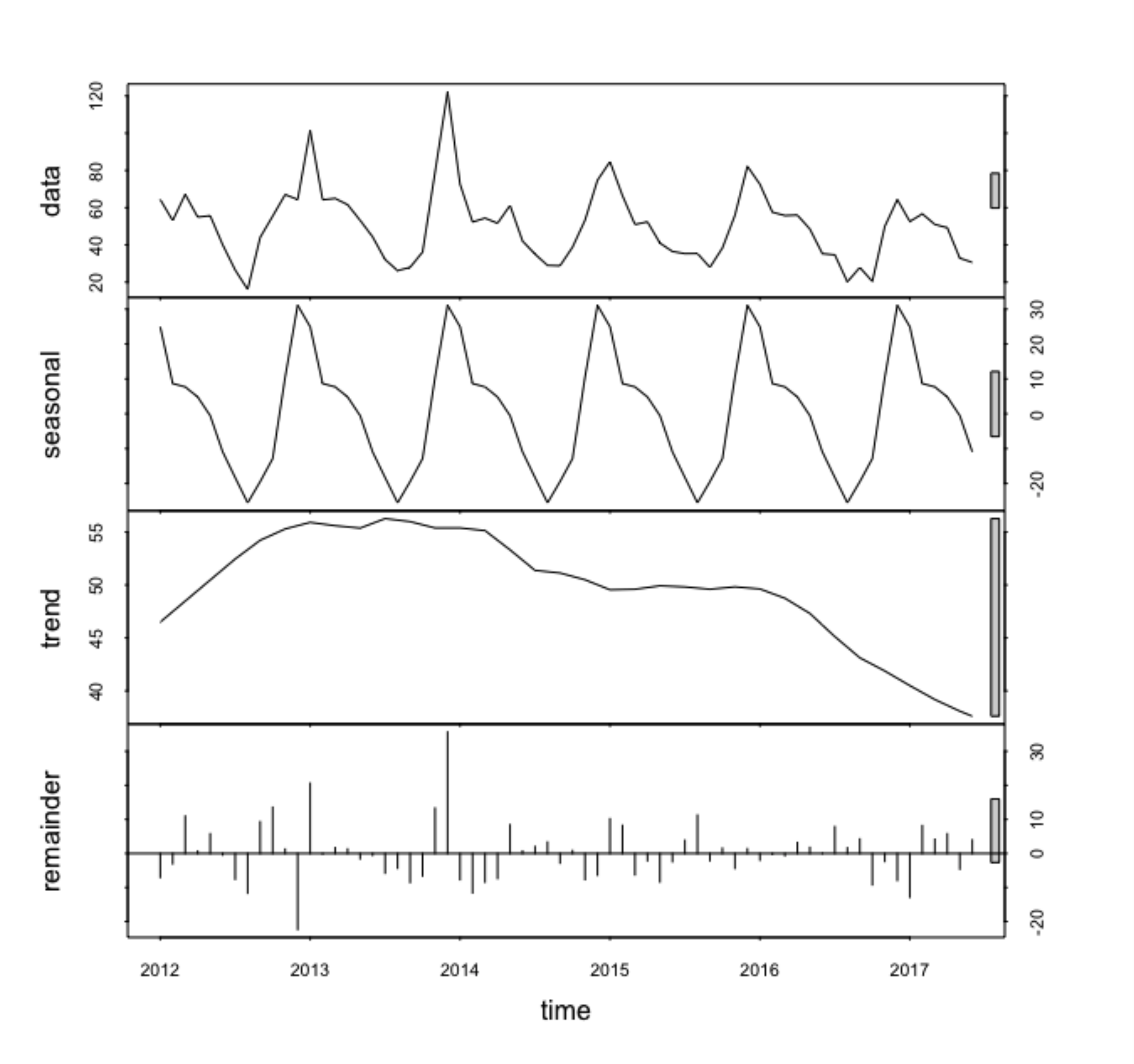
STATISTICAL ANALYSIS

TRENDS ANALYSIS

- **Time series components**
  - $y_t = S_t + T_t + R_t$
  - y_t is the data
  - S_t the seasonal component
  - T_t is the trend-cycle component
  - R_t is the remainder component

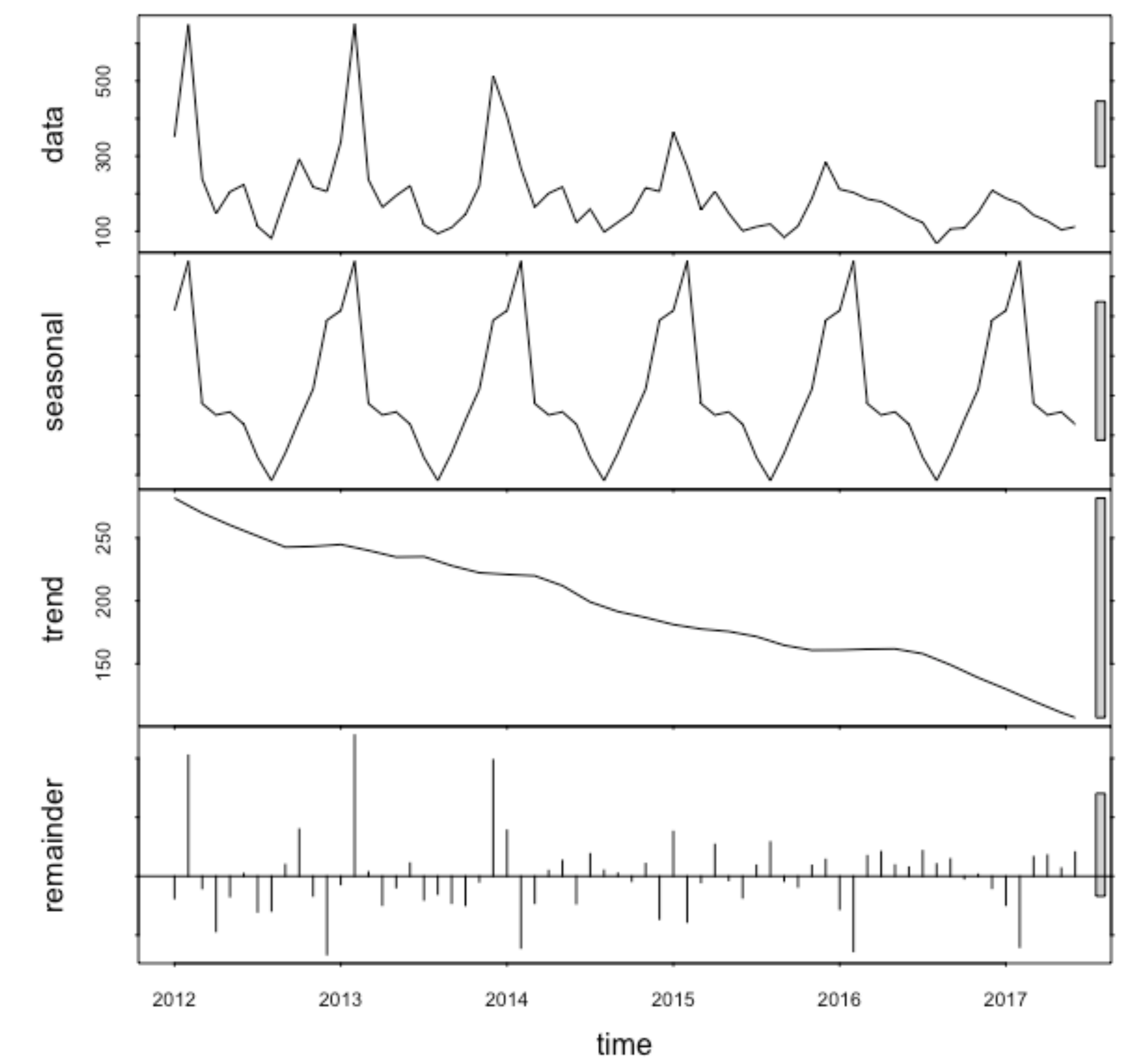
TRENDS ANALYSIS

- Monthly Average



TRENDS ANALYSIS

➤ Monthly Maximum



TRENDS ANALYSIS

- **Unit Root Test**
  - **Augmented Dickey-Fuller test (ADF)**
    - Test the null hypothesis that a unit root is present in a time series sample
    - Test statistic: $DF_\tau = \dfrac{\hat{\gamma}}{SE(\hat{\gamma})}$
    - $H_0 : \gamma = 0$ v.s. $H_a : \gamma \neq 0$

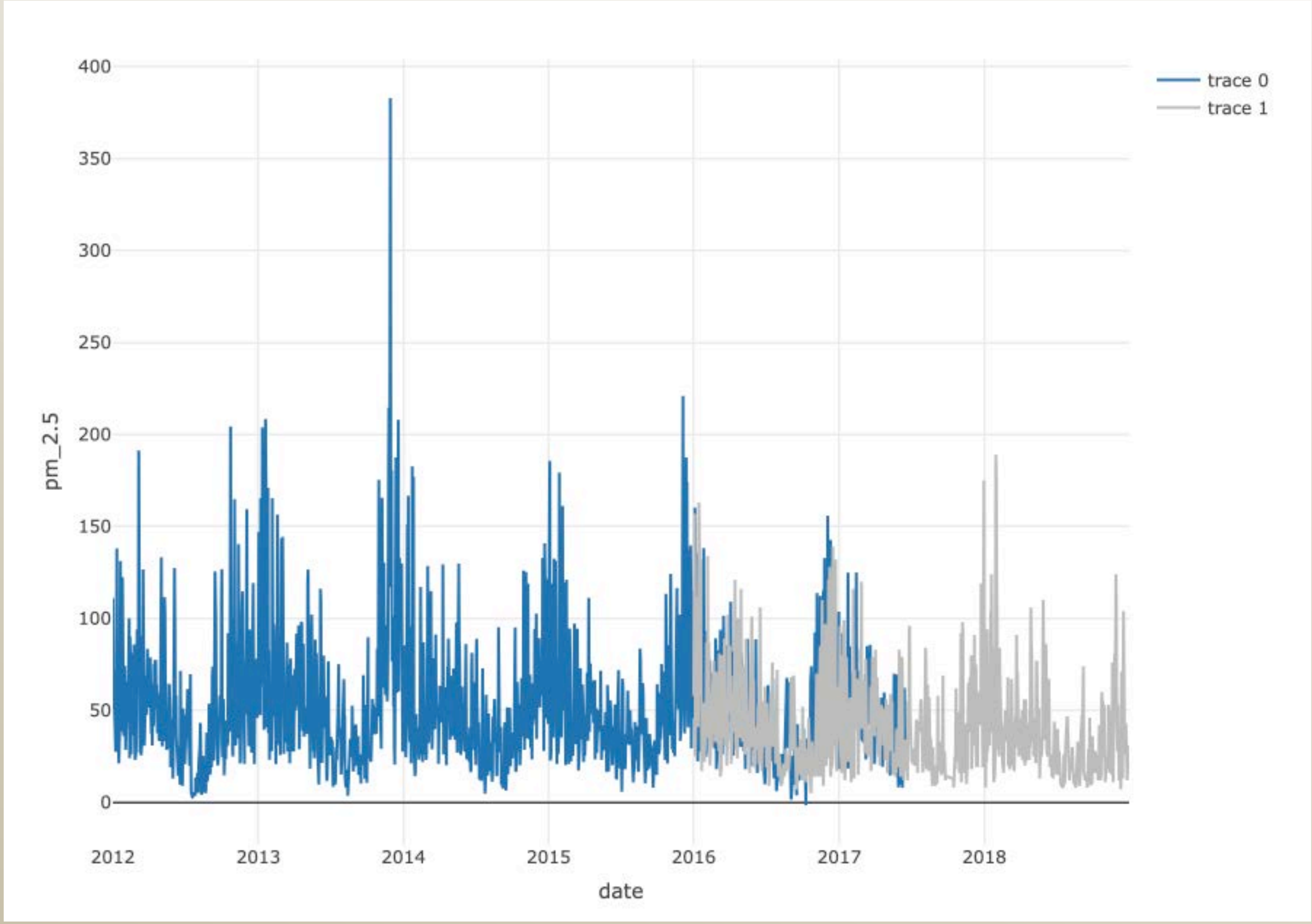
TRENDS ANALYSIS

- Test results

Augmented Dickey-Fuller Test			
	Dickey Fuller	Lag order	p-value
Avg.	-5.5315	4	0.01
Max.	-5.1485	4	0.01

- Stationary

FITTING MODELS



FITTING MODELS

- **MODEL 1: ARIMA MODEL**
  - Autoregressive integrated moving average
  - AR: Variable of interest is regressed on its own lagged (i.e., prior) values
  - MA: The regression error is actually a linear combination of error terms whose values occurred simultaneously and at various times in the past
  - I: Integrated, difference between current values and previous values

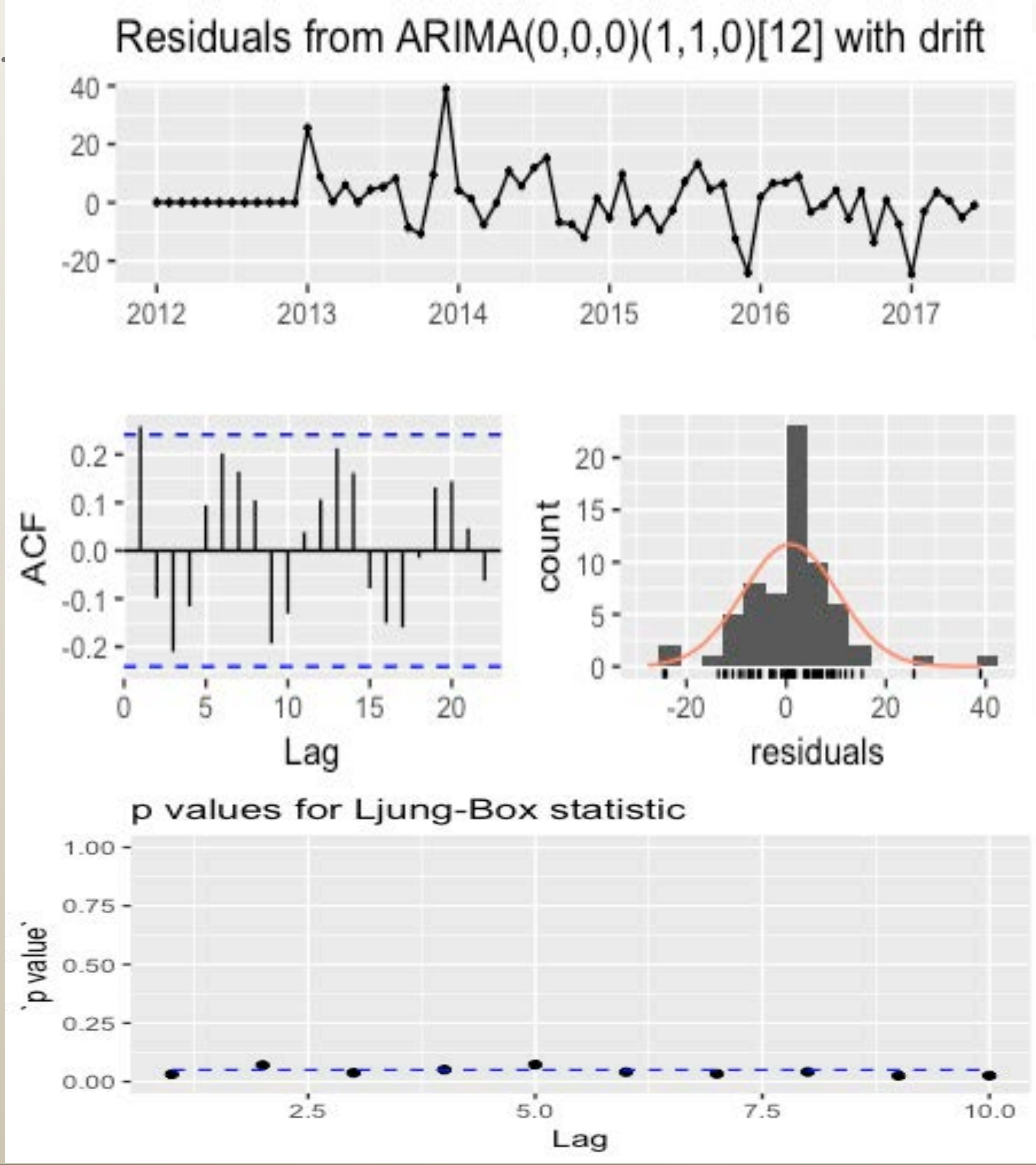
FITTING MODELS

- Monthly Average
- ARIMA(0,0,0)(1,1,0)[12] with drift

```
ARIMA(0,0,0)(1,1,0)[12] with drift

Coefficients:
         sar1    drift
      -0.7644  -0.2215
s.e.   0.0854   0.0752

sigma^2 estimated as 114.7:  log likelihood=-208.92
AIC=423.84   AICc=424.32   BIC=429.81

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 0.8567694 9.507755 6.325077 0.4967536 13.55823 0.6083306 0.2588349
```

FITTING MODELS

- Monthly Average



FITTING MODELS

- **MODEL 2: Exponential Smoothing Methods**
  - Recent observations are given relatively more weight in forecasting than the older observations.

$$s_0 = x_0$$
$$s_t = \alpha x_t + (1-\alpha)s_{t-1}, t > 0$$

  - Well known models such as simple or single exponential smoothing, Holt's linear trend method and the Holt-Winters seasonal method

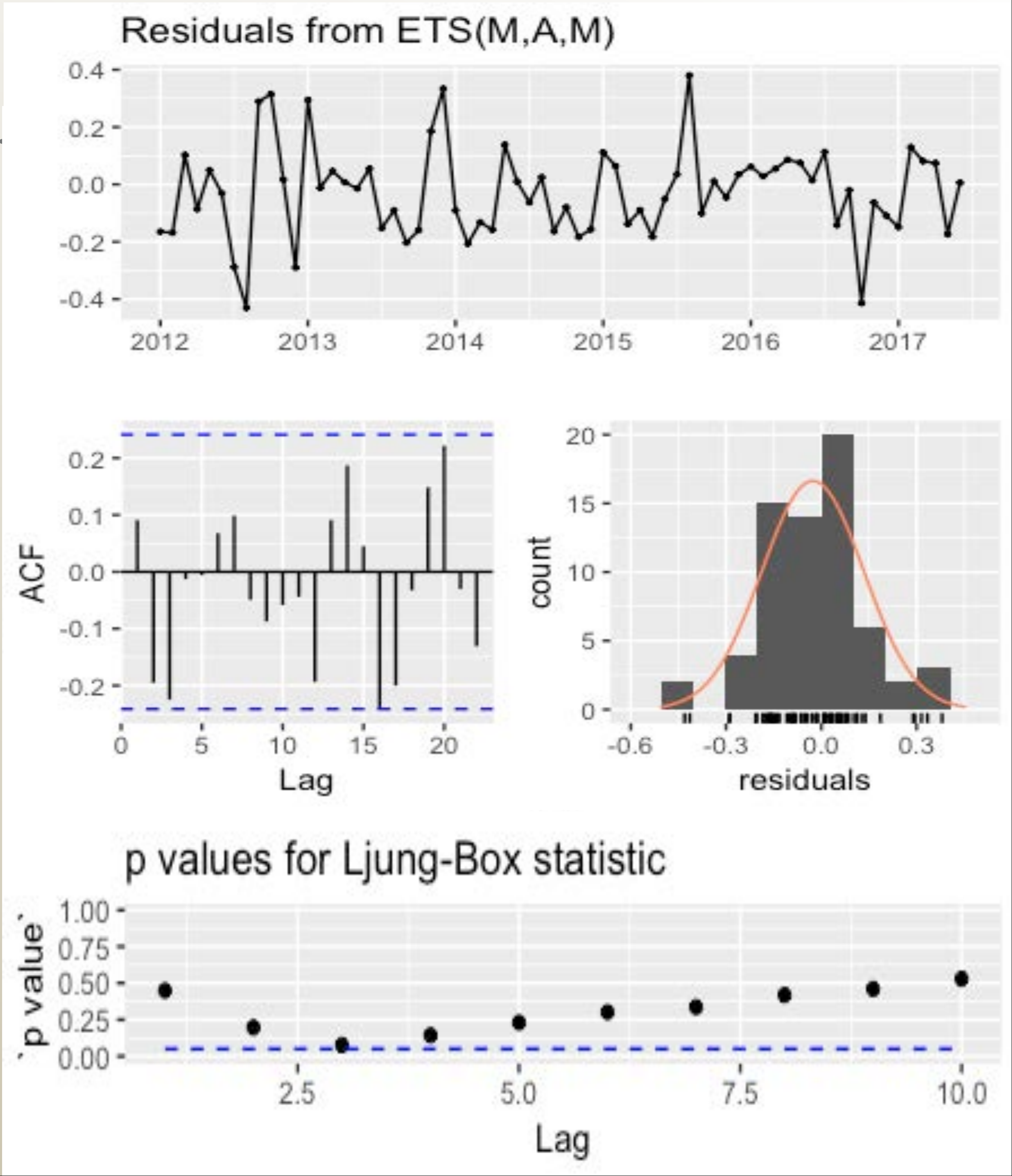
FITTING MODELS

- Monthly Average
- Multiplicative Holt-Winters' method with multiplicative errors

```
ETS(M,A,M)

Call:
 ets(y = pm2.5mon)

  Smoothing parameters:
    alpha = 0.0098
    beta  = 0.0098
    gamma = 1e-04

  Initial states:
    l = 53.6291
    b = 0.3473
    s = 1.6466 1.2017 0.7724 0.6273 0.515 0.678
           0.7539 0.971 1.1032 1.1231 1.1792 1.4287

  sigma:  0.1855

     AIC     AICc      BIC
583.5733 596.3233 620.7975

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set -1.221448 8.840635 6.511173 -5.720211 13.89125 0.6262288 0.02893665
```

FITTING MODELS

- Monthly Average



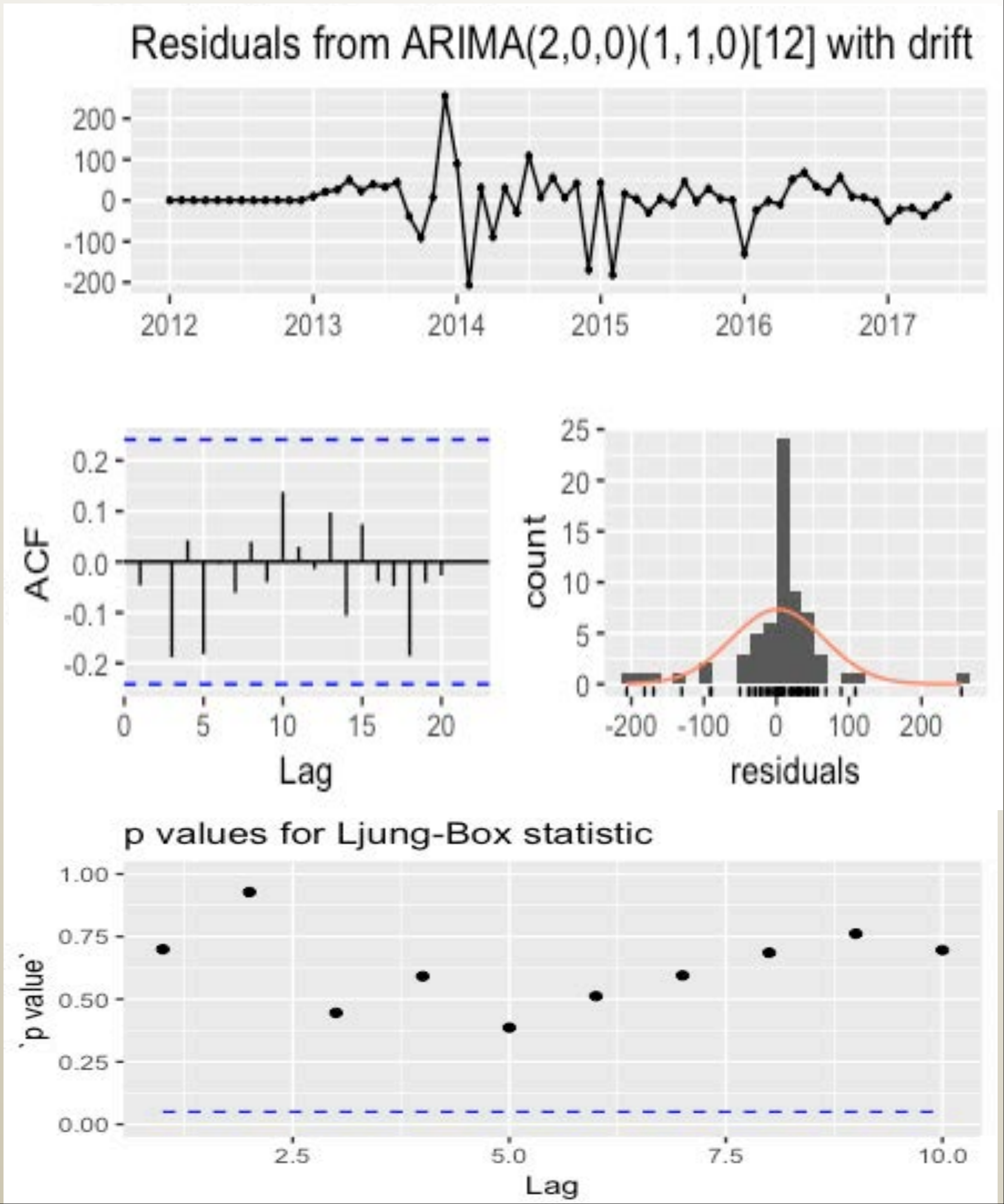
FITTING MODELS

- Monthly Maximum
- ARIMA(2,0,0)(1,1,0)[12] with drift

```
ARIMA(2,0,0)(1,1,0)[12] with drift

Coefficients:
         ar1      ar2     sar1    drift
      0.0627  -0.4892  -0.3654  -2.1277
s.e.  0.1179   0.1181   0.1183   0.4412

sigma^2 estimated as 5286:  log likelihood=-307.14
AIC=624.27   AICc=625.52   BIC=634.22

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 1.973795 63.28218 36.97267 1.776031 19.88257 0.7079874 -0.0464978
```

FITTING MODELS

- Monthly Maximum



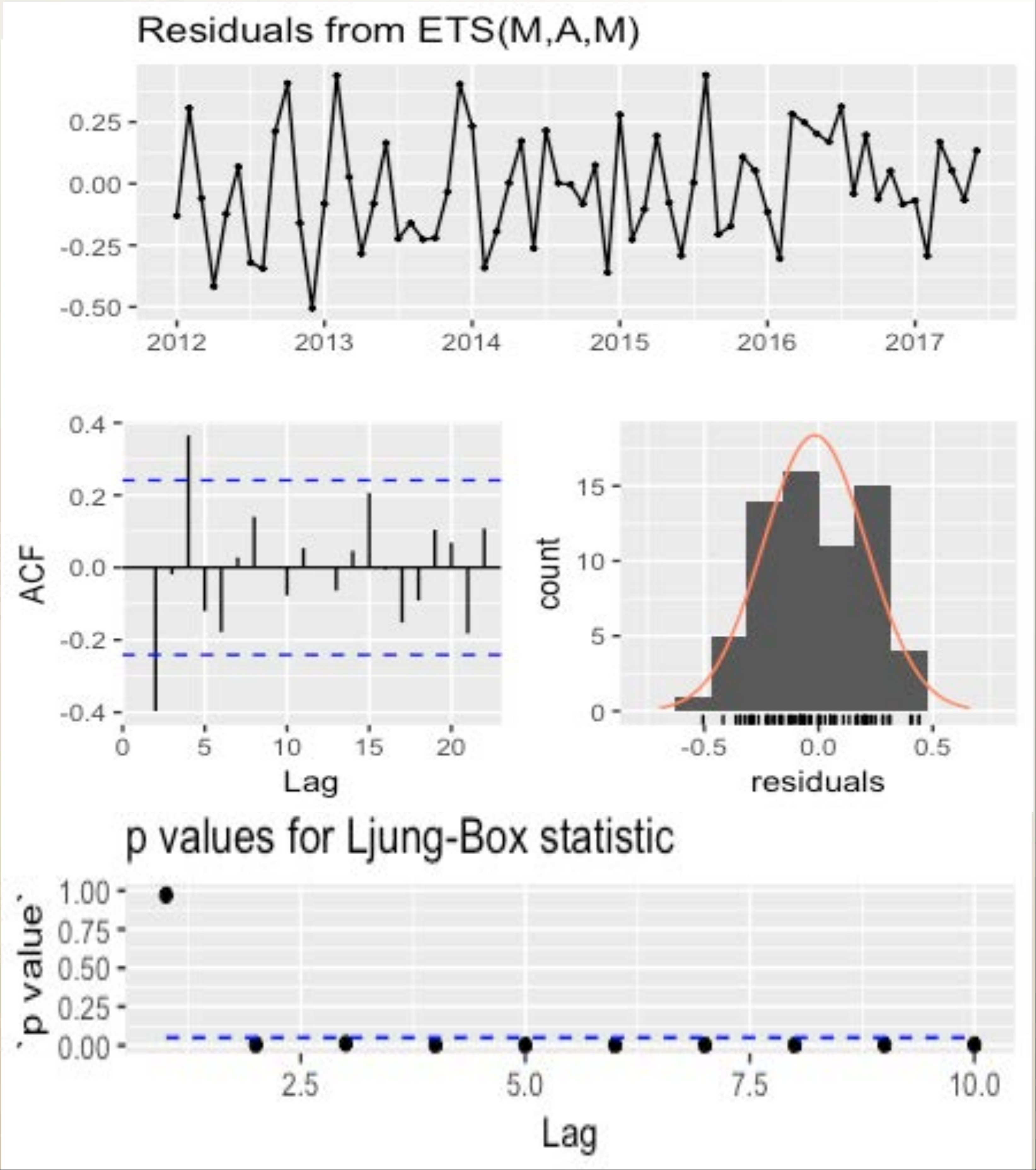
FITTING MODELS

- Monthly Maximum
- Multiplicative Holt-Winters' method with multiplicative errors

```
ETS(M,A,M)

Call:
 ets(y = tsmax)

  Smoothing parameters:
    alpha = 0.0207
    beta  = 0.003
    gamma = 1e-04

  Initial states:
    l = 275.1913
    b = -1.1773
    s = 1.6374 1.0102 0.8096 0.6099 0.4739 0.63
           0.7895 0.8737 0.931 0.9311 1.8297 1.4741

  sigma:  0.2584

     AIC     AICc      BIC
801.8906 814.6406 839.1147

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set -5.478884 61.93863 42.51398 -7.531015 21.01618 0.8140975 -0.03853985
```

FITTING MODELS

- Monthly Maximum



FITTING MODELS

- **MODEL 3: VAR MODEL**
  - Vector Autoregressions VAR(p)
  - For multivariate time series model
  - For example: VAR(1)

$$y_{i,t} = c_i + \phi_{11,1} y_{i,t-1} + \phi_{12,1} y_{i+1,t-1} + e_{i,t}$$

FITTING MODELS

- Monthly pm2.5 and weather data
- Residuals test

```
        Portmanteau Test (asymptotic)

data:  Residuals of VAR object var1
Chi-squared = 670.48, df = 324, p-value < 2.2e-16


data:  Residuals of VAR object var2
Chi-squared = 429.35, df = 288, p-value = 1.235e-07

data:  Residuals of VAR object var3
Chi-squared = 357.46, df = 252, p-value = 1.361e-05
```

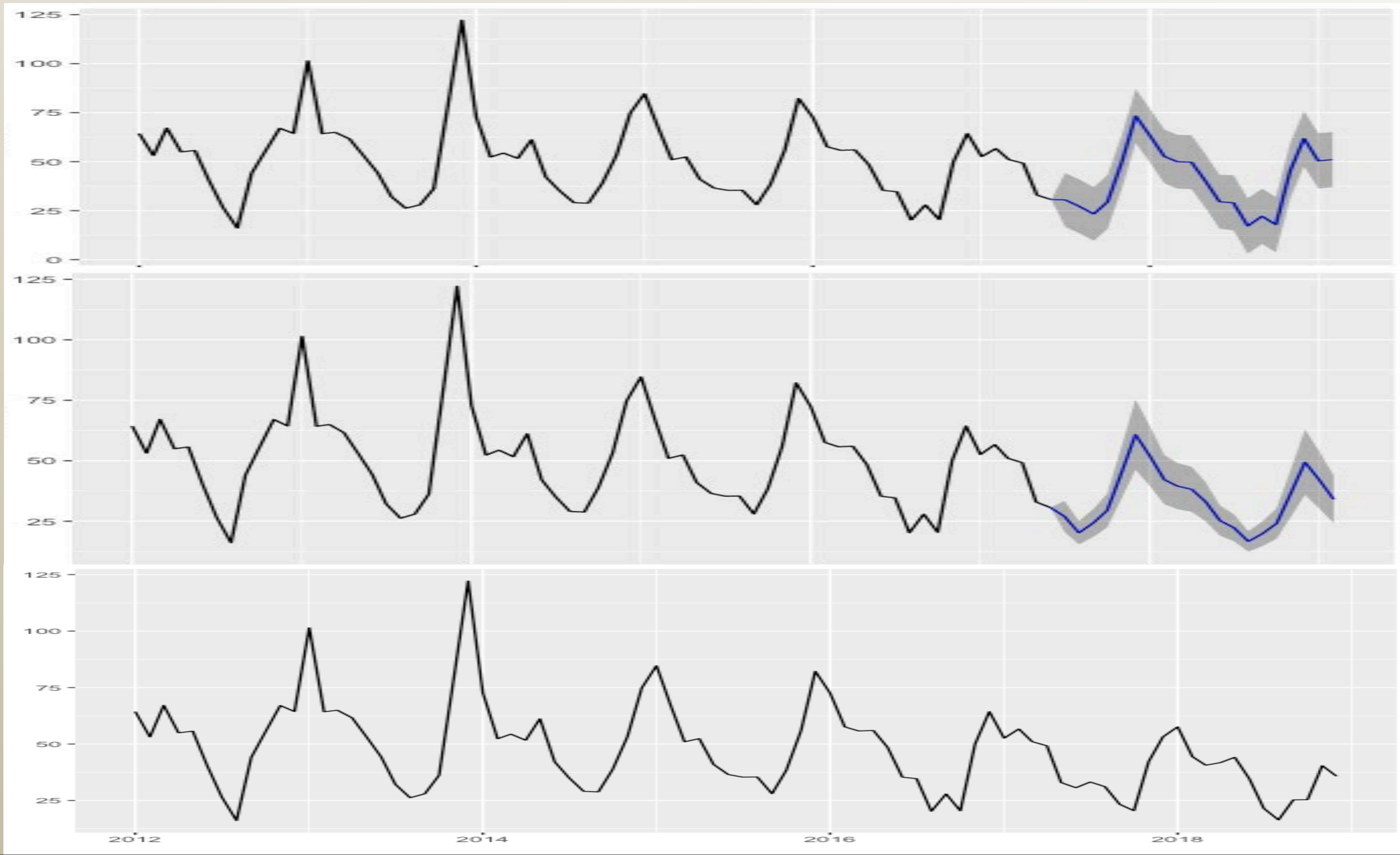
FITTING MODELS

- Monthly pm2.5 and weather data

```
Estimation results for equation pm2.5:
======================================
pm2.5 = pm2.5.l1 + T.l1 + H.l1 + PP.l1 + VV.l1 + V.l1 + pm2.5.l2 + T.l2 + H.l2 +
PP.l2 + VV.l2 + V.l2 + const

         Estimate Std. Error t value Pr(>|t|)
pm2.5.l1  0.34284    0.03727   9.199  < 2e-16 ***
T.l1     -0.37807    0.25625  -1.475  0.14039
H.l1     -0.17543    0.07400  -2.371  0.01793 *
PP.l1    -0.18835    0.05342  -3.526  0.00044 ***
VV.l1    -1.59028    0.15095 -10.535  < 2e-16 ***
V.l1     -1.30059    0.21922  -5.933 4.00e-09 ***
pm2.5.l2 -0.01613    0.03580  -0.451  0.65230
T.l2      0.01554    0.25828   0.060  0.95203
H.l2      0.05091    0.06969   0.731  0.46523
PP.l2     0.04163    0.05387   0.773  0.43979
VV.l2     0.61466    0.15704   3.914 9.64e-05 ***
V.l2      0.55917    0.22474   2.488  0.01300 *
const    62.38183    7.38870   8.443  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 18.83 on 1081 degrees of freedom
Multiple R-Squared: 0.4853,     Adjusted R-squared: 0.4796
F-statistic: 84.94 on 12 and 1081 DF,  p-value: < 2.2e-16
```

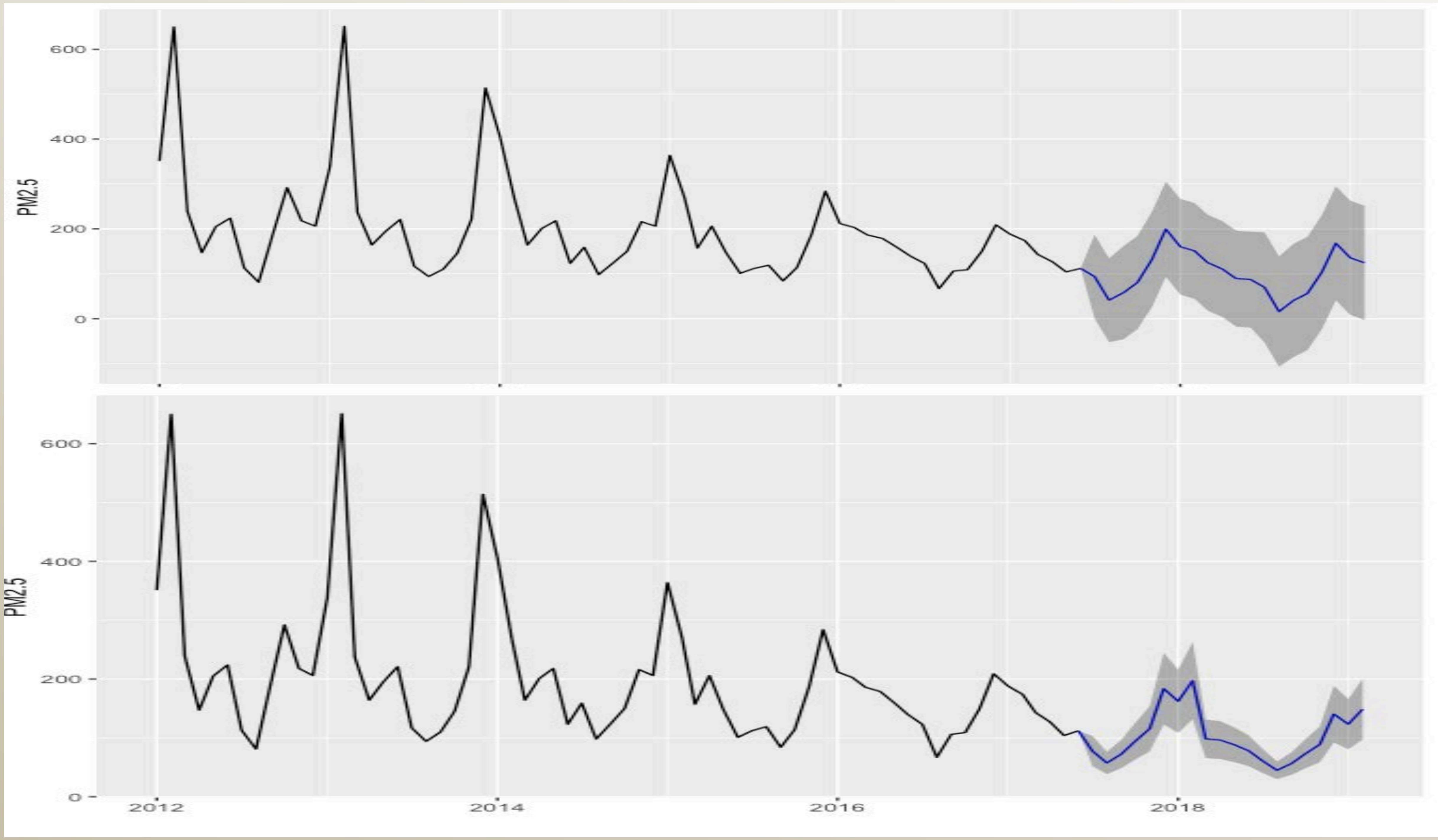
FORECAST

➤ Monthly Average



FORECAST

- Monthly Maximum



SUGGESTION

REASONS FOR POOR AIR QUALITY IN WINTER

- Shanghai gets wind mainly from South East during summer months, while winds from the North dominate during winter months.
- In winter, the Northern people use more coal-burning boilers for central heating systems, so northern winds carry dirty emissions south.
- Trees density decreases in Shanghai because of losing leaves in late Autumn which can no longer trap dust.
- Straw burning in the provinces nearby during October to December

ACTIONS TAKEN TO IMPROVE SHANGHAI AIR QUALITY

- Shanghai Clean Air Action Plan 2013-2017
  - On October 18th, 2013, the local Shanghai government unveiled this comprehensive plan.
  - Aiming to reduce the concentration of PM2.5 by 20 percent in five years.
  - The first regulation in China to prevent and control volatile compounds.
  - The new plan increases punishments for factories and vehicles and restaurants which emit excessive pollutants.

ACTIONS TAKEN TO IMPROVE SHANGHAI AIR QUALITY

- In late November 2013, Shanghai announced rules for a carbon emissions trading scheme. (industrial emissions)
- In early January 2014, Shanghai launched a joint effort rule with its three closest provinces (Zhejiang, Jiangsu and Anhui), to tackle air pollution. (outside shanghai)
- In late January 2014, Shanghai announced a ban on the burning of straw and other bonfires within all of Shanghai. (previous certain areas) (agricultural)

ACTIONS TAKEN TO IMPROVE SHANGHAI AIR QUALITY

- Shanghai government extended subsidies for renewable energy 'green cars'. (Buyer will get a subsidy of RMB 40,000, plus a free Shanghai license plate, worth about RMB 70,000, and also get the central government subsidy of RMB 60,000.) (cars and ships)
- In April 2014, Shanghai adopted the V emission standards for all new vehicles. (cars and ships)
- Shanghai tightened its ban on Yellow Label vehicles from outer ring roads (previous inner ring roads) and a complete ban will come in force in 2015. (cars and ships)

YEAR 2015

- Shanghai environmental protection plan for 2015 to 2017
  - The government intends to invest 100 billion RMB on more than 200 projects to reduce pollution.
  - The goal is to reduce the average PM2.5 concentration to 48 μg/m^3 by the end of 2017.
  - Method mainly is by upgrading companies' facilities to reduce carbon emissions.

YEAR 2016

- Trucks that don't meet IV emission standard will not be allowed during the daytime.
- Shanghai banned fireworks within the Outer Ring Road. This ban was strictly enforced during the 2016 Chinese New Year celebrations.
- A new goal in the development blueprint for Shanghai is reducing Shanghai PM2.5 yearly average to 42 ug/m3 by 2020.

TO SUMMARIZE

We still got a lot to explore!

FOR NEXT STAGE, WE ARE GOING TO:

- Try other methods to analyze the trends, correlations
- Try more methods dealing with weather data
- …

REFERENCE

- Ole Raaschou-Nielsen; et al. (July 10, 2013). "Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)". The Lancet Oncology. 14 (9): 813–22. doi:10.1016/S1470-2045(13)70279-1.
- Ghassan B. Hamra,1 Neela Guha,1 Aaron Cohen; et al. (September 2014). "Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis". Environmental Health Perspectives. 122 (9): 906–11. doi:10.1289/ehp.1408092.
- Hyndman, Rob J. & Athanasopoulos, George. & OTexts.com, issuing body. (2014). Forecasting : principles and practice. [Heathmont?, Victoria] : OTexts.com

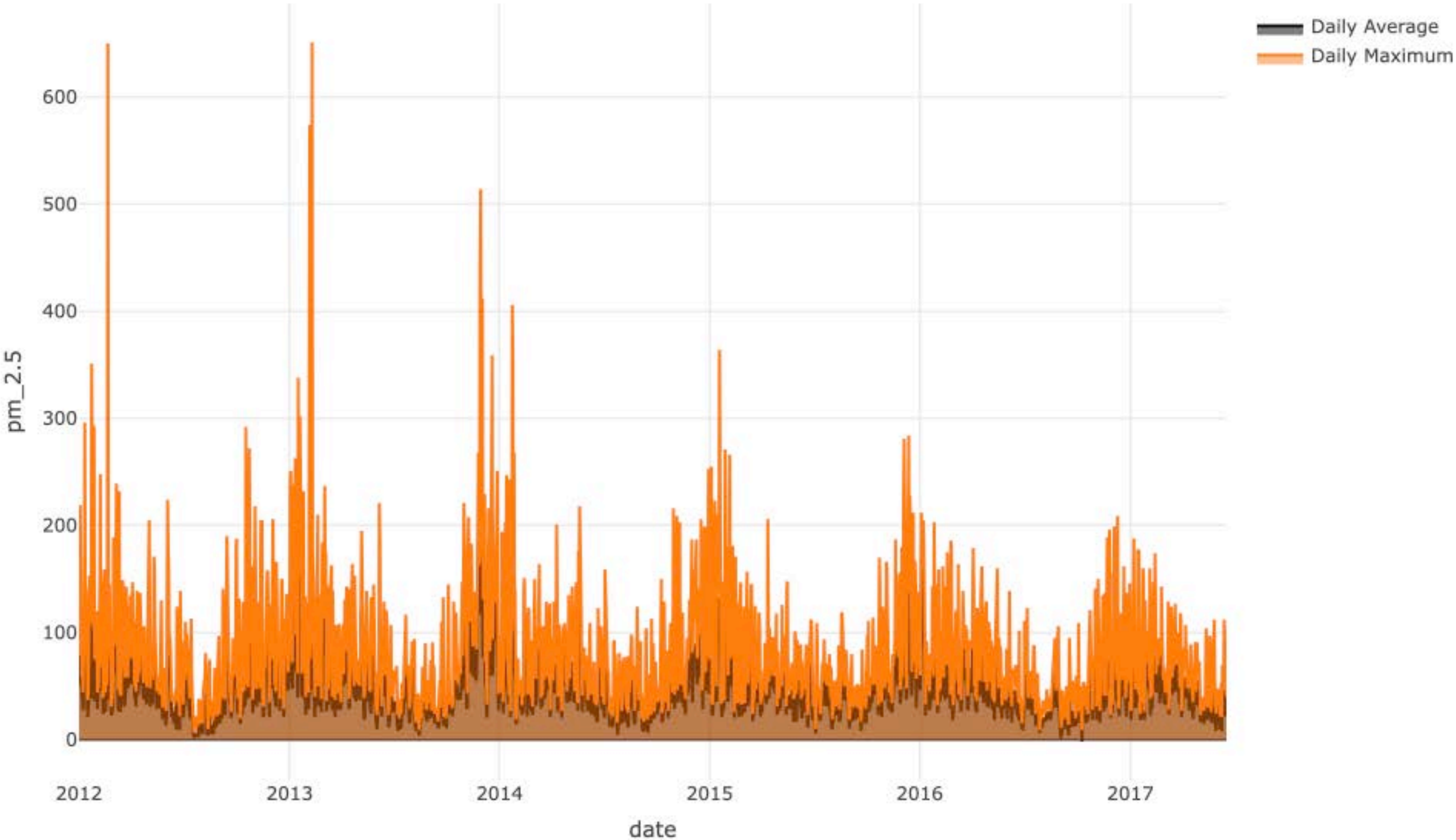
Have a nice spring break!

PM2.5 SHANGHAI

ZHU WANG (王祝)
WEI ZHANG (张薇)

STATISTICAL ANALYSIS

TRENDS IN RECENT YEARS

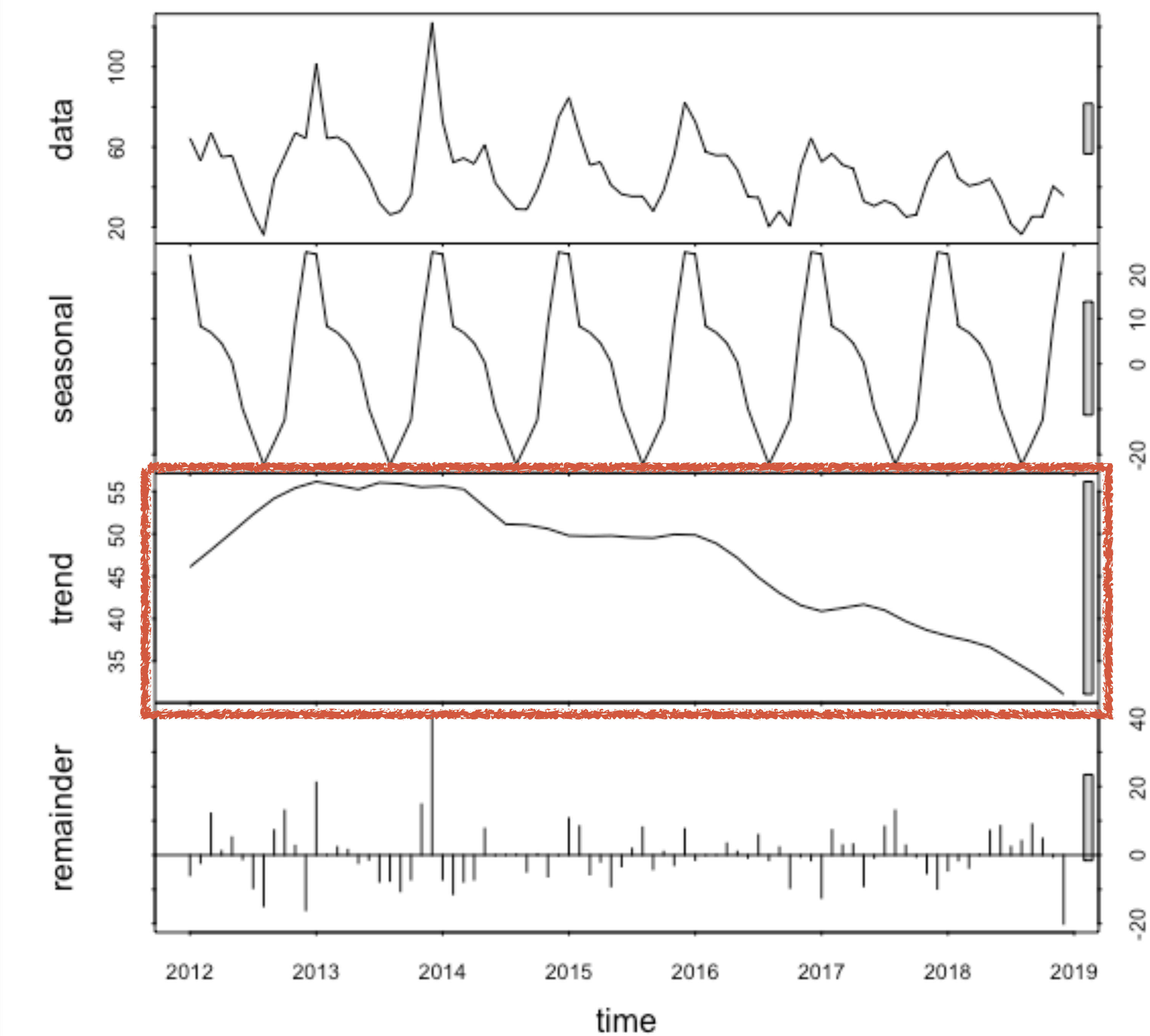


TRENDS ANALYSIS

- **Linear Model with Time Series Components**
  - $y_t = S_t + T_t + R_t$
  - y_t is the data
  - S_t the seasonal component
  - T_t is the trend-cycle component
  - R_t is the remainder component

TRENDS ANALYSIS

- Monthly Average
  - Two-sample t-test
  - Fit a linear model using months and trend components



TRENDS ANALYSIS

- Monthly Average
  - Two-sample t-test

```
        Welch Two Sample t-test

data:  comb3$Y2013 and comb3$Y2018
t = 2.577, df = 14.296, p-value = 0.02166
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  4.036098 43.636839
sample estimates:
mean of x mean of y
 59.48023  35.64376


        Welch Two Sample t-test

data:  comb3$Y2014 and comb3$Y2018
t = 2.4919, df = 20.559, p-value = 0.02134
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  2.278282 25.445876
sample estimates:
mean of x mean of y
 49.50584  35.64376
```

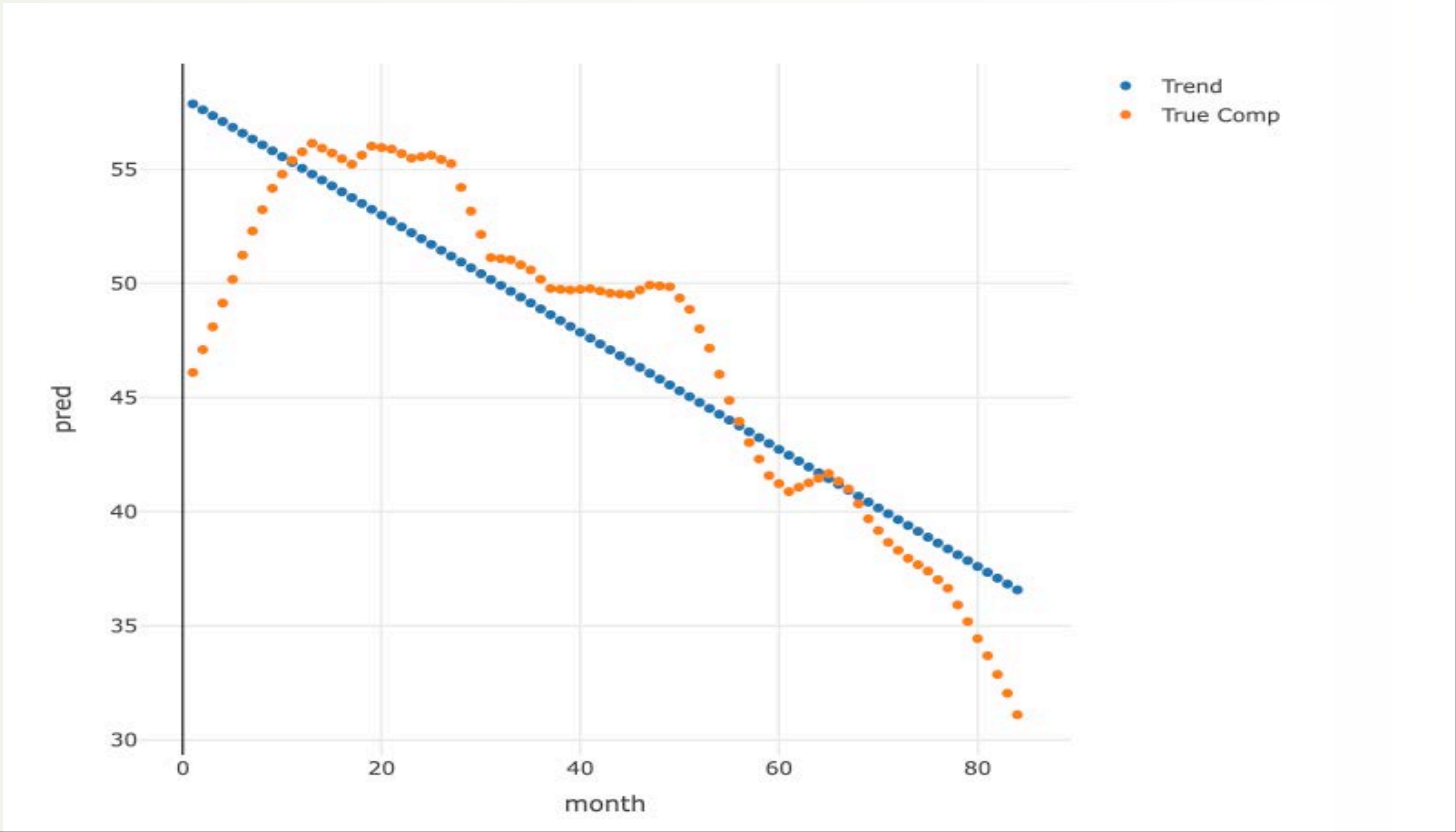
TRENDS ANALYSIS

- Monthly Average
  - Fit a linear model using months and trend components

```
lm(formula = trendmon ~ month, data = trendmon)

Residuals:
     Min       1Q   Median       3Q      Max
-11.7661  -1.4674   0.9118   2.4802   4.3040

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 58.12926    0.74520   78.00   <2e-16 ***
month       -0.25655    0.01523  -16.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.384 on 82 degrees of freedom
Multiple R-squared:  0.7758,    Adjusted R-squared:  0.7731
F-statistic: 283.8 on 1 and 82 DF,  p-value: < 2.2e-16
```

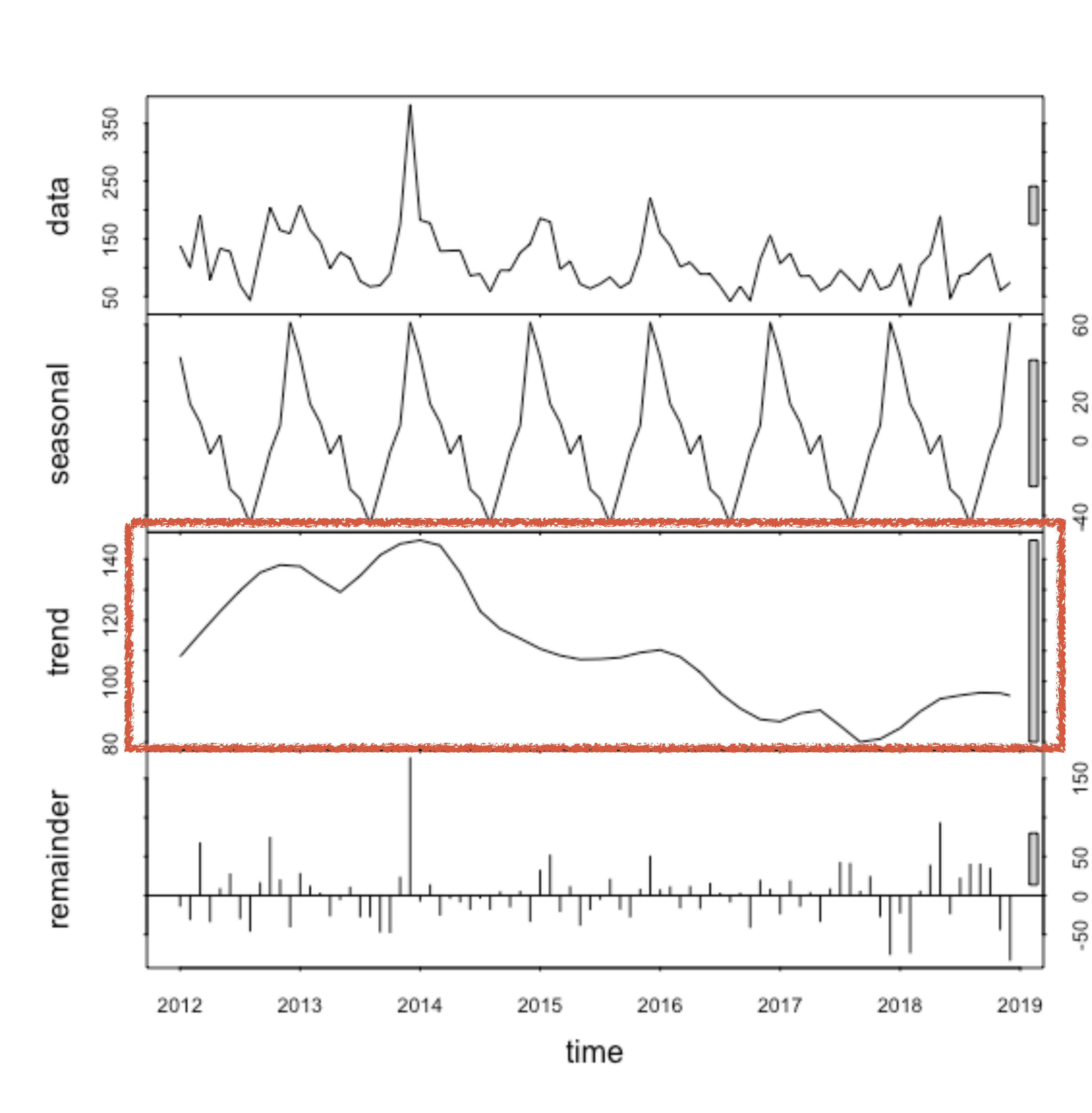
TRENDS ANALYSIS

- Monthly Average
  - Fit a linear model using months and trend components



TRENDS ANALYSIS

- Monthly Maximum
  - Two-sample t-test
  - Fit a linear model using months and trend components



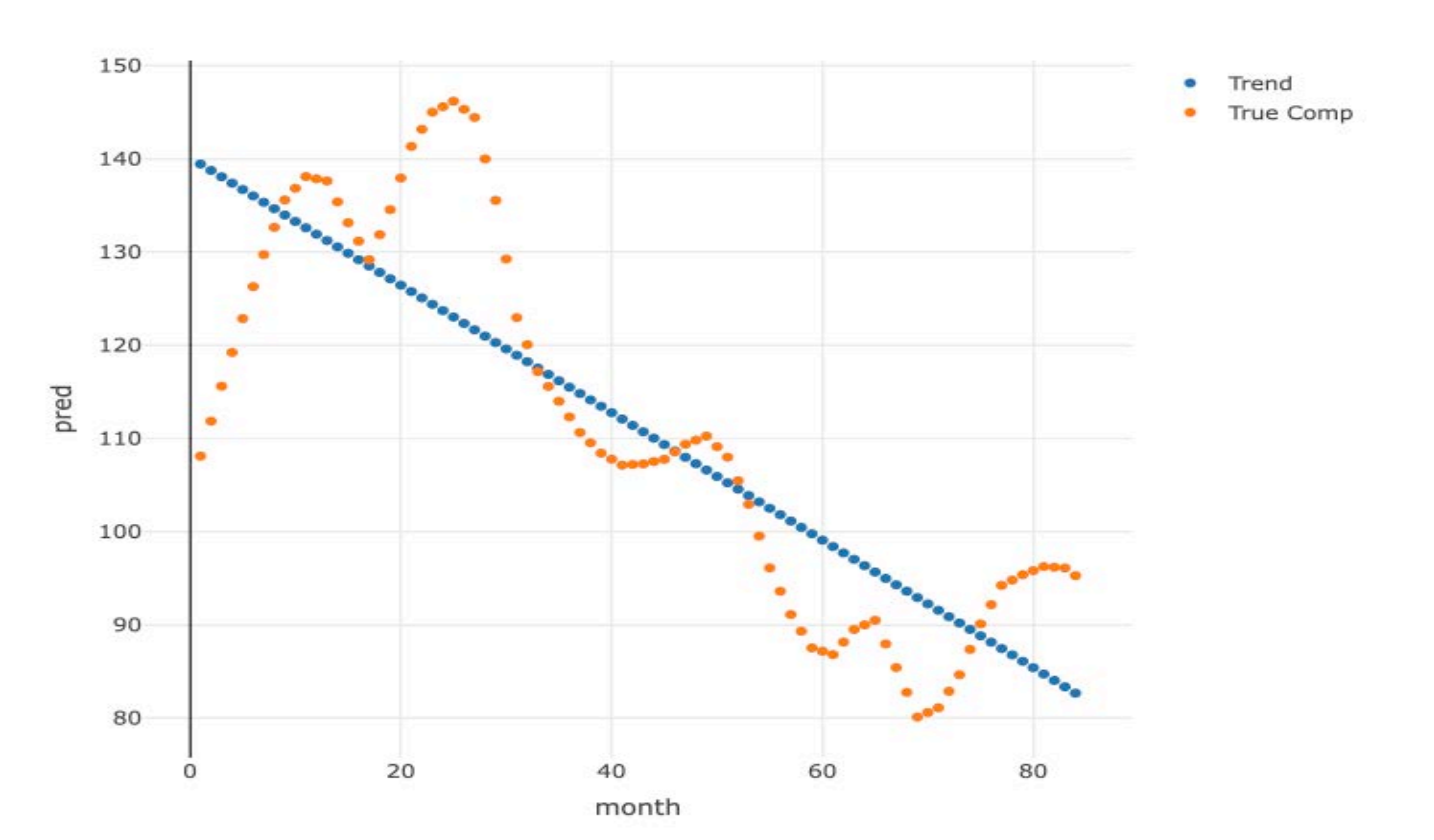
TRENDS ANALYSIS

- Monthly Maximum
  - Fit a linear model using months and trend components

```
lm(formula = trendmon ~ month, data = trendmon)

Residuals:
     Min       1Q   Median       3Q      Max 
-31.3389  -6.5586  -0.6639   6.0561  23.1572 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 140.13382    2.41539   58.02   <2e-16 ***
month        -0.68401    0.04936  -13.86   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.97 on 82 degrees of freedom
Multiple R-squared:  0.7007,	Adjusted R-squared:  0.6971 
F-statistic:   192 on 1 and 82 DF,  p-value: < 2.2e-16
```

TRENDS ANALYSIS

- Monthly Maximum
  - Fit a linear model using months and trend components



FITTING MODELS

- **MODEL 4: Generalized Additive Models**
  - $y_t = S_t + T_t + R_t$
  - $y = \beta_0 + f_{seasonal}\left(x_1\right) + f_{trend}\left(x_2\right) + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2 \mathbf{\Lambda}\right)$

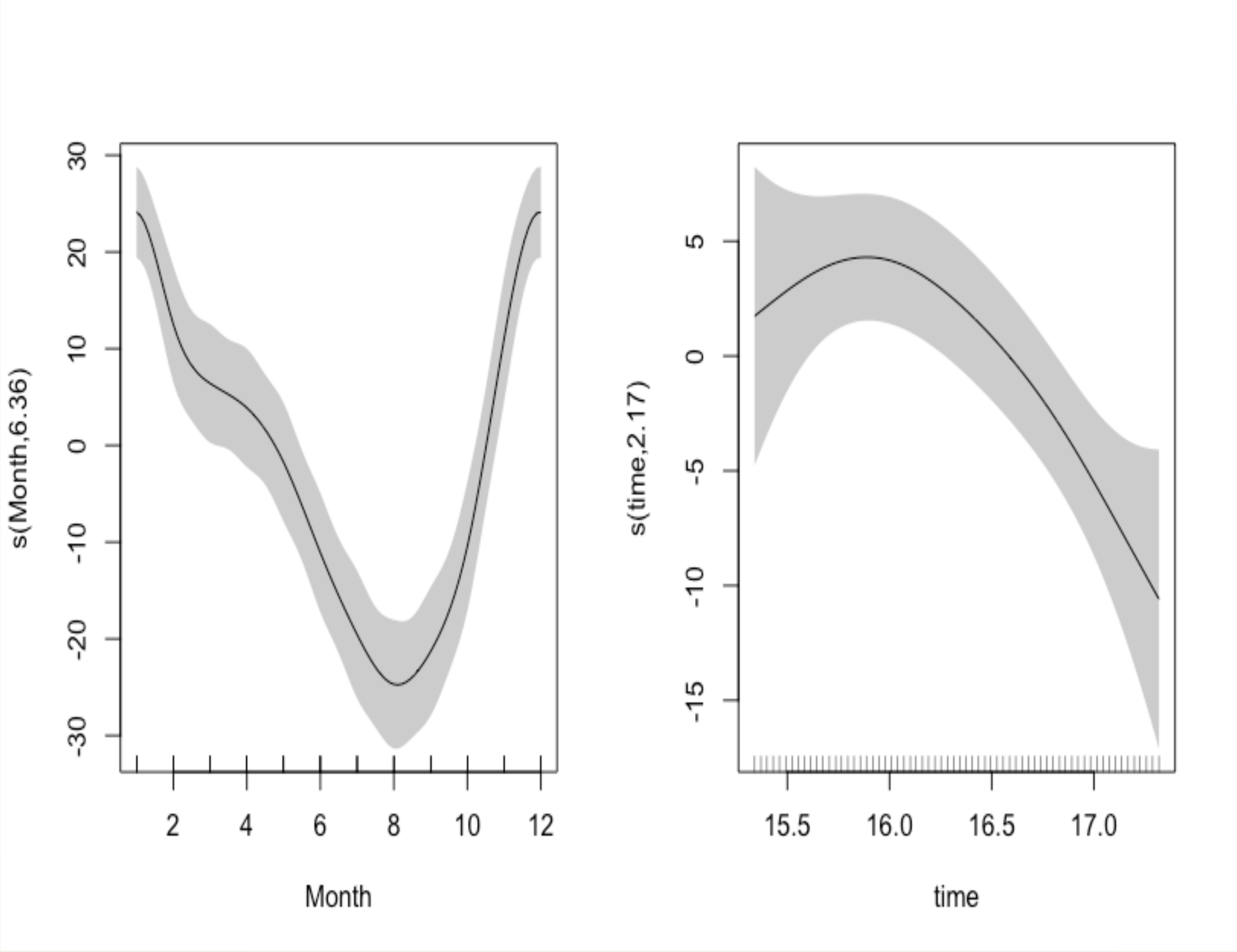
FITTING MODELS

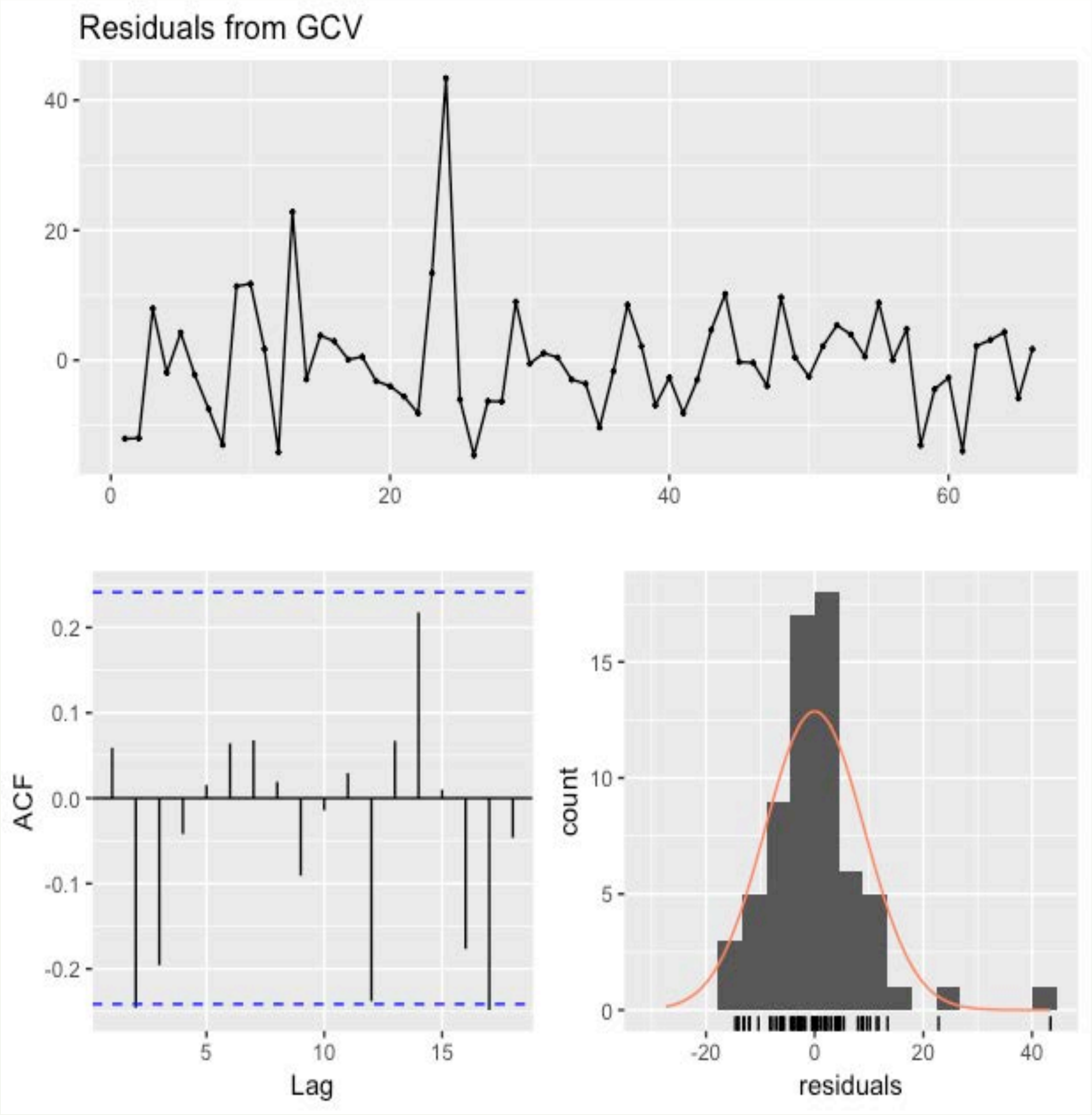
- Monthly Average

```
Family: gaussian
Link function: identity

Formula:
pm_2.5mon ~ s(Month, bs = "cc", k = 12) + s(time)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   50.615      1.206   41.96   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
           edf Ref.df      F p-value
s(Month) 6.357 10.000 18.001 < 2e-16 ***
s(time)  2.170  2.707  5.207 0.00365 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.745   Deviance explained = 77.9%
GCV = 112.24  Scale est. = 96.038    n = 66
```

FITTING MODELS

➤ Monthly Average





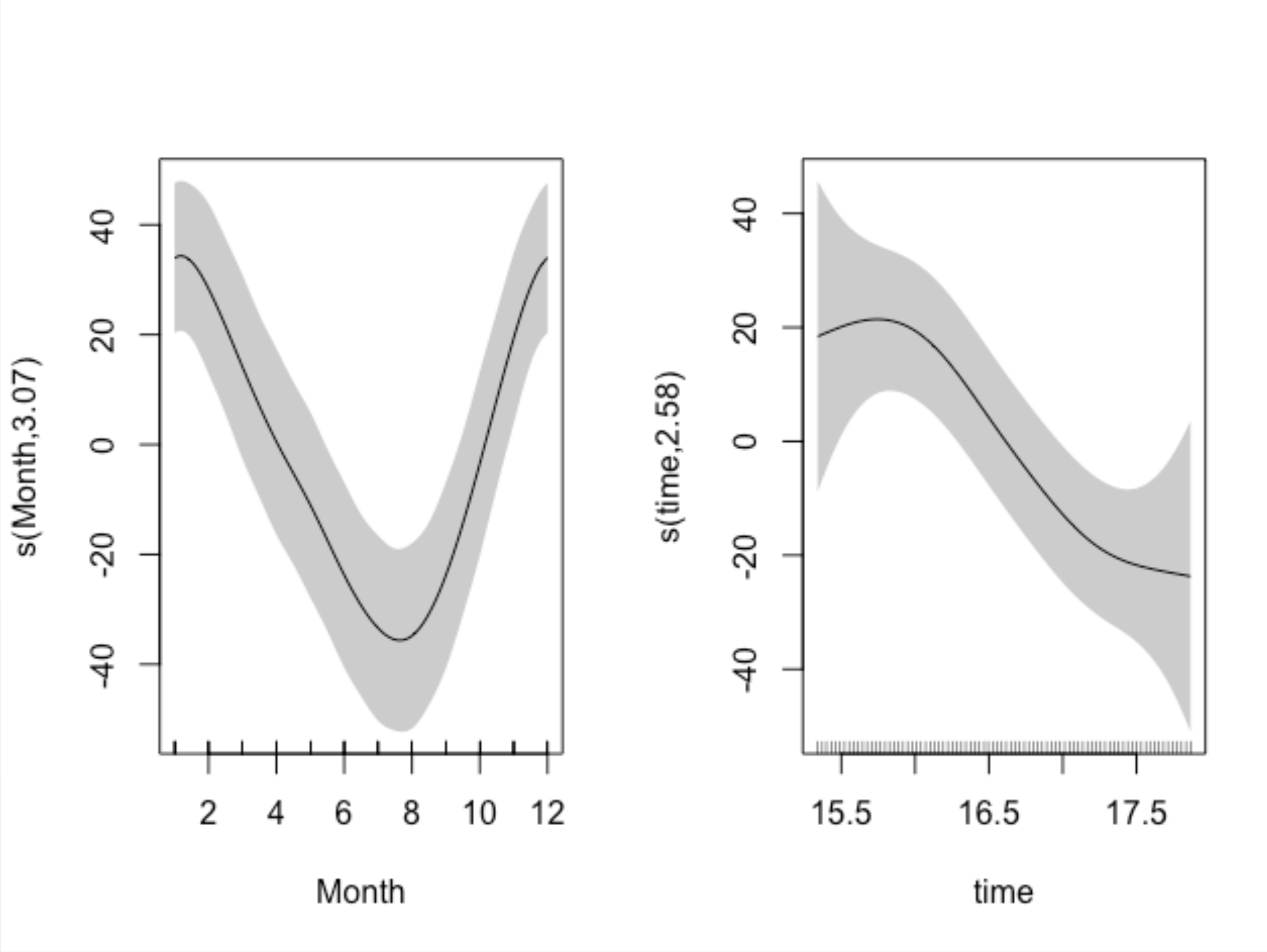
FITTING MODELS

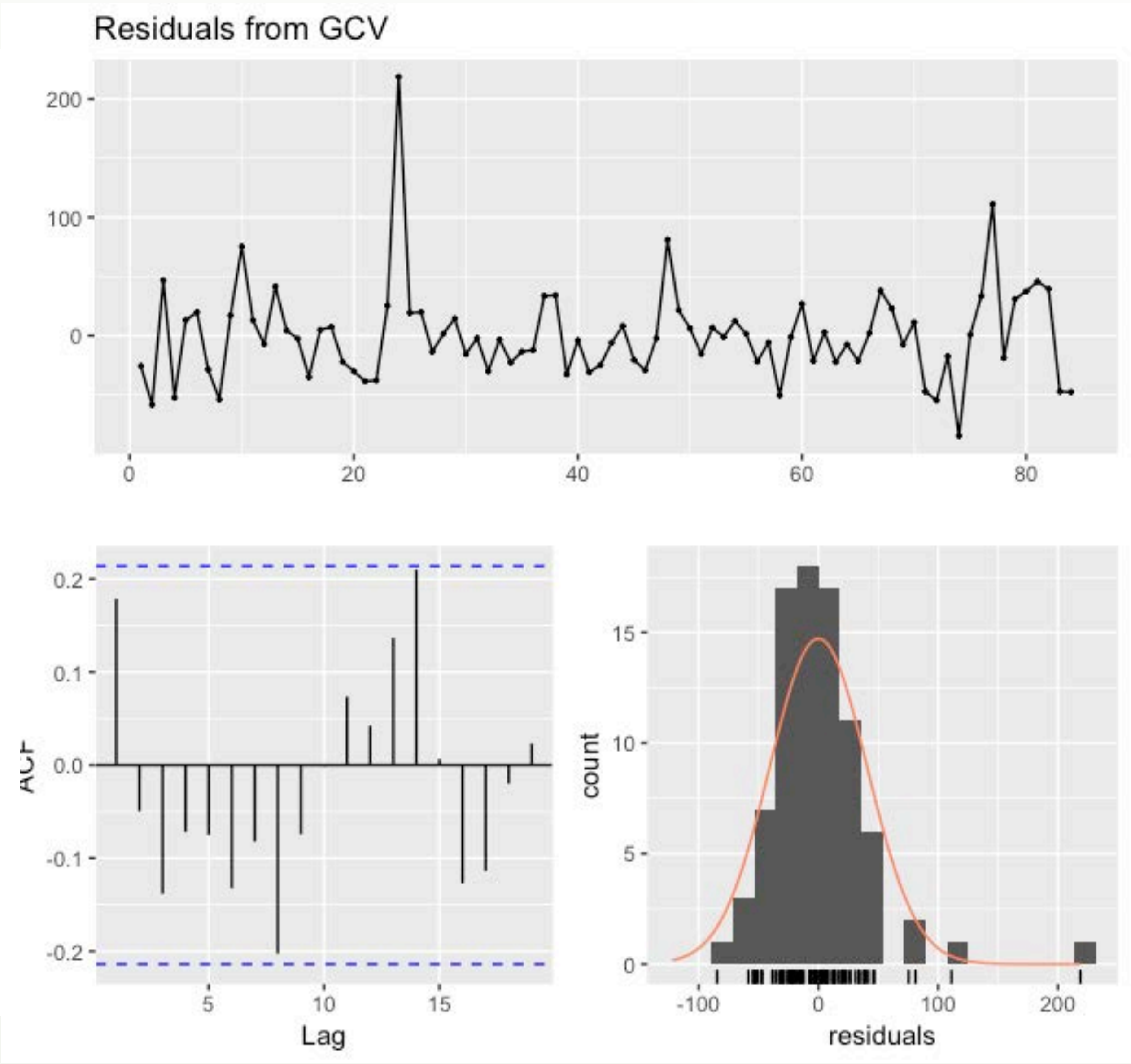
- Monthly Maximum

```
Family: gaussian
Link function: identity

Formula:
pm2.5max ~ s(Month, bs = "cc", k = 12) + s(time)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  111.432      4.591   24.27   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
           edf Ref.df     F  p-value
s(Month) 3.068  10.00 3.228 5.59e-07 ***
s(time)  2.579   3.21 4.630  0.00395 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.363   Deviance explained = 40.6%
GCV = 1922.4  Scale est. = 1770.3    n = 84
```

FITTING MODELS

- Monthly Maximum





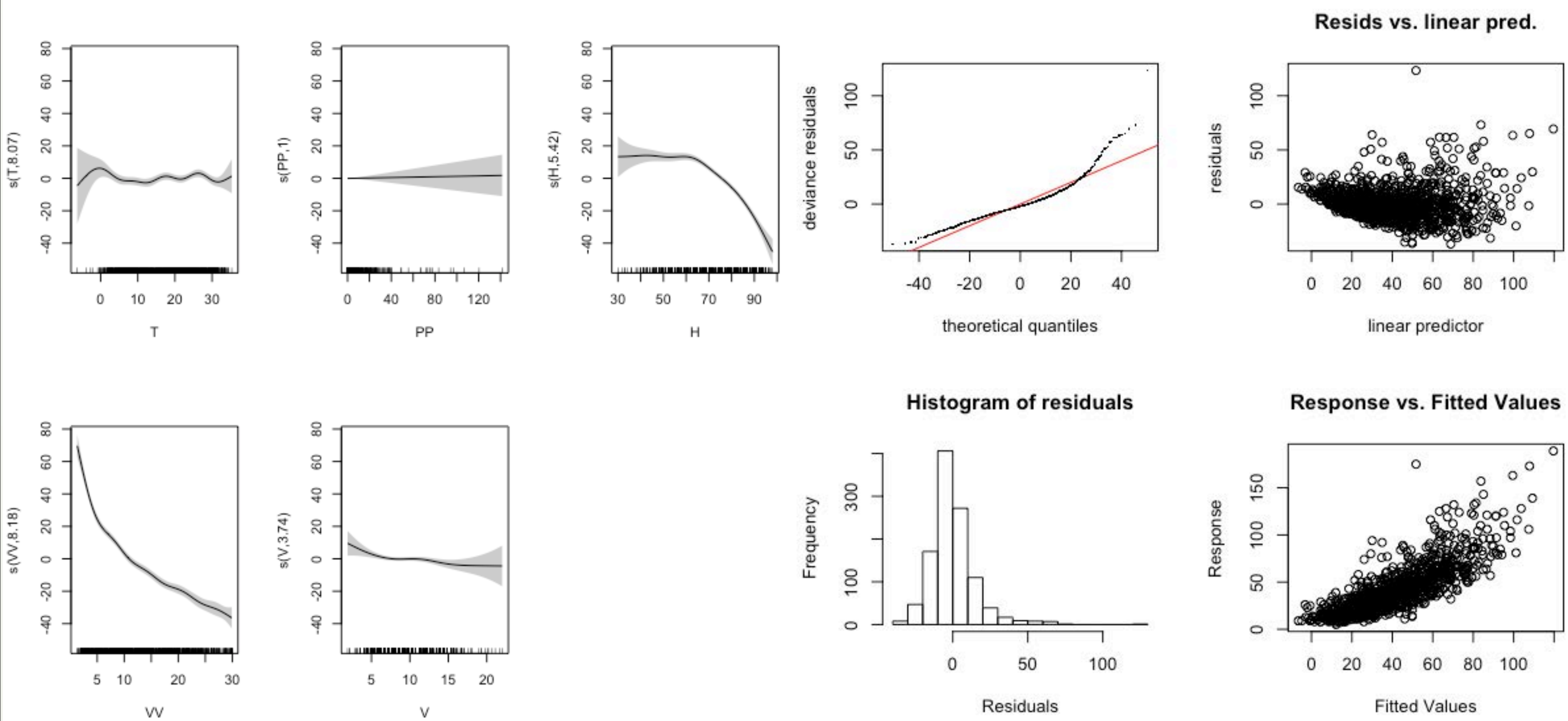
FITTING MODELS

- **Daily pm2.5 and weather**
  - $y = \beta_0 + f_{Temp}(x_1) + f_{Humidity}(x_2) + f_{Rainfall}(x_3) + f_{Visibility}(x_4) + f_{Wind}(x_5) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda})$

```
Family: gaussian
Link function: identity

Formula:
pm2.5 ~ s(T) + s(PP) + s(H) + s(VV) + s(V)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   39.418      0.459   85.89   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
        edf Ref.df       F p-value
s(T)  8.073  8.767   2.076  0.0480 *
s(PP) 1.000  1.000   0.077  0.7812
s(H)  5.422  6.567  80.088  <2e-16 ***
s(VV) 8.180  8.813 164.852  <2e-16 ***
s(V)  3.742  4.683   2.930  0.0151 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.662   Deviance explained =   67%
GCV = 236.79  Scale est. = 230.86    n = 1096
```

FITTING MODELS

- Daily pm2.5 and weather



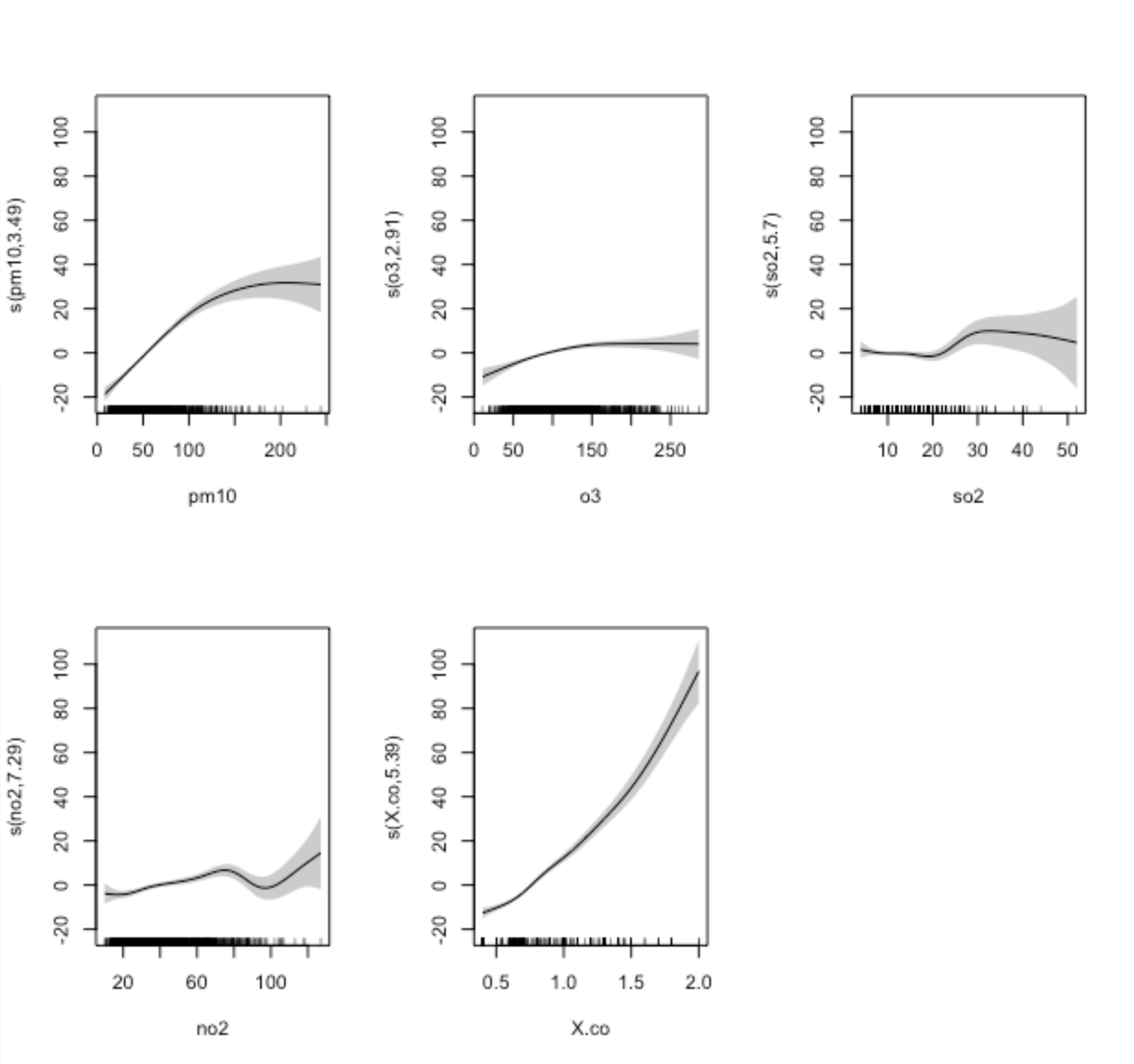
FITTING MODELS

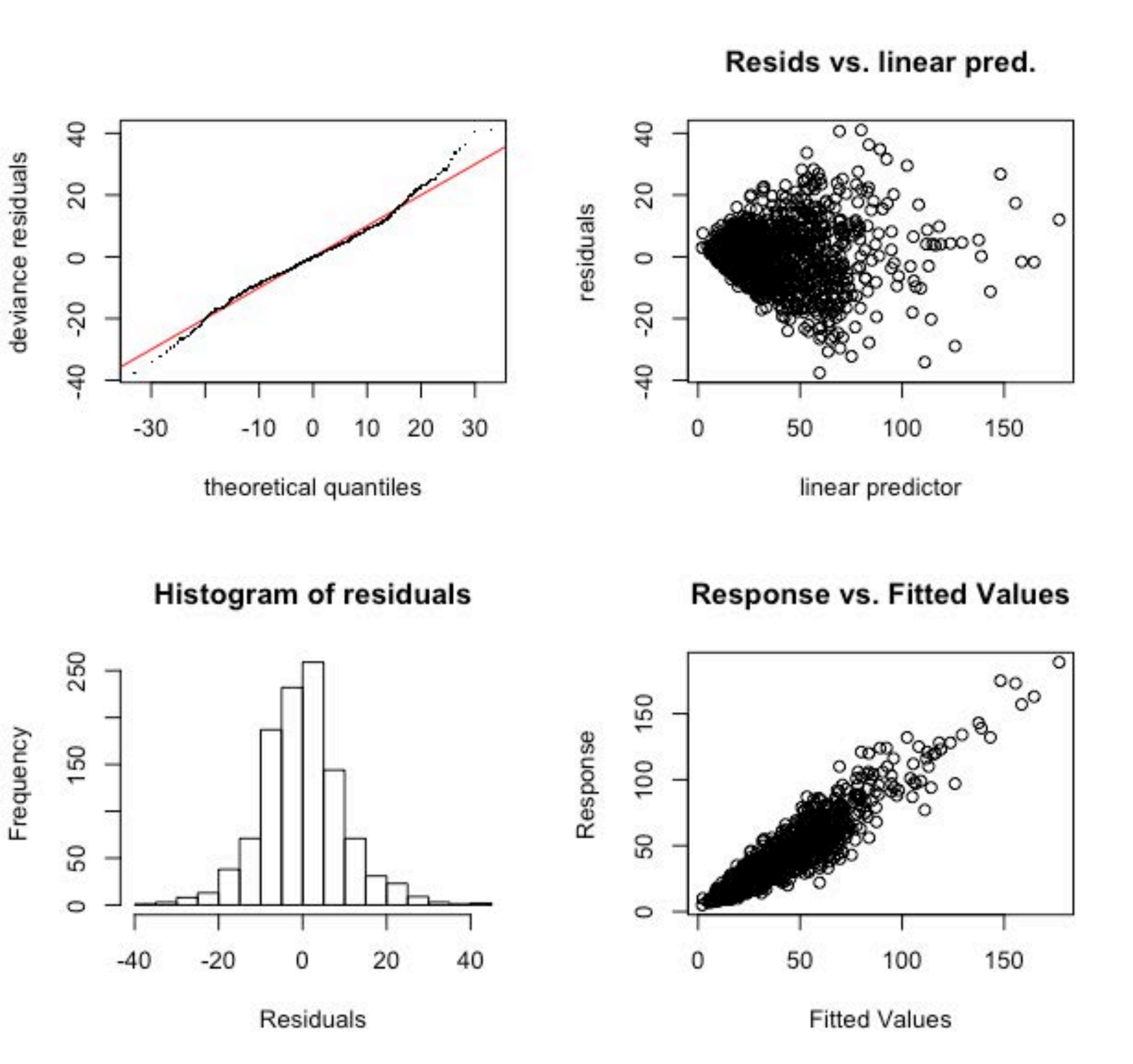
- Daily pm2.5 and other pollutions

```
Family: gaussian
Link function: identity

Formula:
pm2.5 ~ s(pm10) + s(o3) + s(so2) + s(no2) + s(X.co)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  39.4183     0.3012   130.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
          edf Ref.df      F  p-value
s(pm10) 3.486  4.400 81.802  < 2e-16 ***
s(o3)   2.912  3.700 16.973 2.52e-12 ***
s(so2)  5.701  6.745  2.442    0.016 *
s(no2)  7.288  8.244  5.900 1.43e-07 ***
s(X.co) 5.388  6.514 87.409  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.855   Deviance explained = 85.8%
GCV = 101.85  Scale est. = 99.459    n = 1096
```

FITTING MODELS

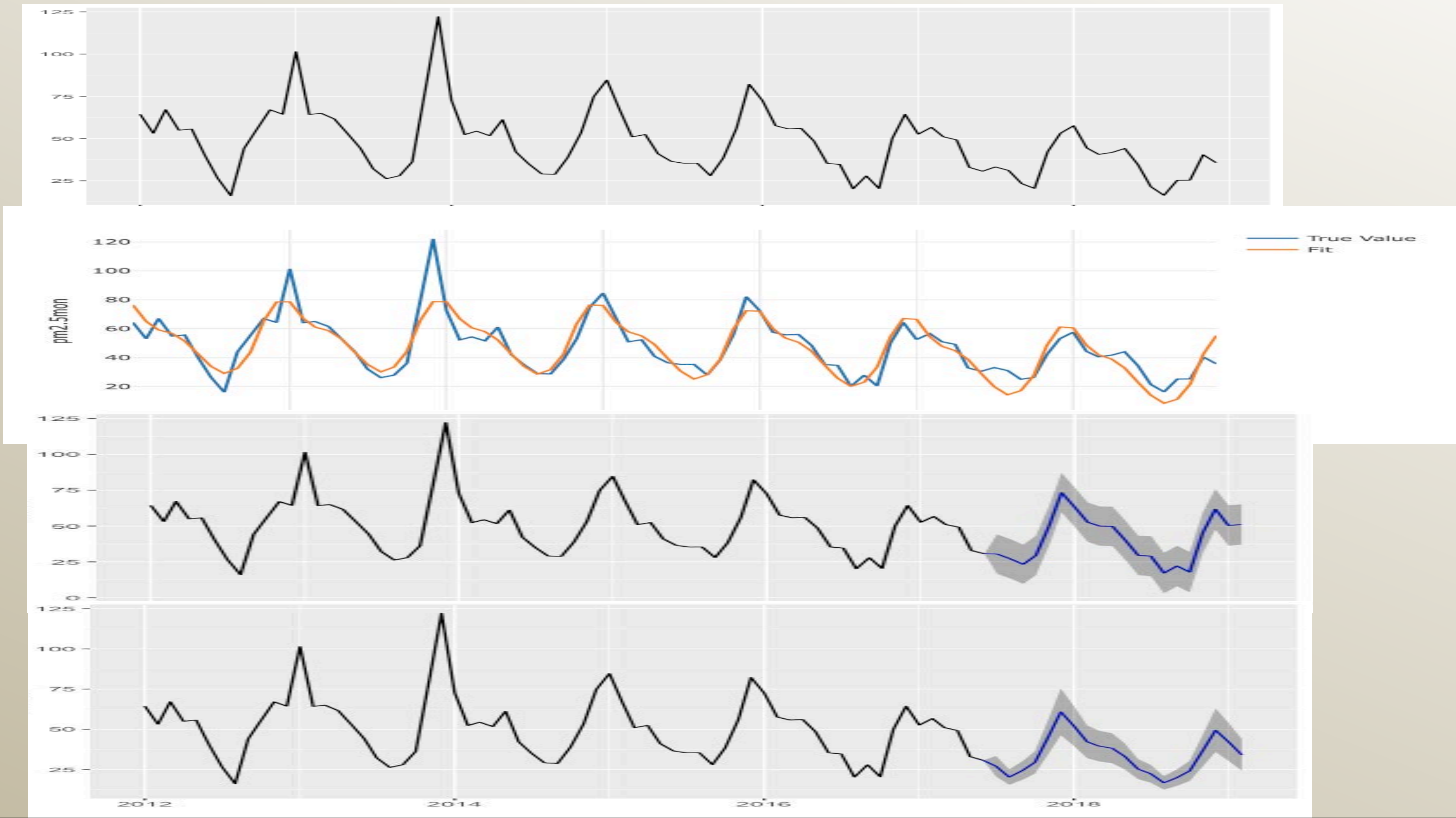
- Daily pm2.5 and other pollutions





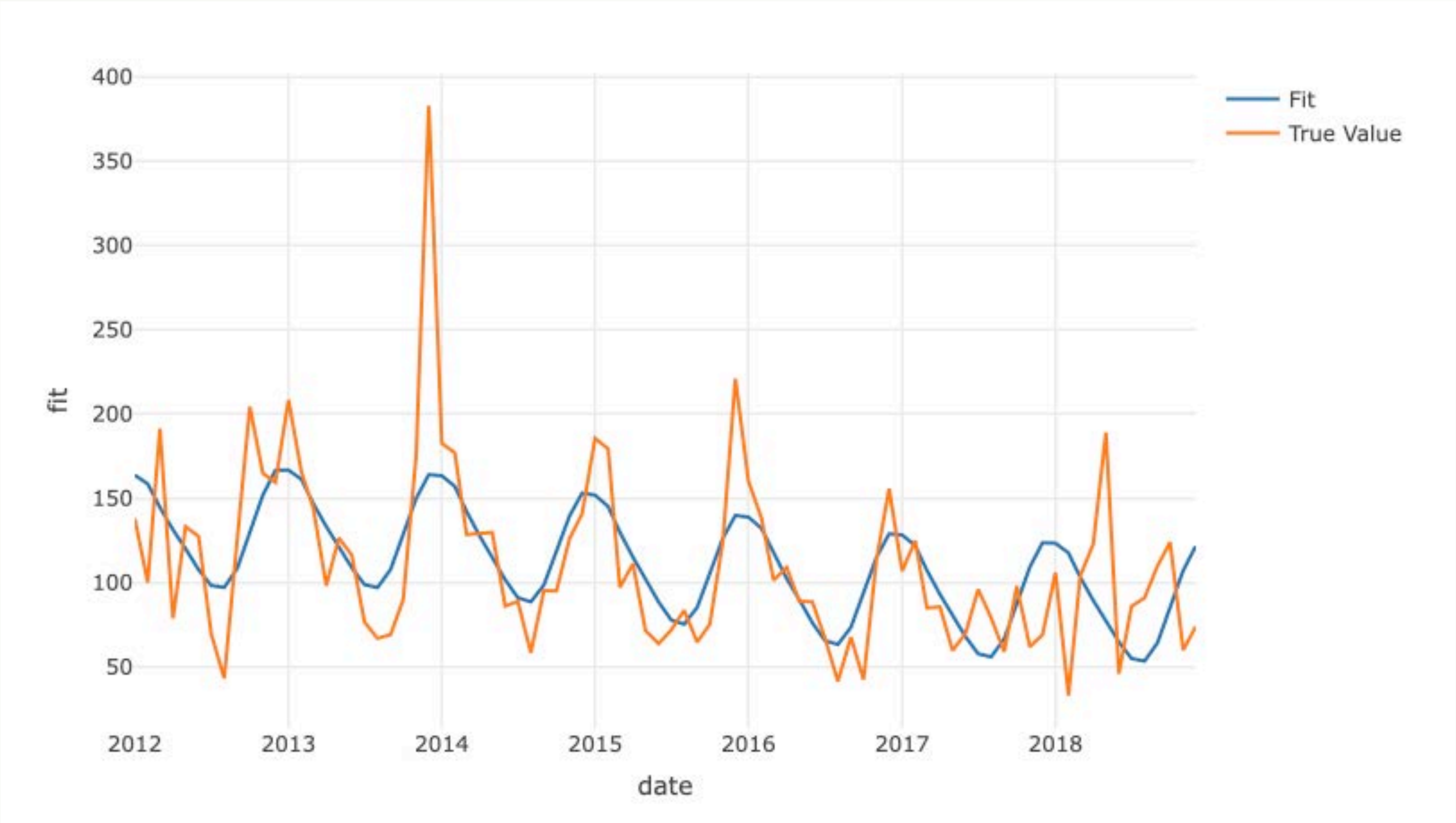
FORECAST

➤ Monthly Average



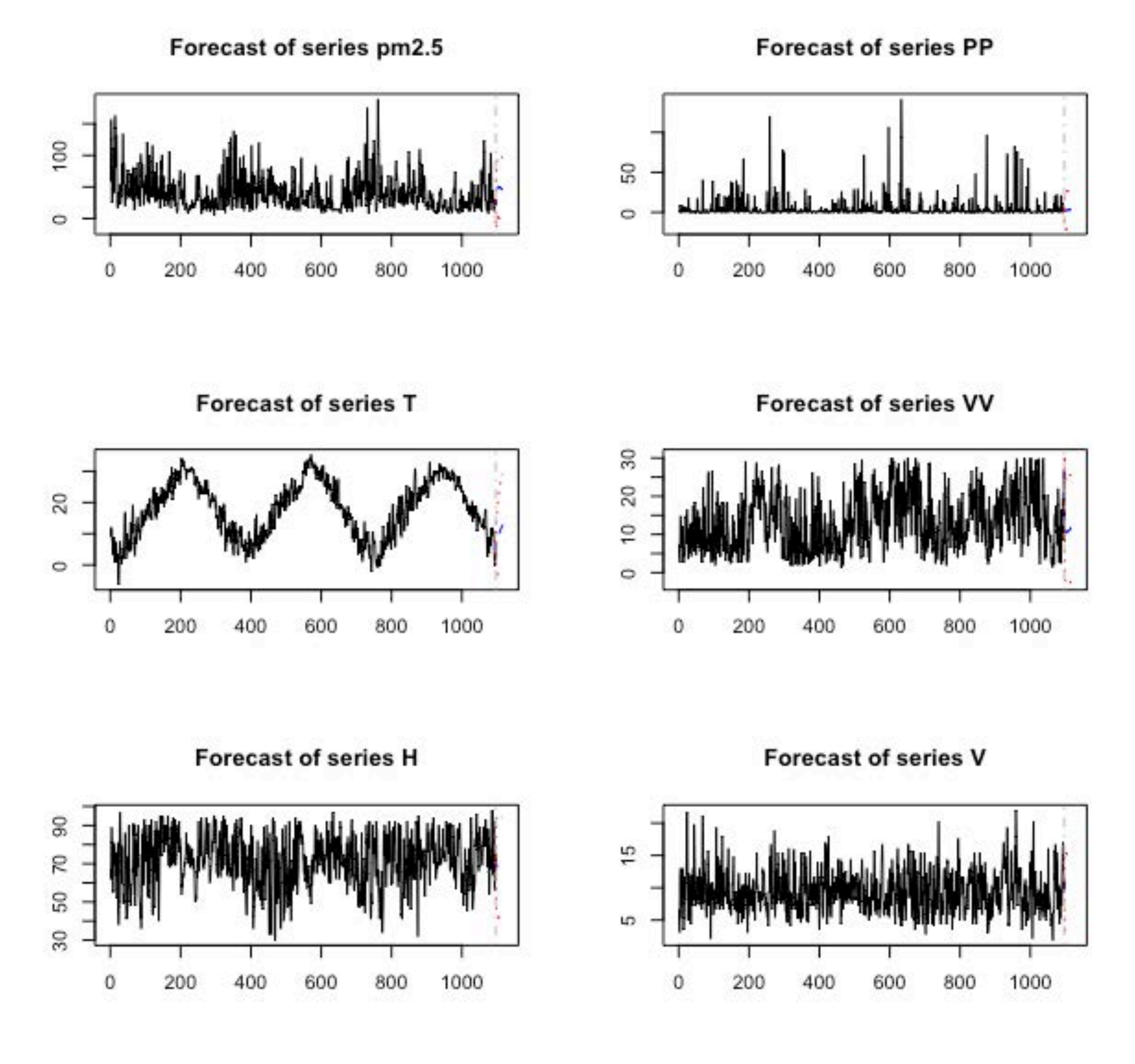
FORECAST

- Monthly Maximum



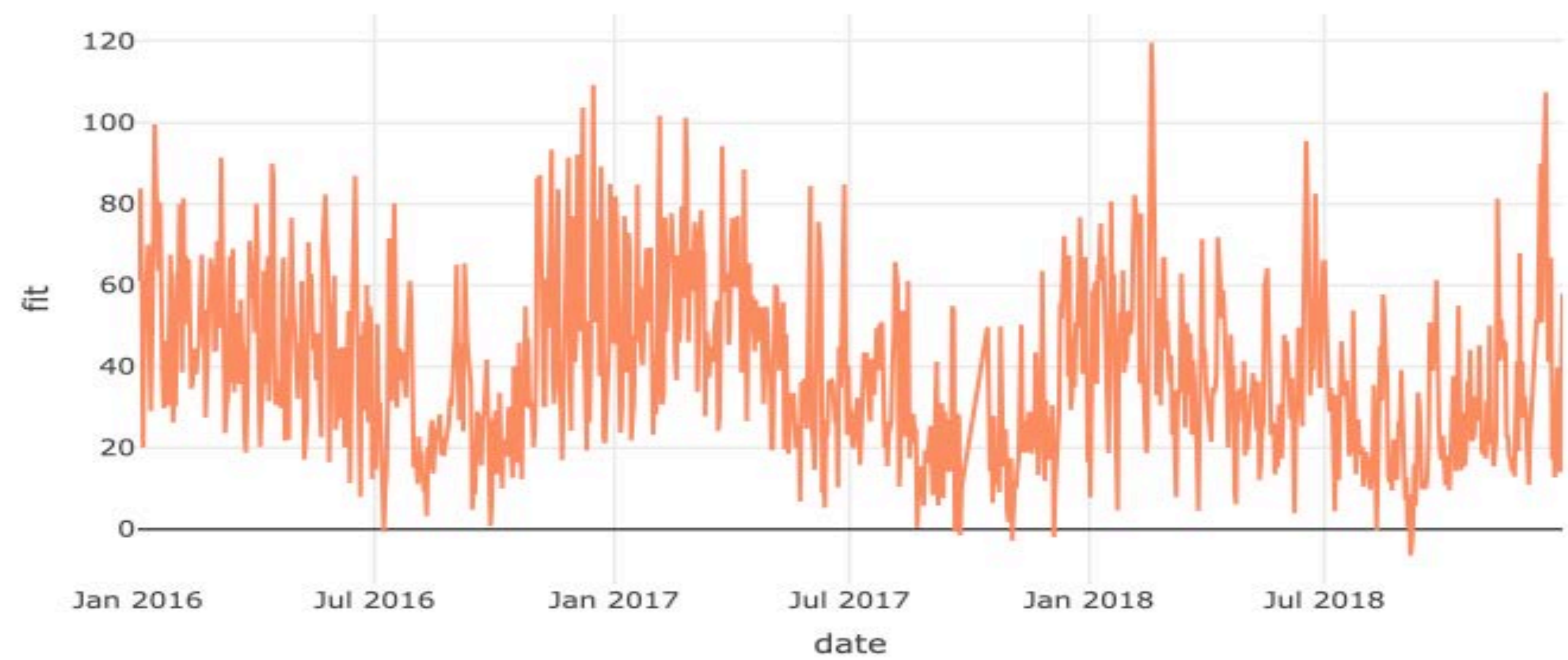
FORECAST

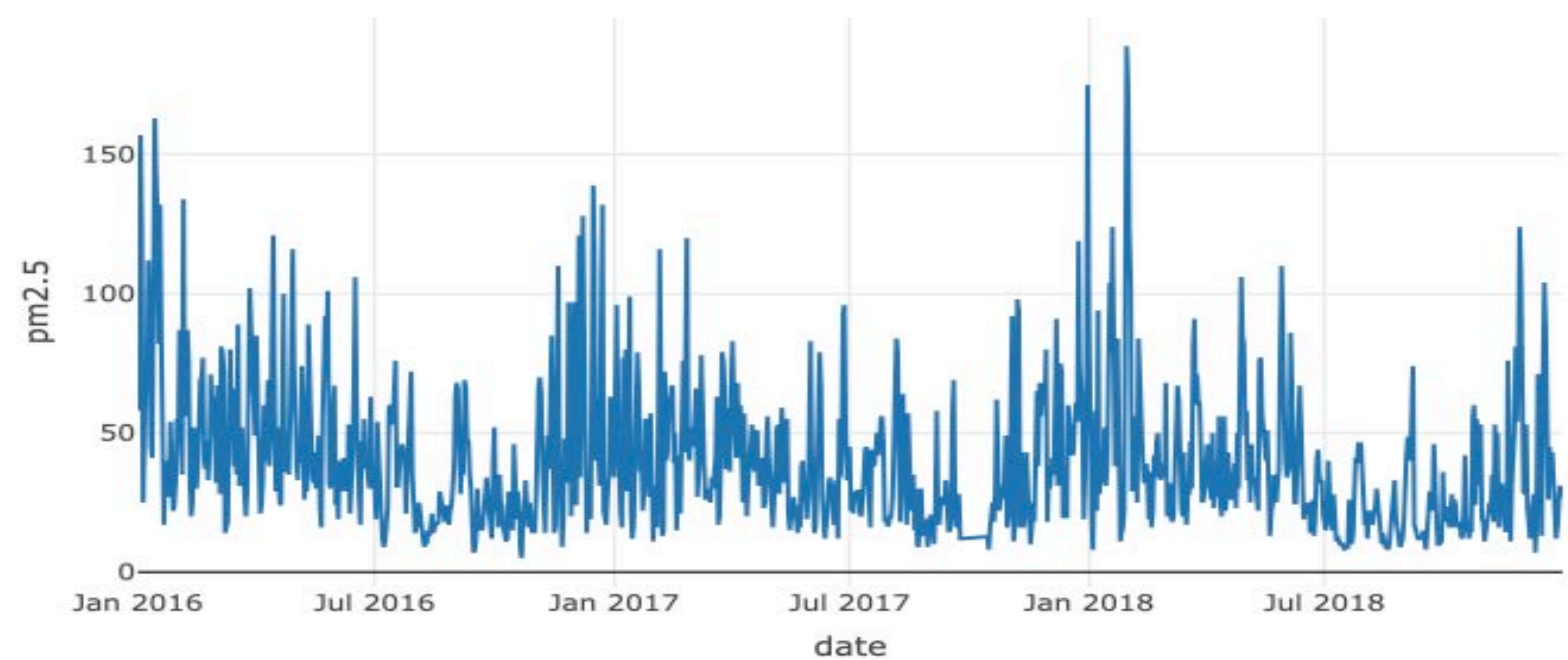
- Daily pm2.5 and weather
    - VAR(1)



FORECAST

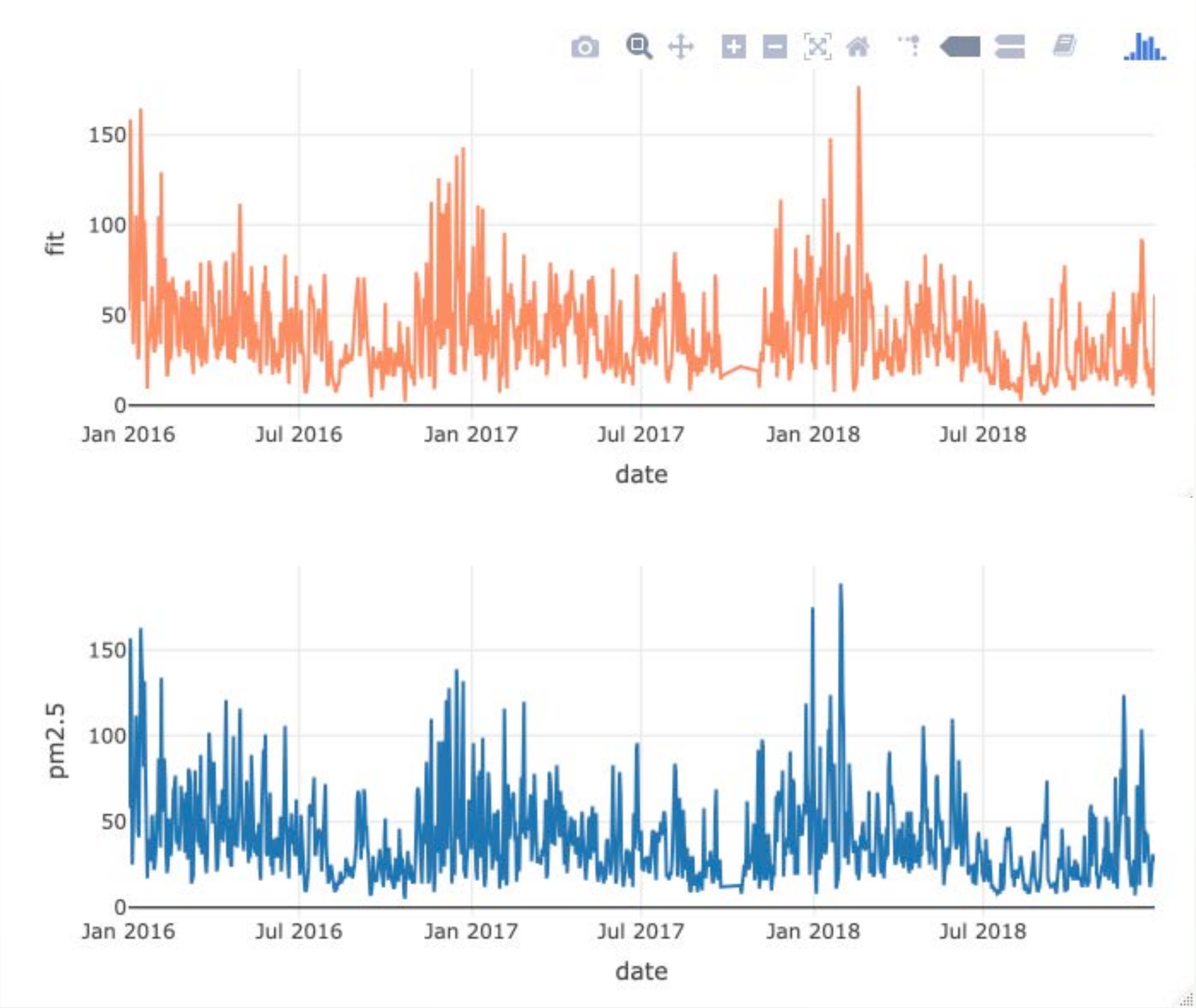
- Daily pm2.5 and weather
    - GAM





FORECAST

- Daily pm2.5 and other pollutions
  - GAM



SUGGESTION

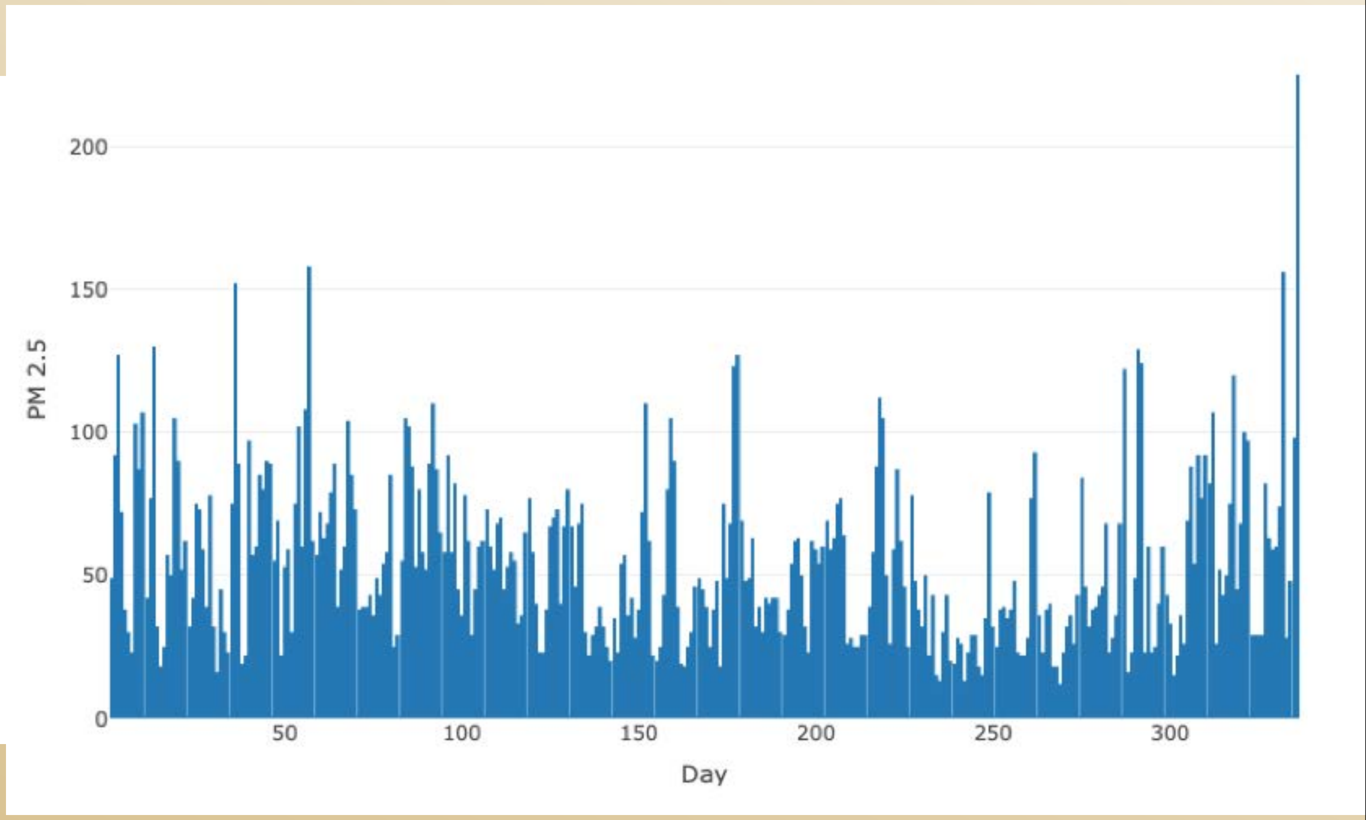
TESTS BASED ON THE POLICY

- Shanghai Clean Air Action Plan 2013-2017
- Aiming to reduce the concentration of PM2.5 by 20 percent in five years.
- Average PM 2.5 in 2013: 59.94 μg/m^3
- Average PM 2.5 in 2017: 54.34 μg/m^3
- The PM2.5 reduced 9.34% from 2013 to 2017

TESTS BASED ON THE POLICY

- Shanghai environmental protection plan for 2015 to 2017
- The goal is to reduce the average PM2.5 concentration to 48 µg/m^3 by the end of 2017.

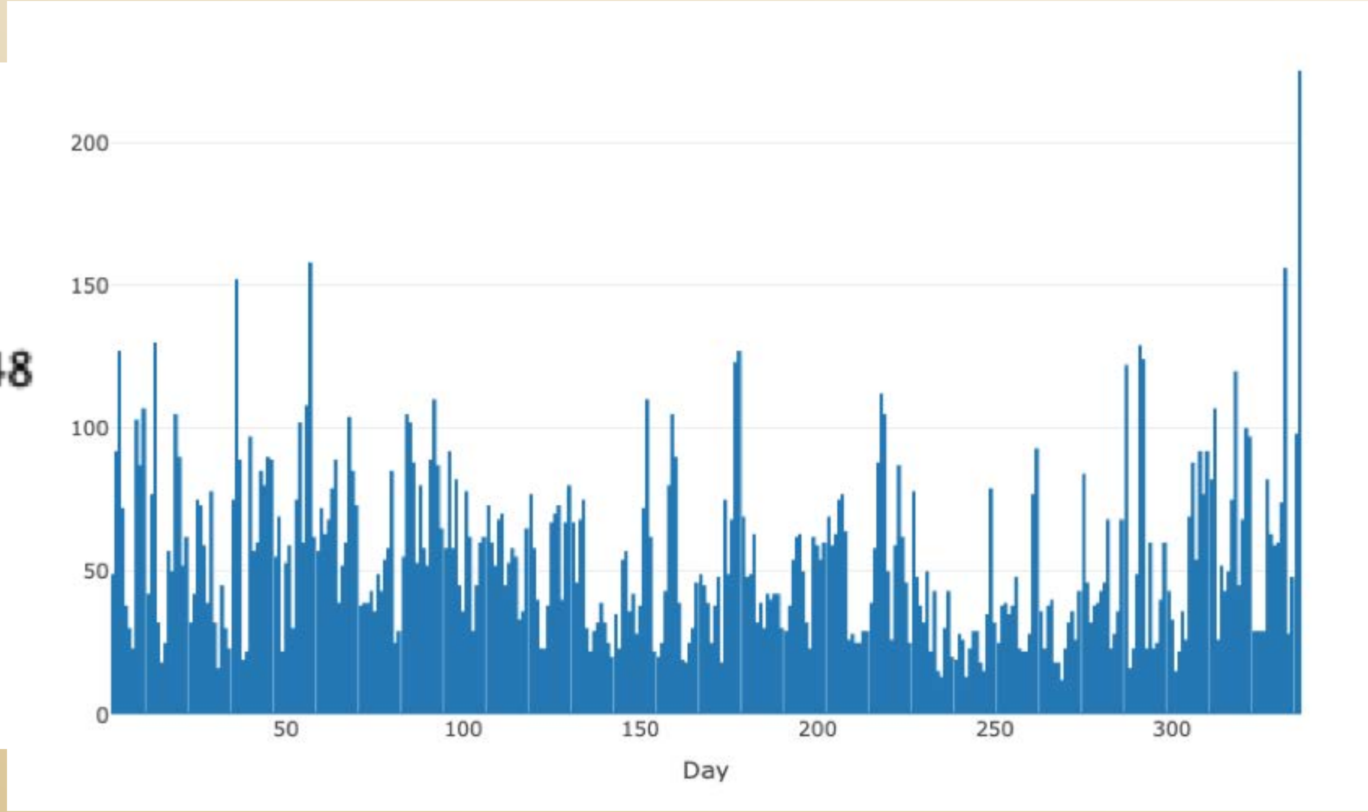
```
        One Sample t-test

data:  daily2017$pm
t = 3.9238, df = 335, p-value = 0.9999
alternative hypothesis: true mean is less than 48
95 percent confidence interval:
     -Inf 56.99986
sample estimates:
mean of x
 54.33631
```



TESTS BASED ON THE POLICY

- Then, use data of 2018 to conduct the same test

```
        One Sample t-test

data:  daily2018$pm
t = 0.89191, df = 363, p-value = 0.8135
alternative hypothesis: true mean is less than 48
95 percent confidence interval:
     -Inf 52.39858
sample estimates:
mean of x
 49.54396
```



TESTS BASED ON THE POLICY

- In late November 2013, Shanghai announced rules for a carbon emissions trading scheme.
- In late January 2014, Shanghai announced a ban on the burning of straw and other bonfires within all of Shanghai.
- Comparing the concentrate of PM2.5 between 2013.11-2014.2 and 2014.11-2015.2.

```
        Two Sample t-test

data:  groupdata$pm by groupdata$group
t = 1.9954, df = 238, p-value = 0.02357
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.121682      Inf
sample estimates:
mean in group 1 mean in group 2
       82.32067        70.01900
```


THANK YOU

PM2.5 SHANGHAI

ZHU WANG (王祝)
WEI ZHANG (张薇)


FINAL PRESENTATION
A conclusion

WHAT WE'VE DONE…

- **Background Introduction**
- **Exploration Data Analysis**
  - Missing data
  - Trends exploration
  - Correlations between pm2.5 and factors
- **Statistical Analysis**
  - **Trends analysis**
  - **Models fitting and prediction**
- **Suggestions**

SHANGHAI

- China's east coast
- Located in the Yangtze River Delta
- A global financial center and transport hub

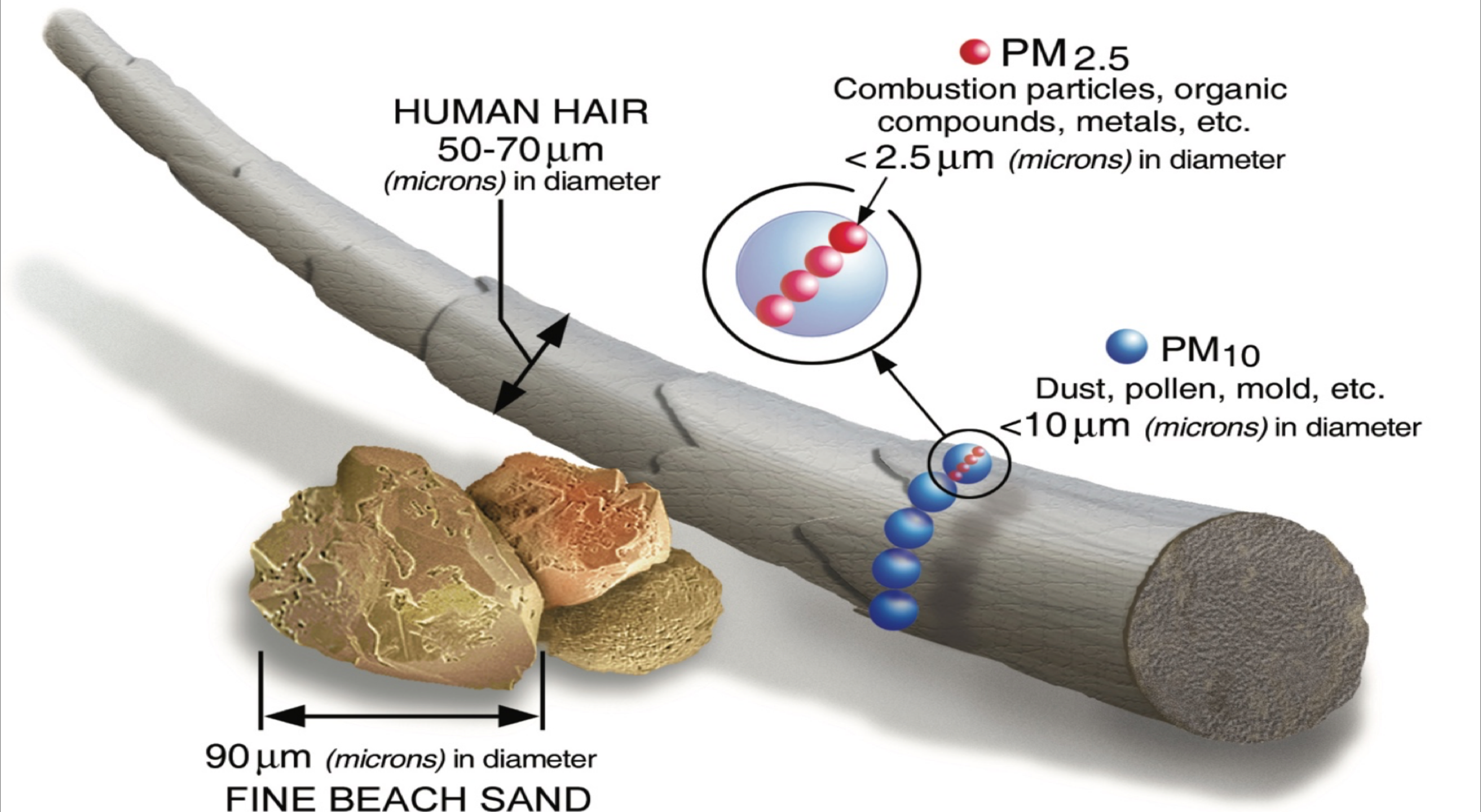
AIR POLLUTION AND GOVERNMENT REACTION

- Not as severe as in many other Chinese cities, but still substantial by world standards.
- On 23 January 2014, the mayor of Shanghai municipality announced that three main measures would be taken to manage the air pollution in Shanghai, along with surrounding Anhui, Jiangsu and Zhejiang provinces.
- The measures involved delivery of the 2013 air cleaning program, linkage mechanism with the three surrounding provinces and improvement of the ability of early warning of emergency situation.

WHAT IS PM

- particulate matter (also called particle pollution)
- a mixture of solid particles and liquid droplets found in the air
- Particle pollution includes:
  - PM10 : inhalable particles, with diameters that are generally 10 micrometers and smaller;
  - PM2.5 : fine inhalable particles, with diameters that are generally 2.5 micrometers and smaller.

WHAT IS PM



PM 2.5
Combustion particles, organic compounds, metals, etc.
< 2.5 µm (microns) in diameter
HUMAN HAIR
50-70 µm
(microns) in diameter
PM 10
Dust, pollen, mold, etc.
<10 µm (microns) in diameter
90 µm (microns) in diameter
FINE BEACH SAND

SOURCE OF PM

- These particles come in many sizes and shapes and can be made up of hundreds of different chemicals.
- From natural:
  - volcanoes, dust storms, forest and grassland fires, living vegetation and sea spray
- From human activities:
  - burning of fossil fuels in vehicles, stubble burning, power plants, and various industrial processes

HARMFUL EFFECTS OF PM

- Health Effect
- Small particles less than 10 micrometers in diameter can get deep into your lungs, and some may even get into your bloodstream.
- Exposure to such particles can affect both your lungs and your heart.
- premature death in people with heart or lung disease
- nonfatal heart attacks
- irregular heartbeat
- aggravated asthma
- decreased lung function

HARMFUL EFFECTS OF PM

- Environmental Effect
- Particles can be carried over long distances by wind and then settle on ground or water.
- making lakes and streams acidic
- changing the nutrient balance in coastal waters and large river basins
- damaging sensitive forests and farm crops
- Affecting the diversity of ecosystems
- contributing to acid rain effects.

AIR QUALITY LEVELS

AQI	
0-50	Good
51-100	Moderate
101-150	Unhealthy for sensitive groups
151-200	Unhealthy
201-300	Very Unhealthy
300+	Hazardous

EXPLORATION DATA ANALYSIS

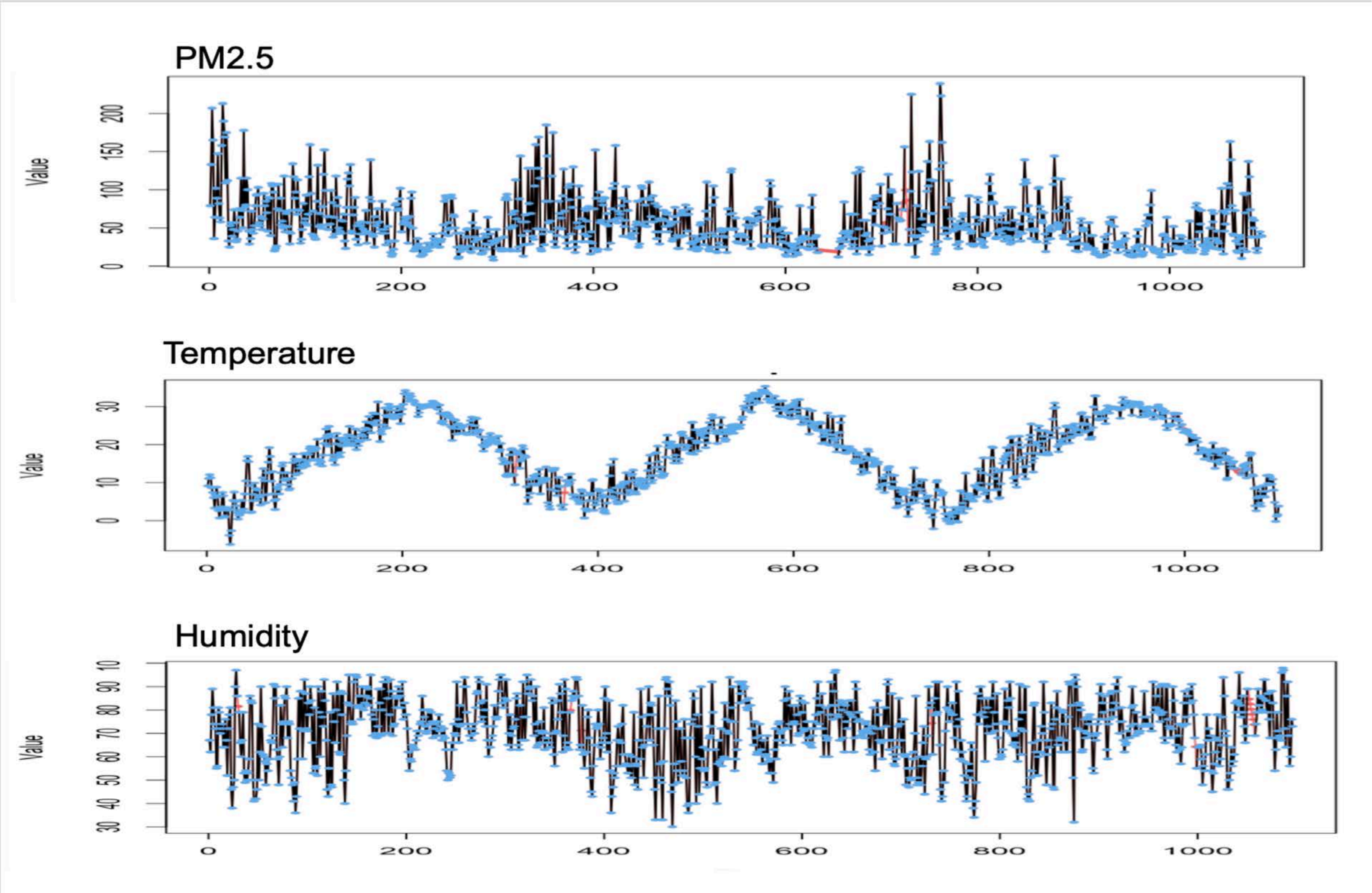
DATA SOURCES

- 2012-2017 Shanghai pm2.5 hourly data
- 2016-2019 Shanghai pollution daily data
- 2016-2018 Shanghai weather data
- 2014-2019 Shanghai pollution daily data (without missing values)

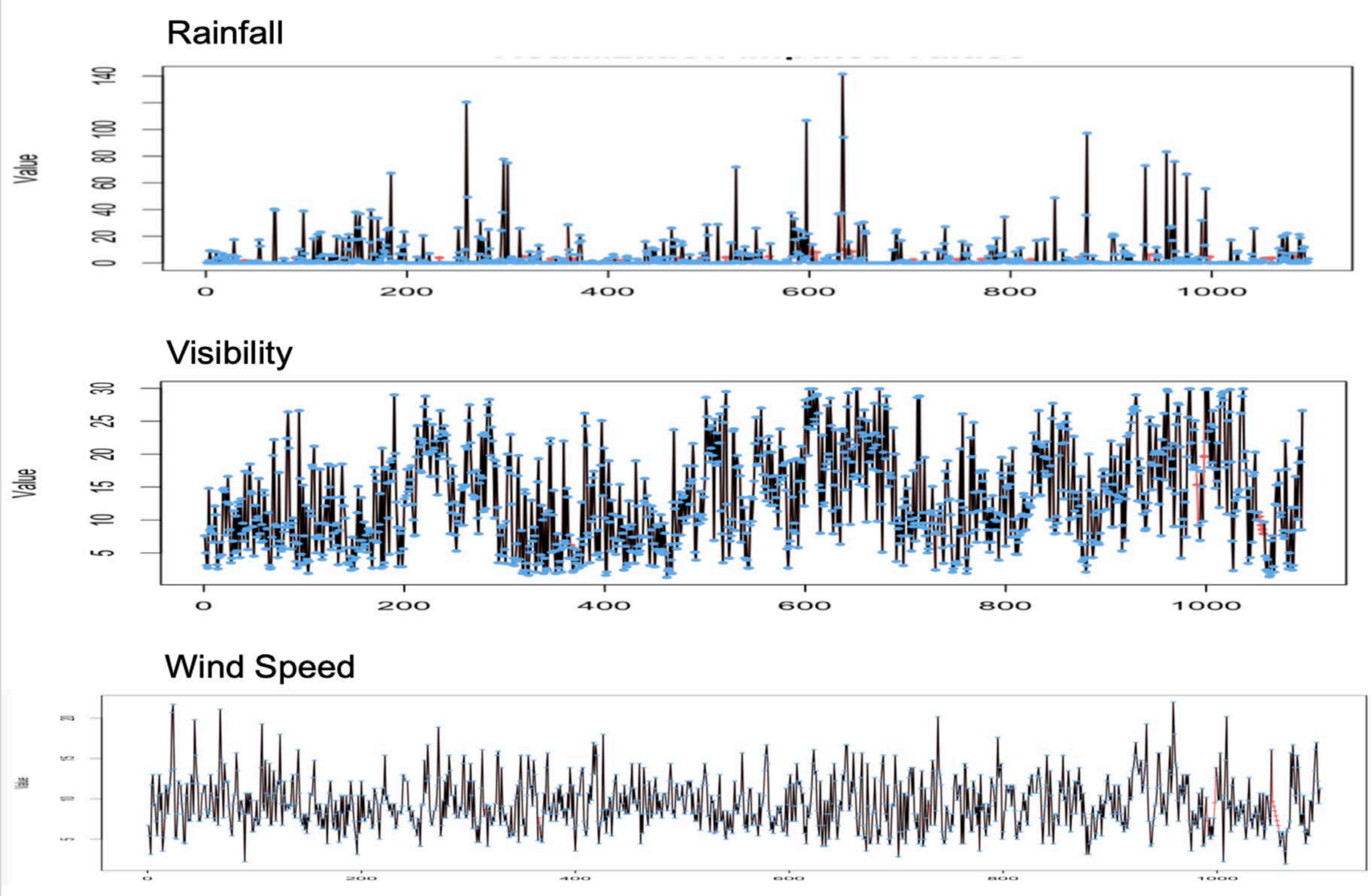
MISSING DATA

- Imputing Method:
  - EM-Algorithm:
    - E-Step: A Kalman filter
    - M-Step: Use the filtered estimates within maximum-likelihood calculations to obtain updated parameter estimates

MISSING DATA

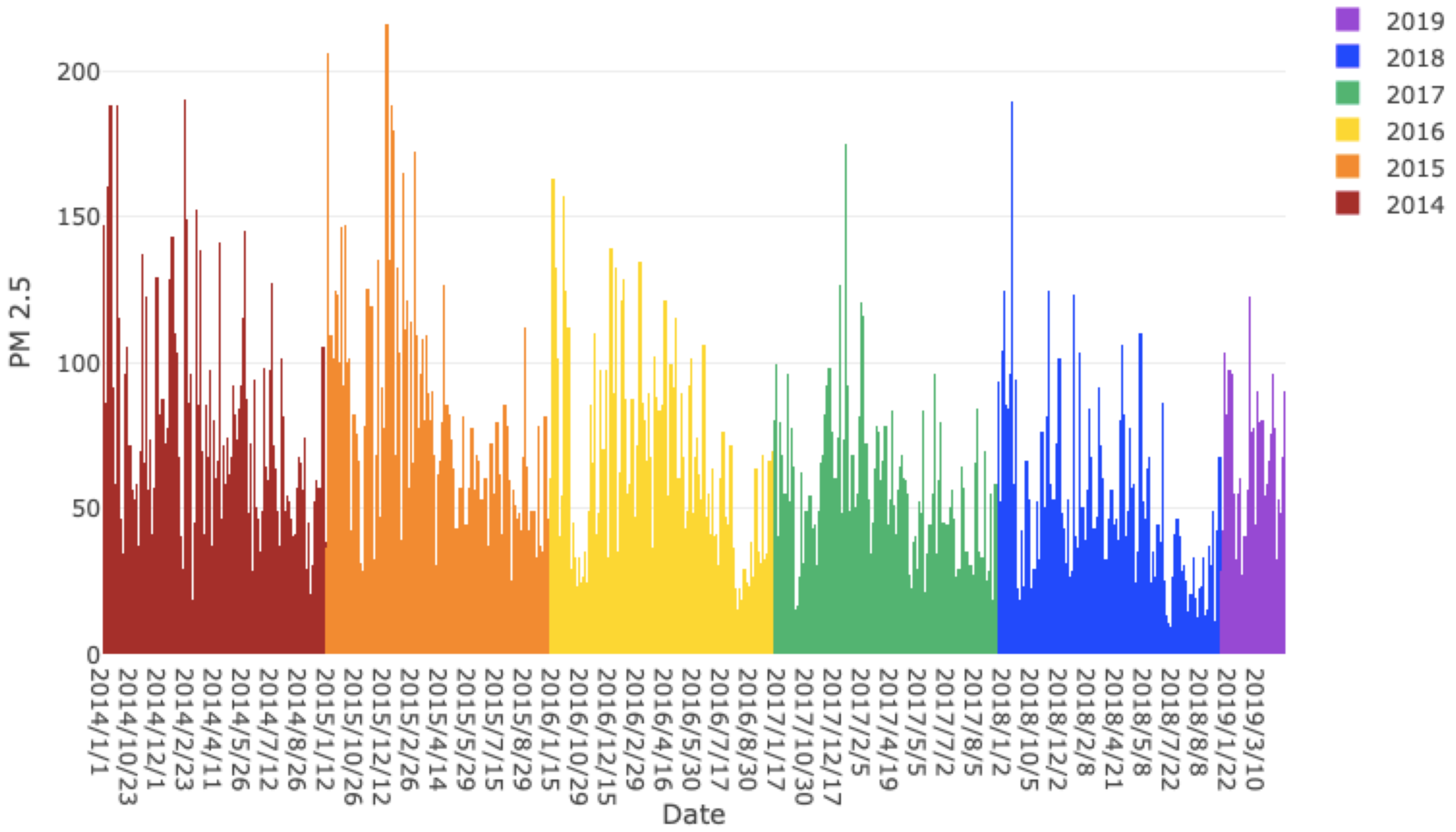


MISSING DATA



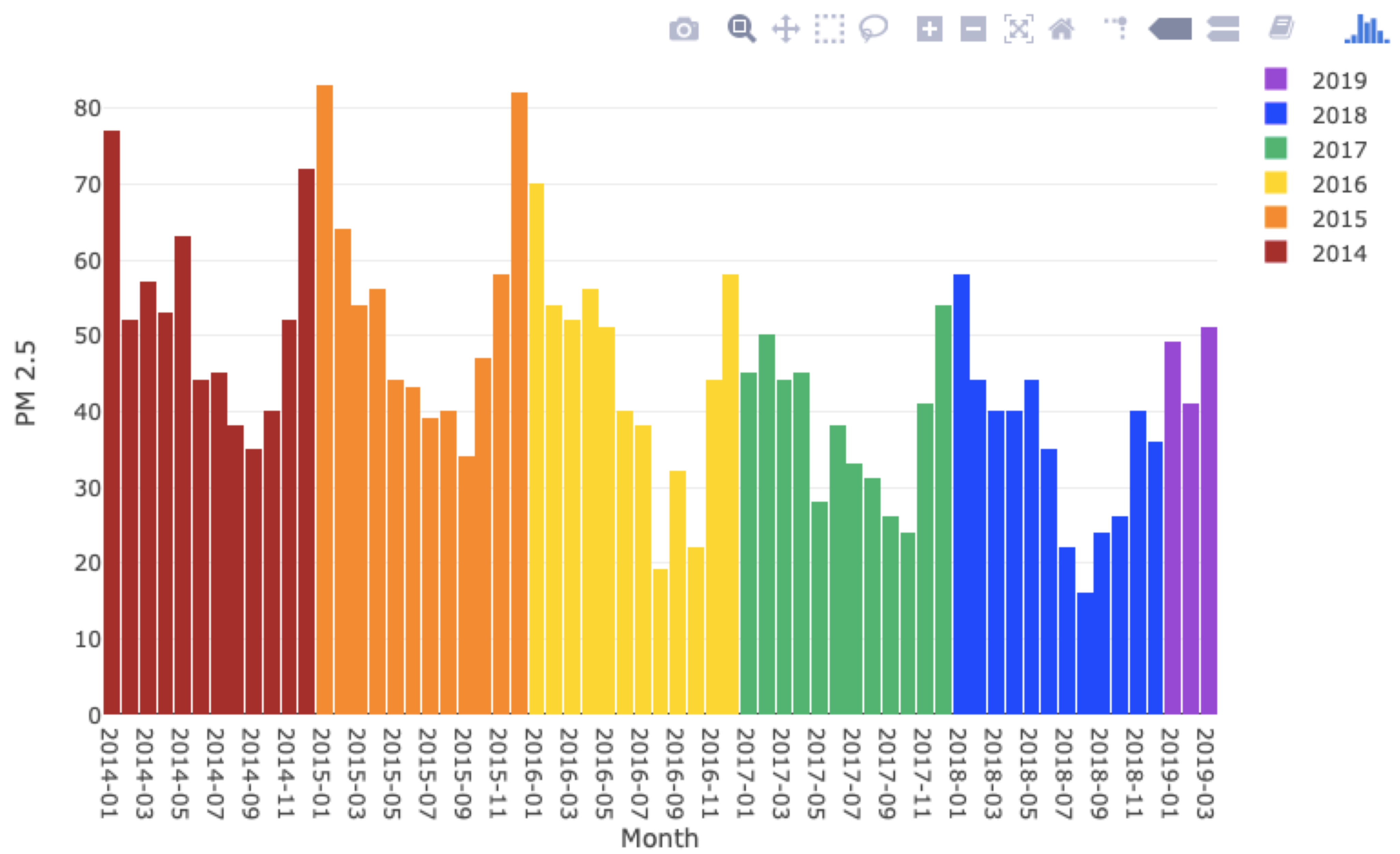
TRENDS EXPLORATION

➤ 2014.01.01-2019.04.15 daily data from aqistudy.cn

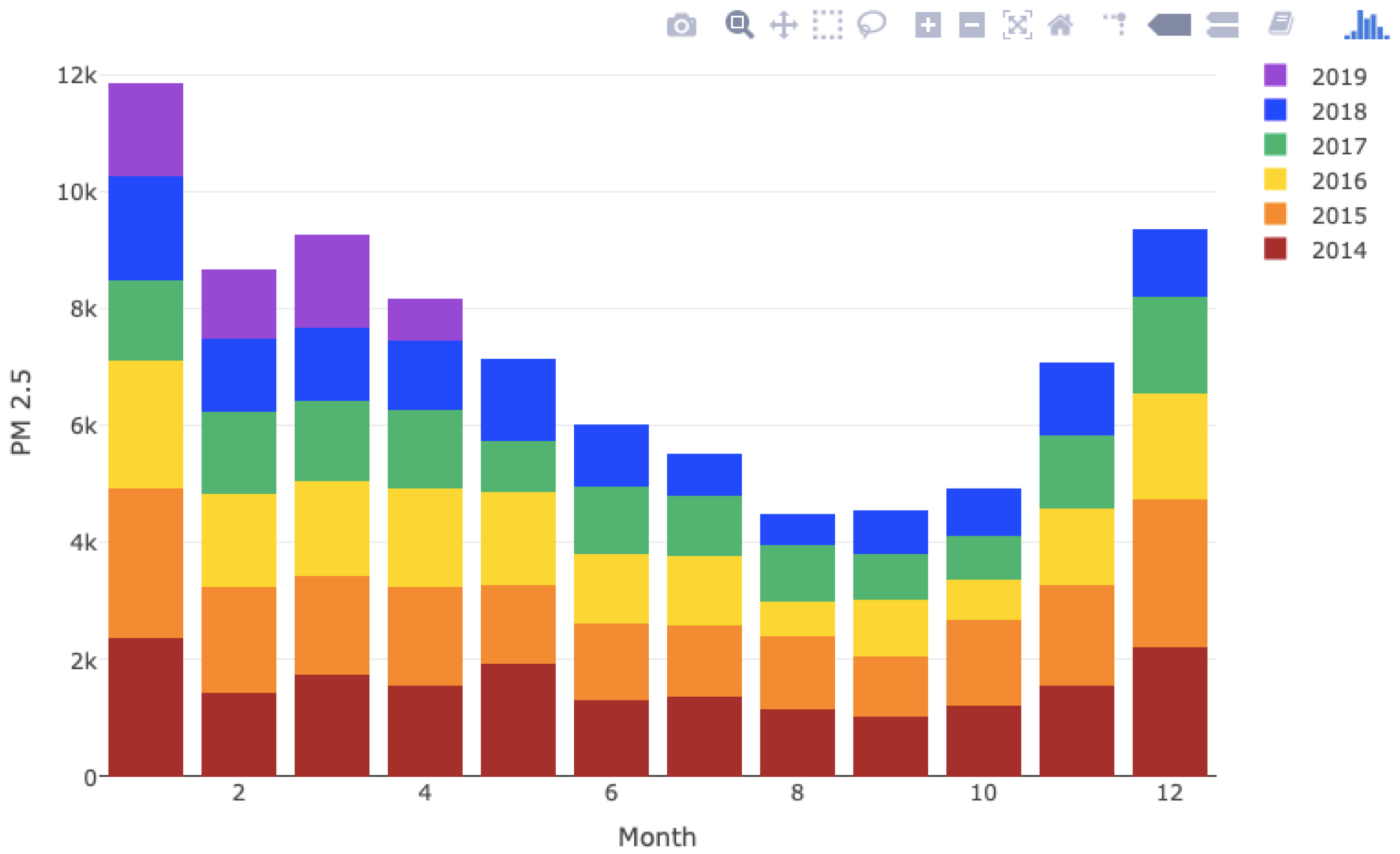


TRENDS EXPLORATION

➤ 2014.01.01-2019.04.15 monthly data from aqistudy.cn

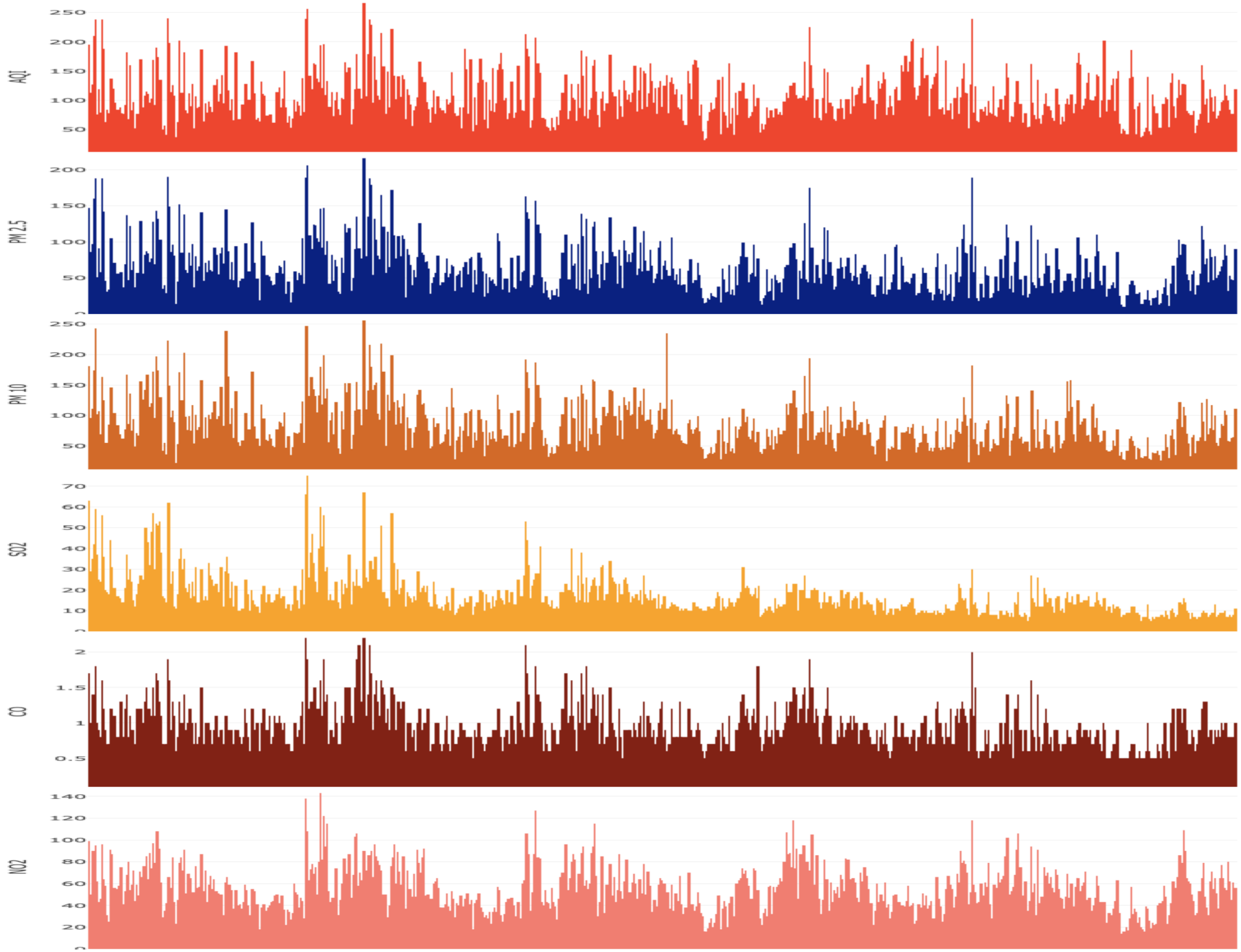


TRENDS EXPLORATION



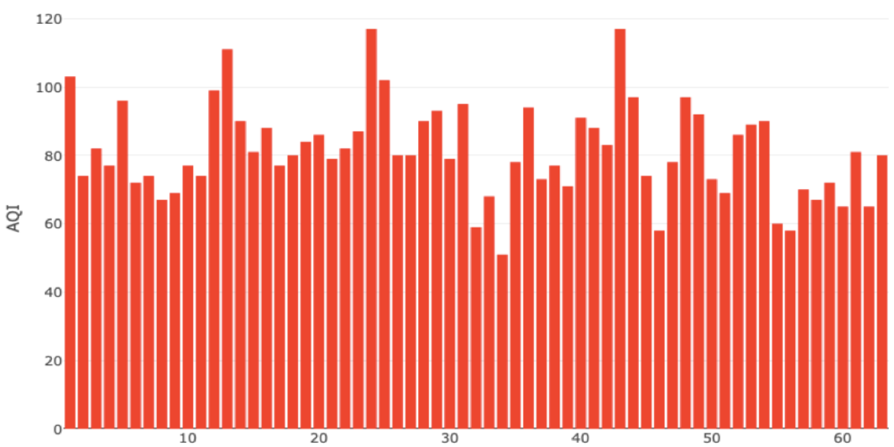
TRENDS EXPLORATION

➤ 2014.01.01-2019.04.15 daily data from aqistudy.cn

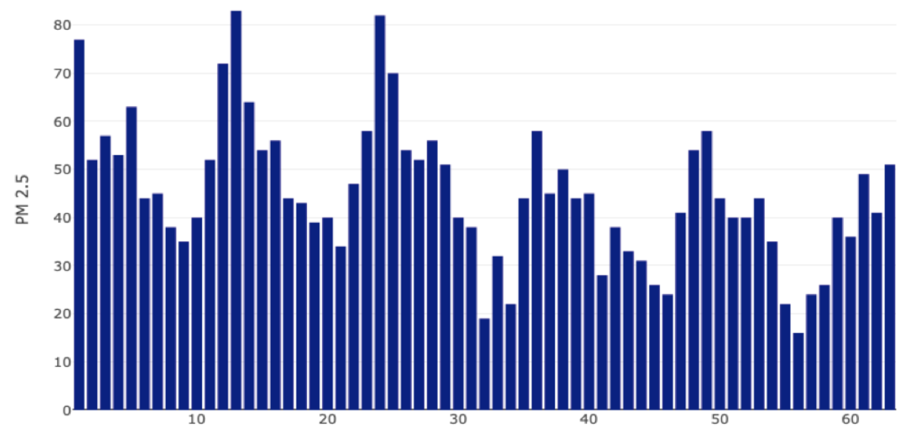


TRENDS EXPLORATION

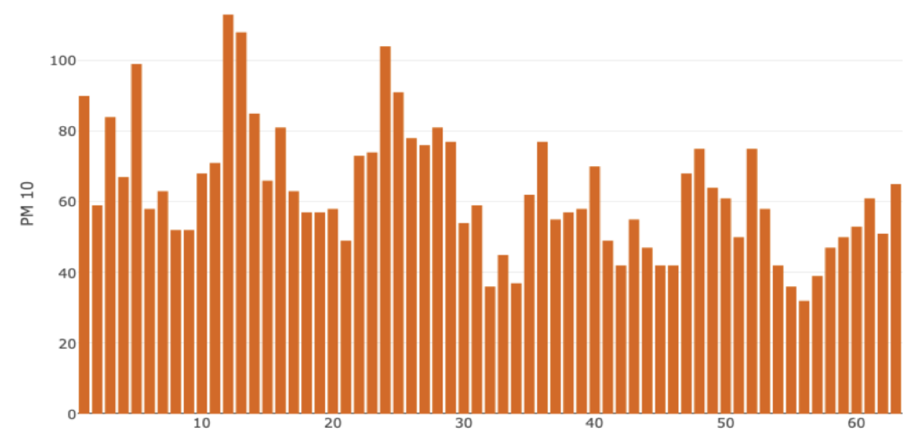
- 2014.01.01-2019.04.15 monthly data from aqistudy.cn



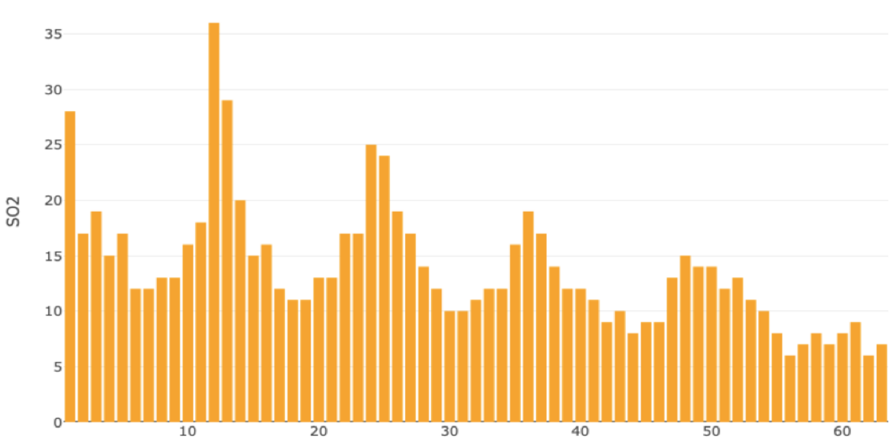
AQI



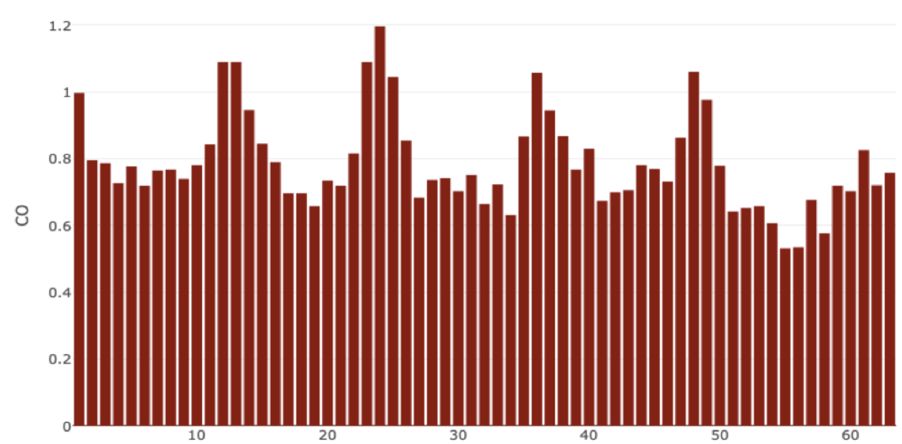
PM 2.5



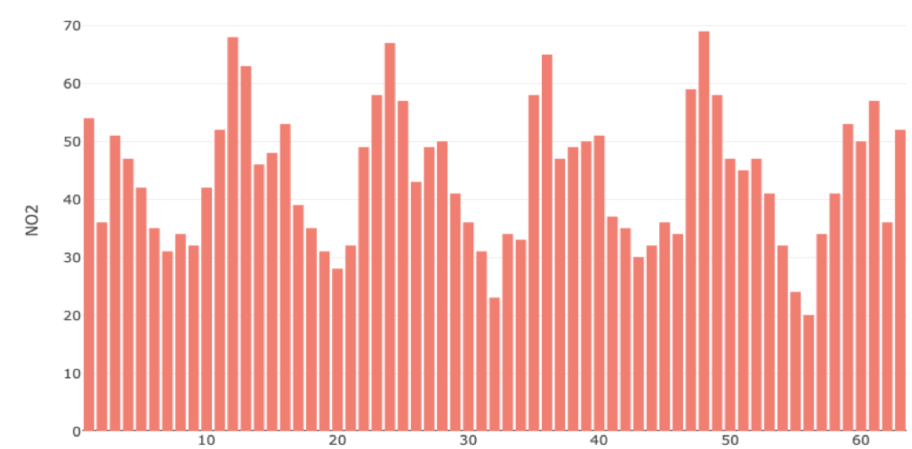
PM 10



SO_2

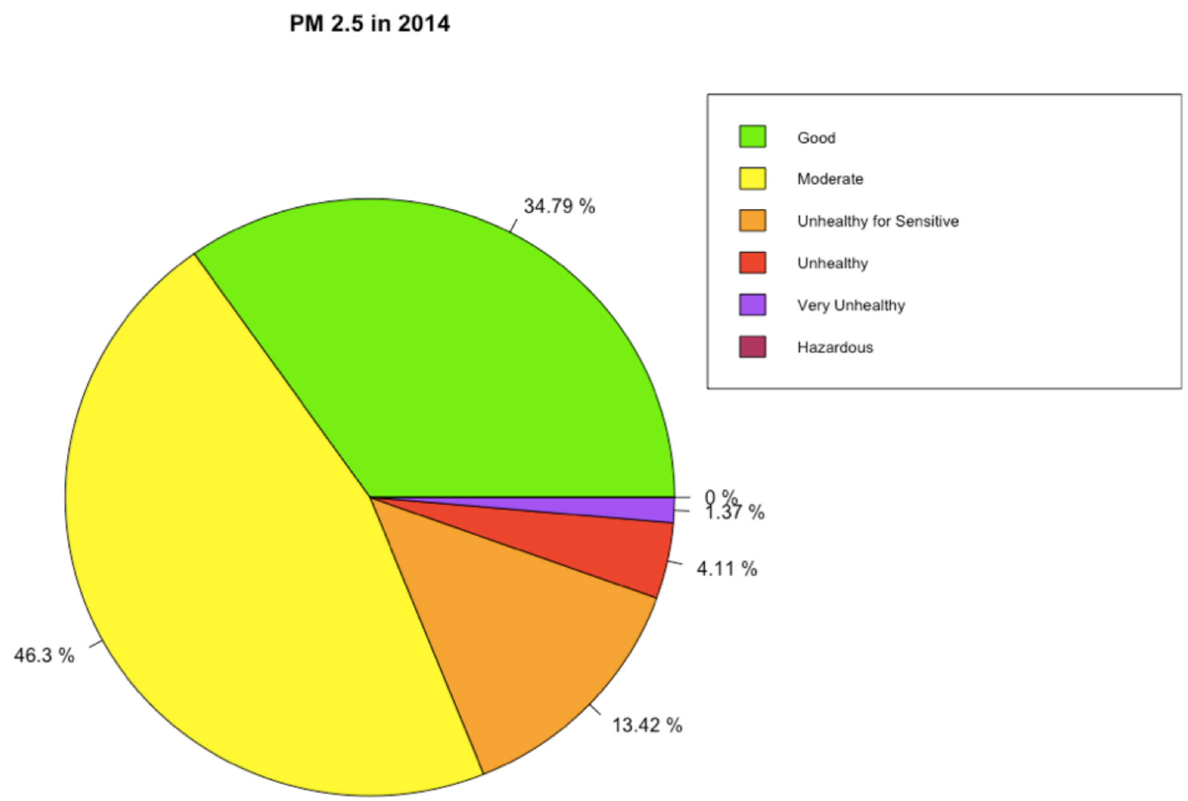


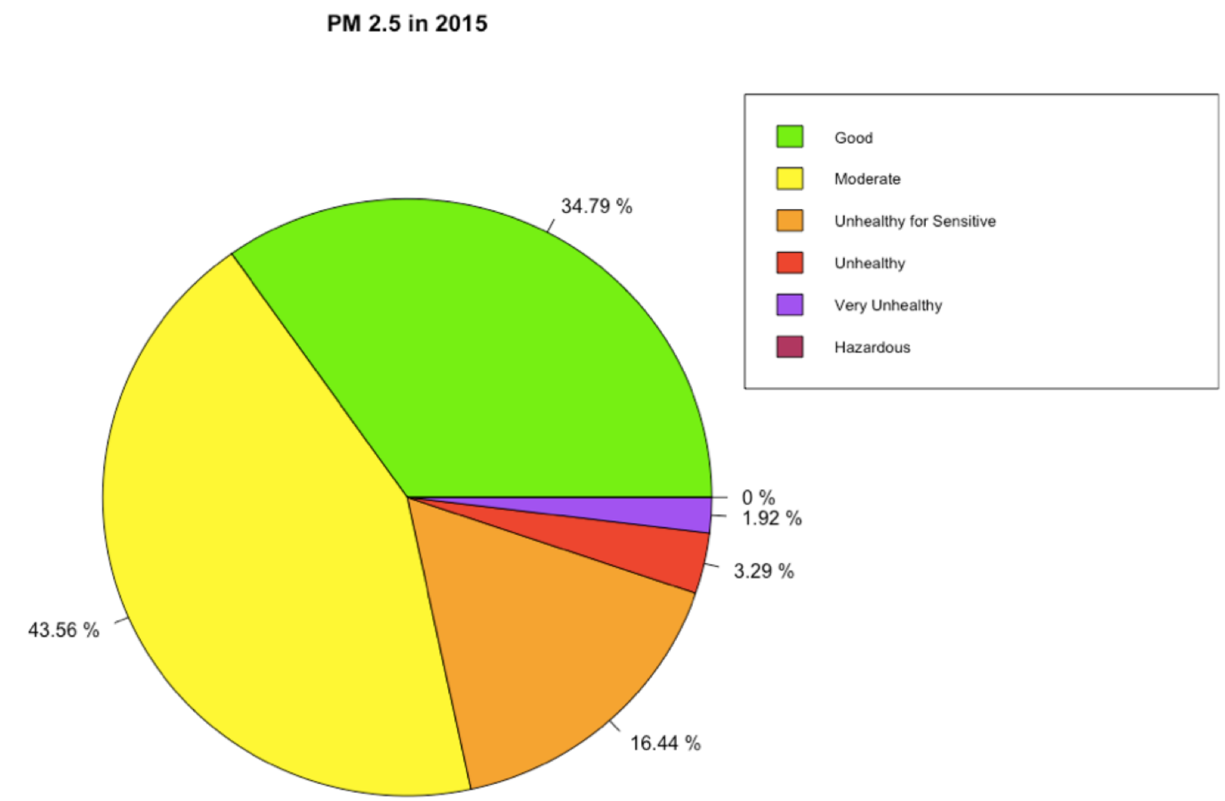
CO

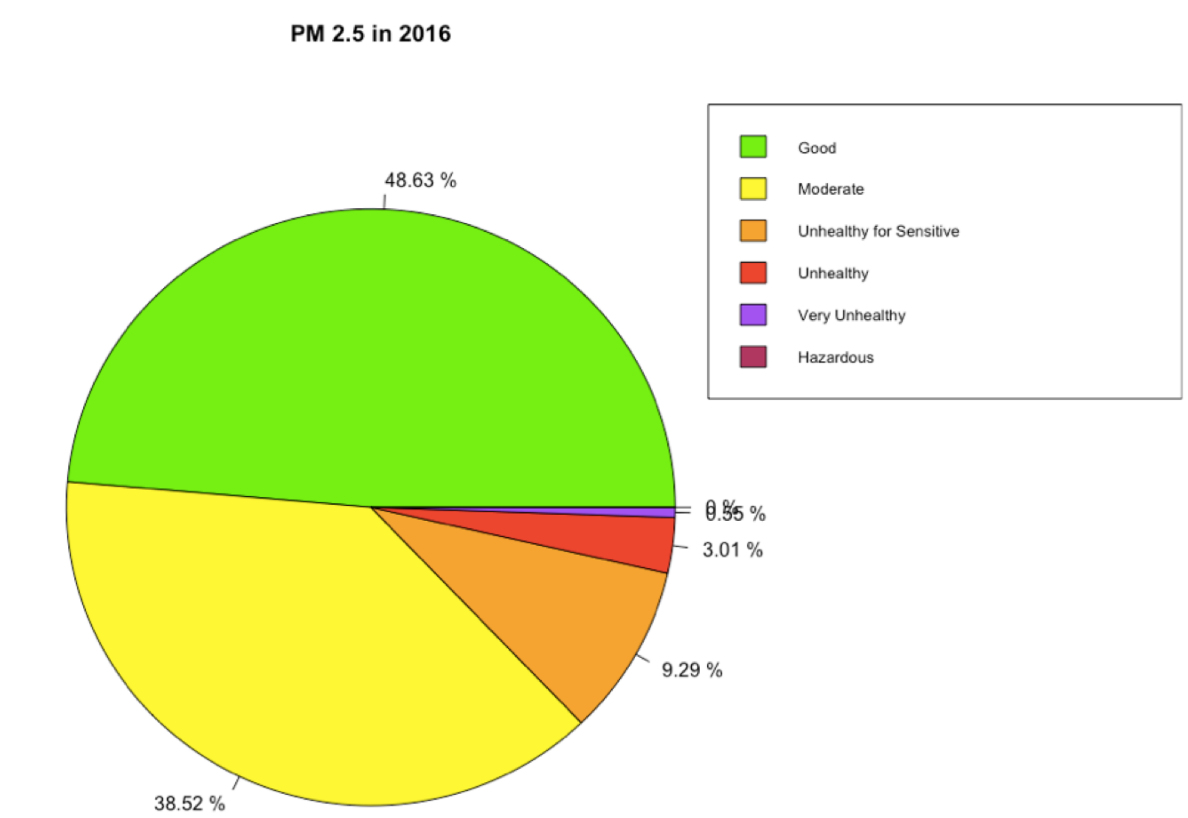


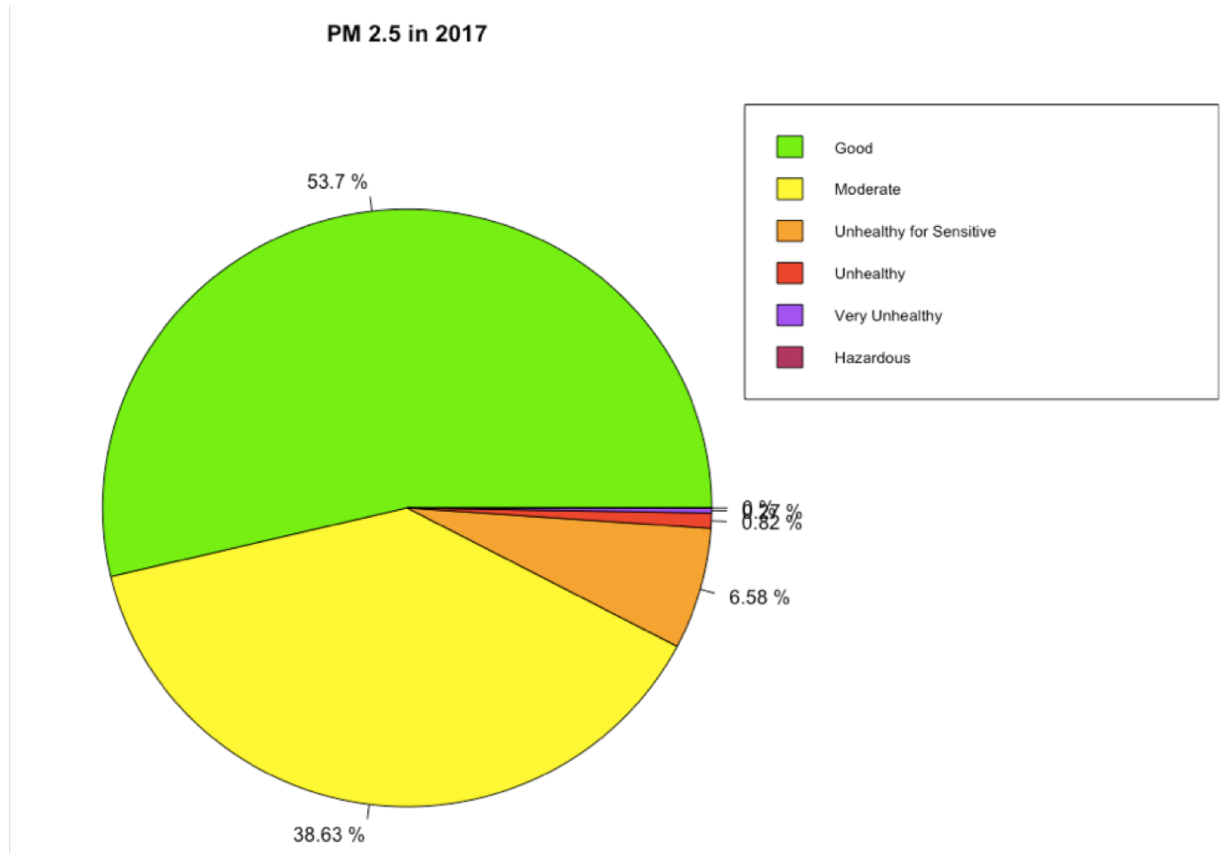
NO_2

TRENDS EXPLORATION

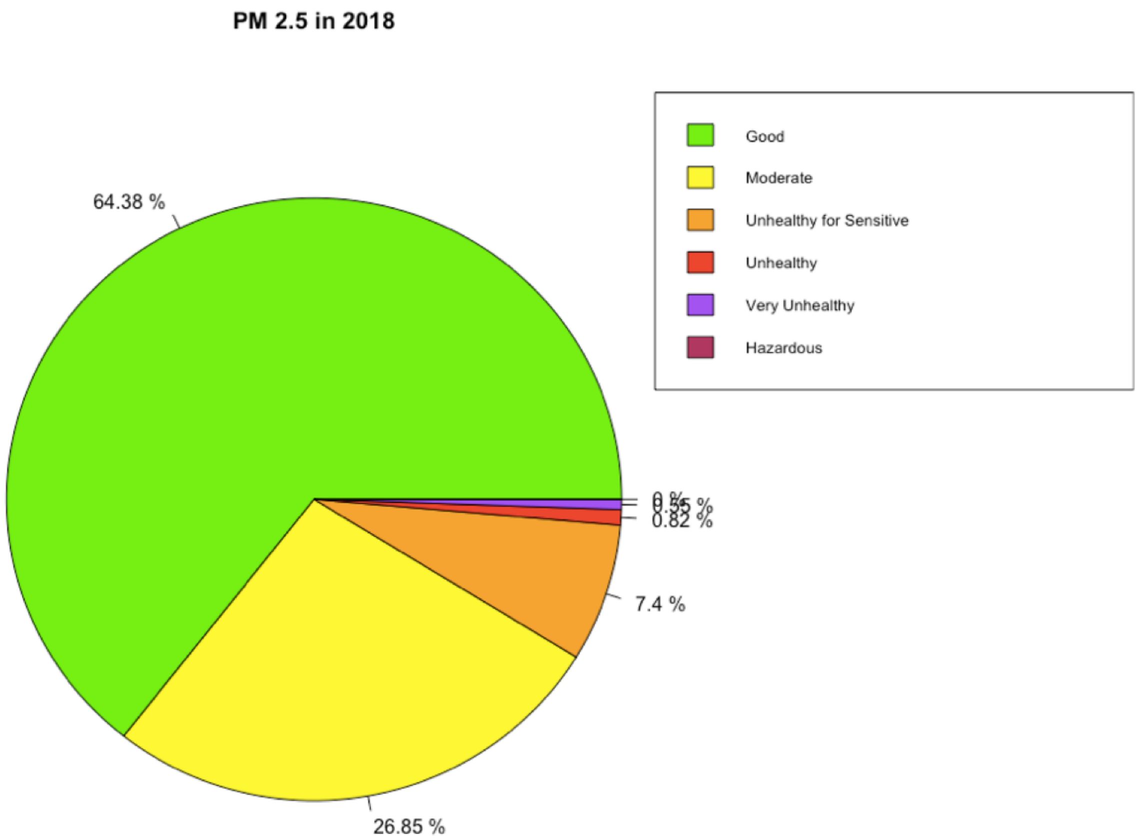


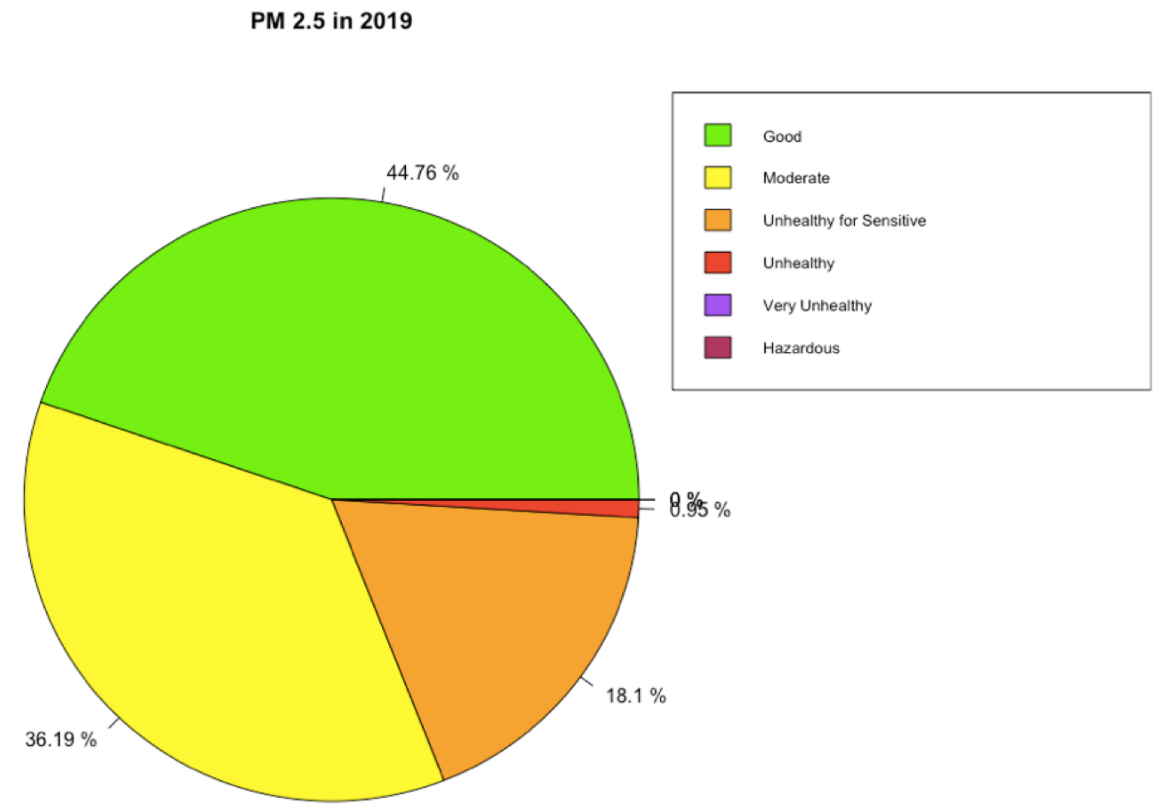




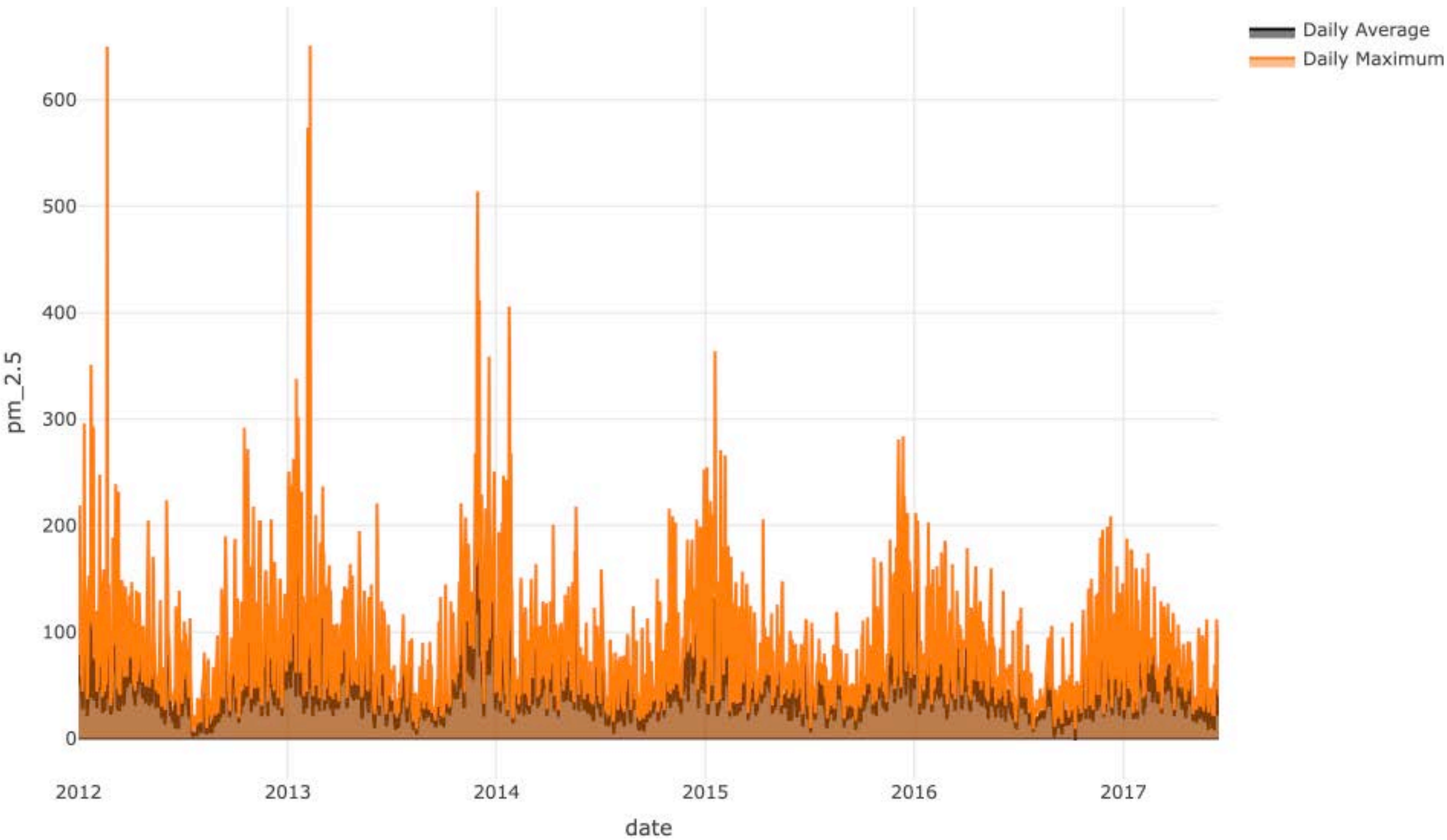


TRENDS EXPLORATION

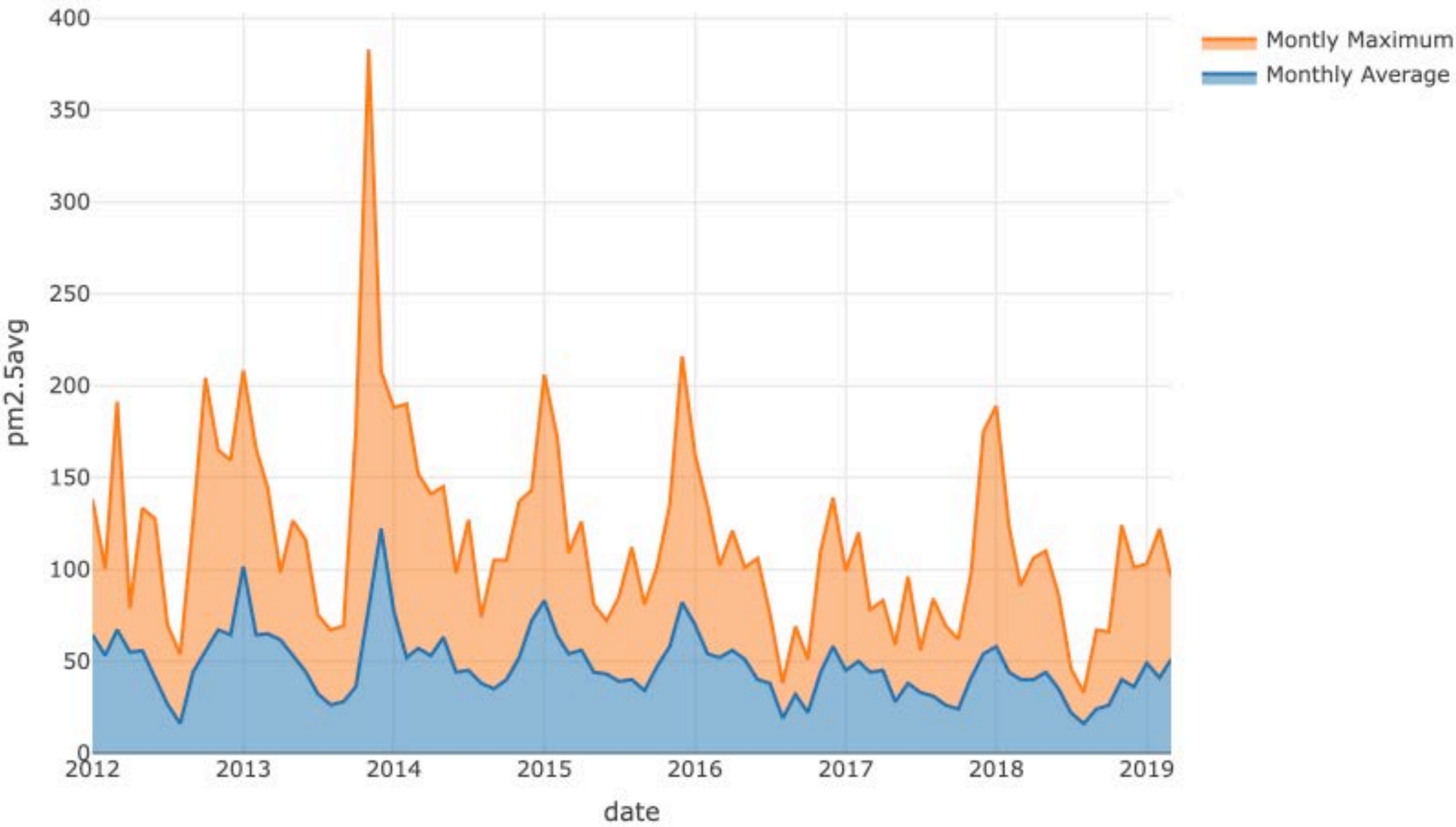




TRENDS EXPLORATION

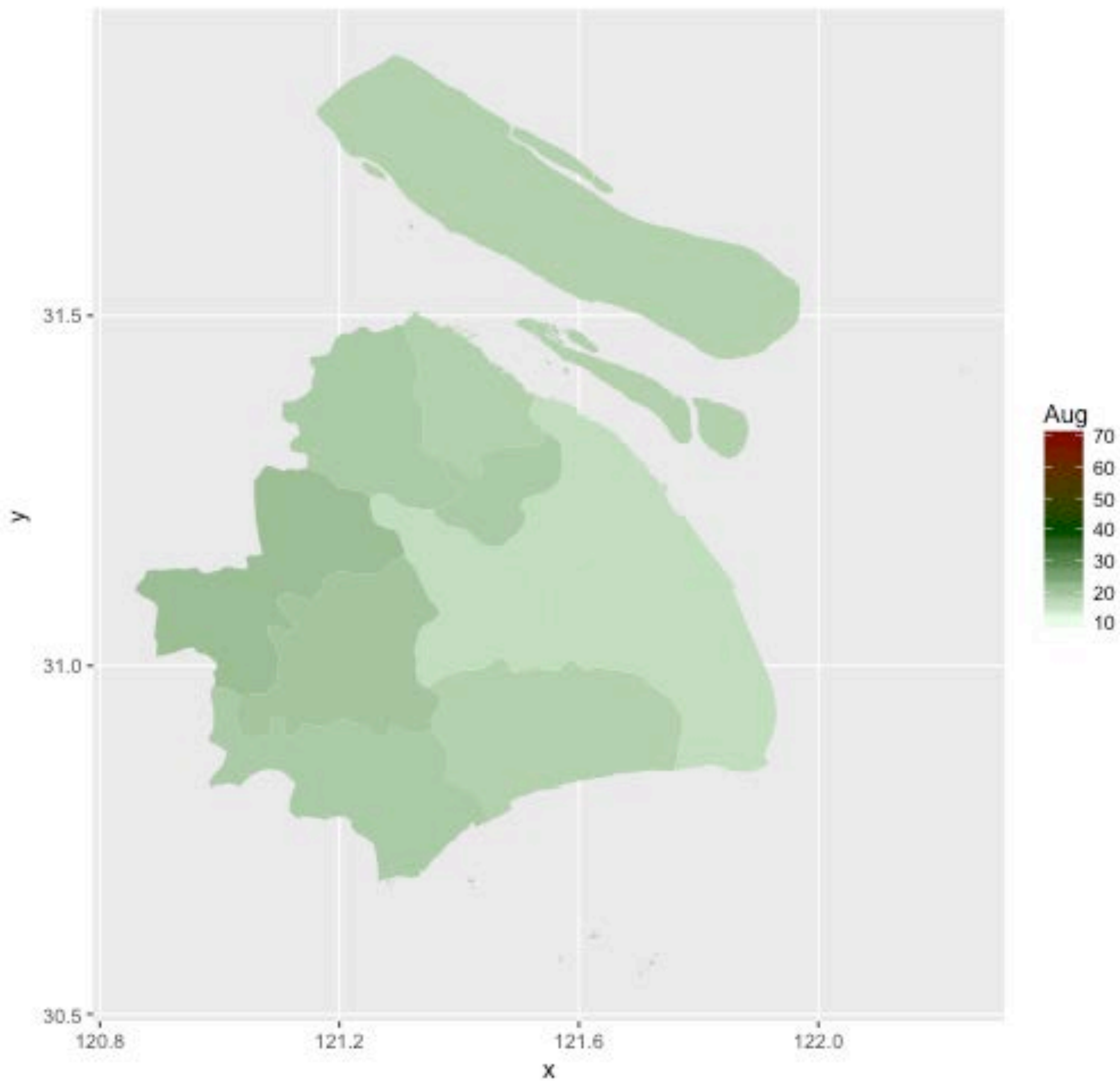


TRENDS EXPLORATION

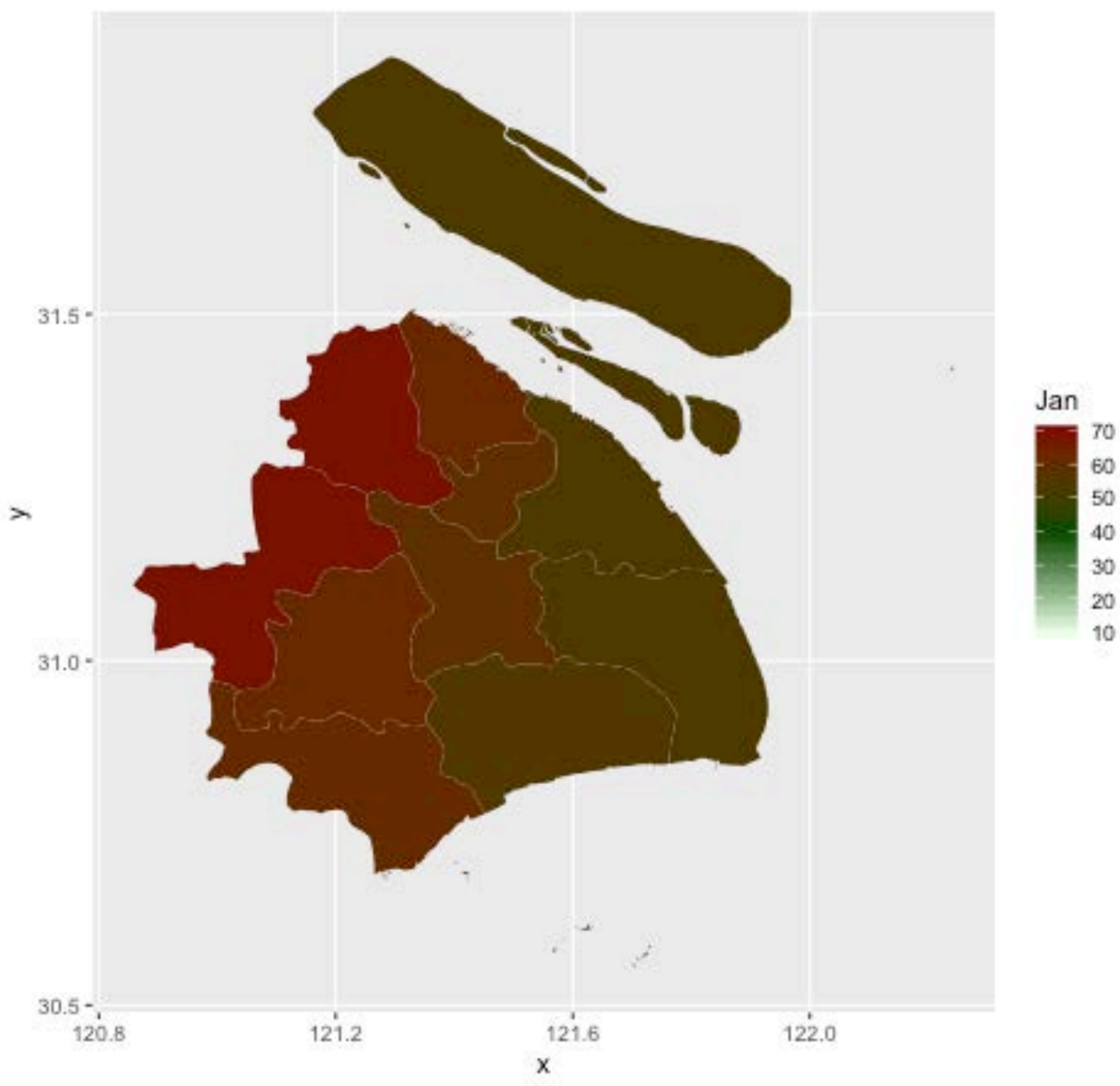


TRENDS EXPLORATION

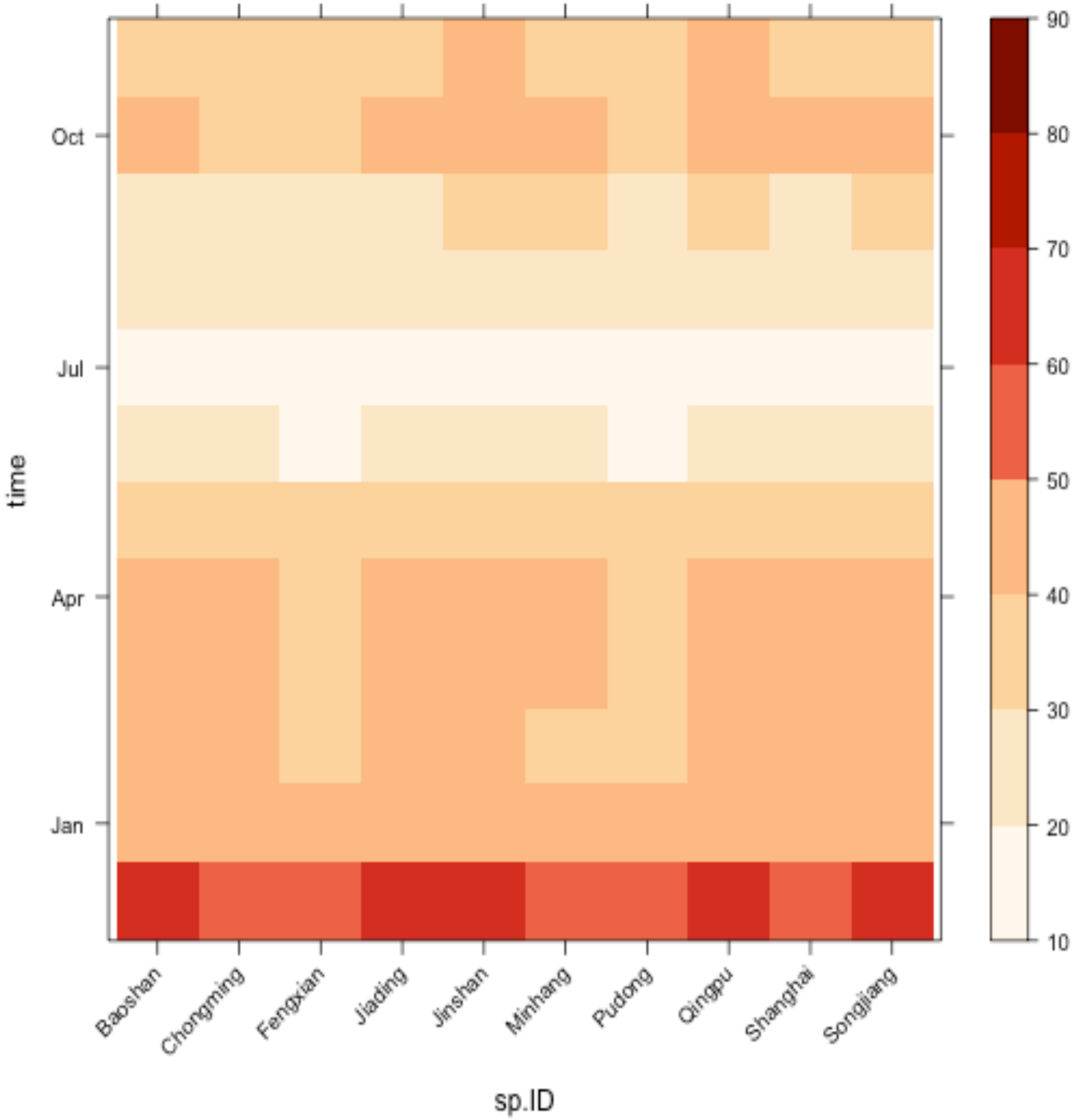
SUMMER

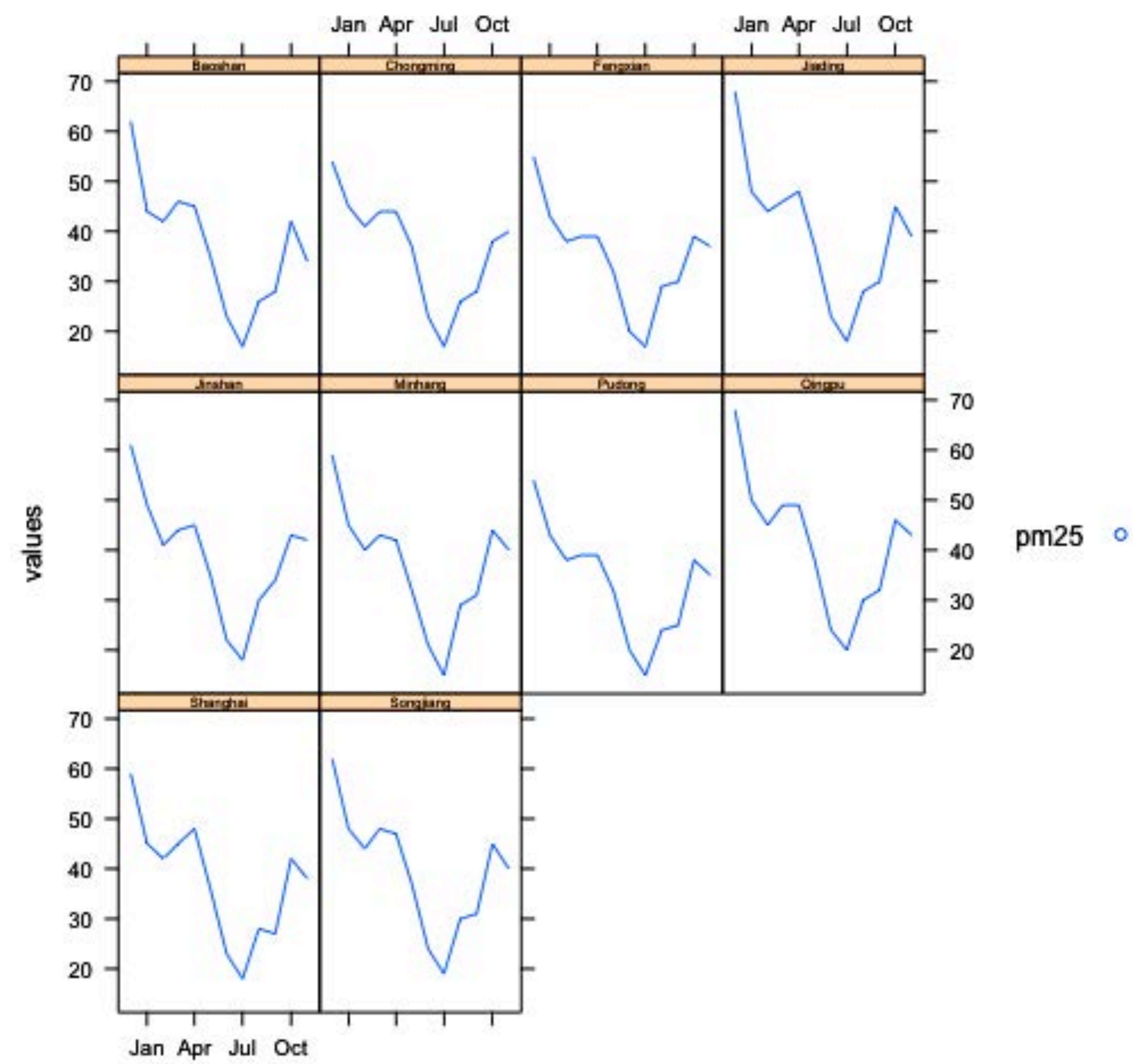


WINTER

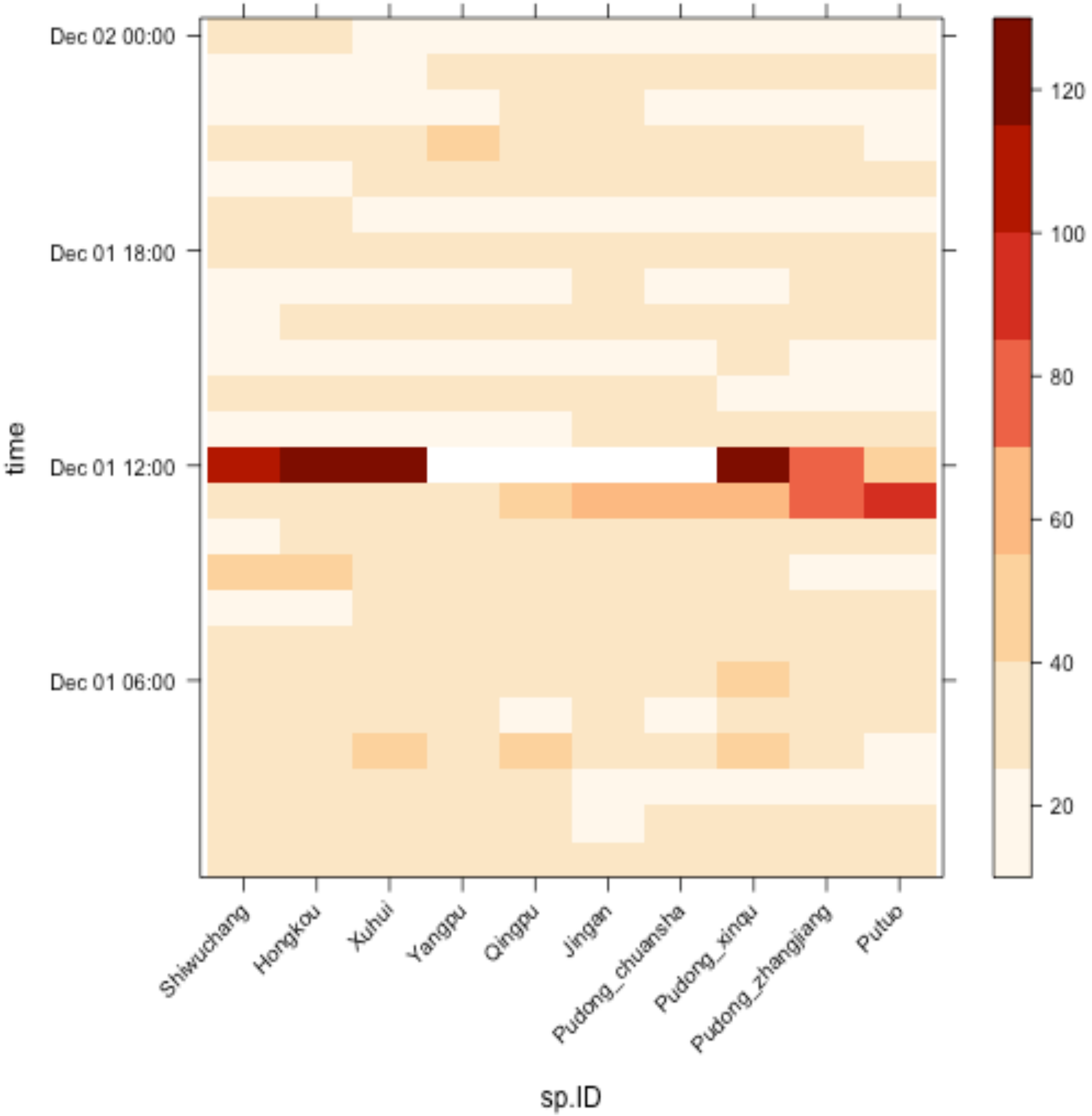


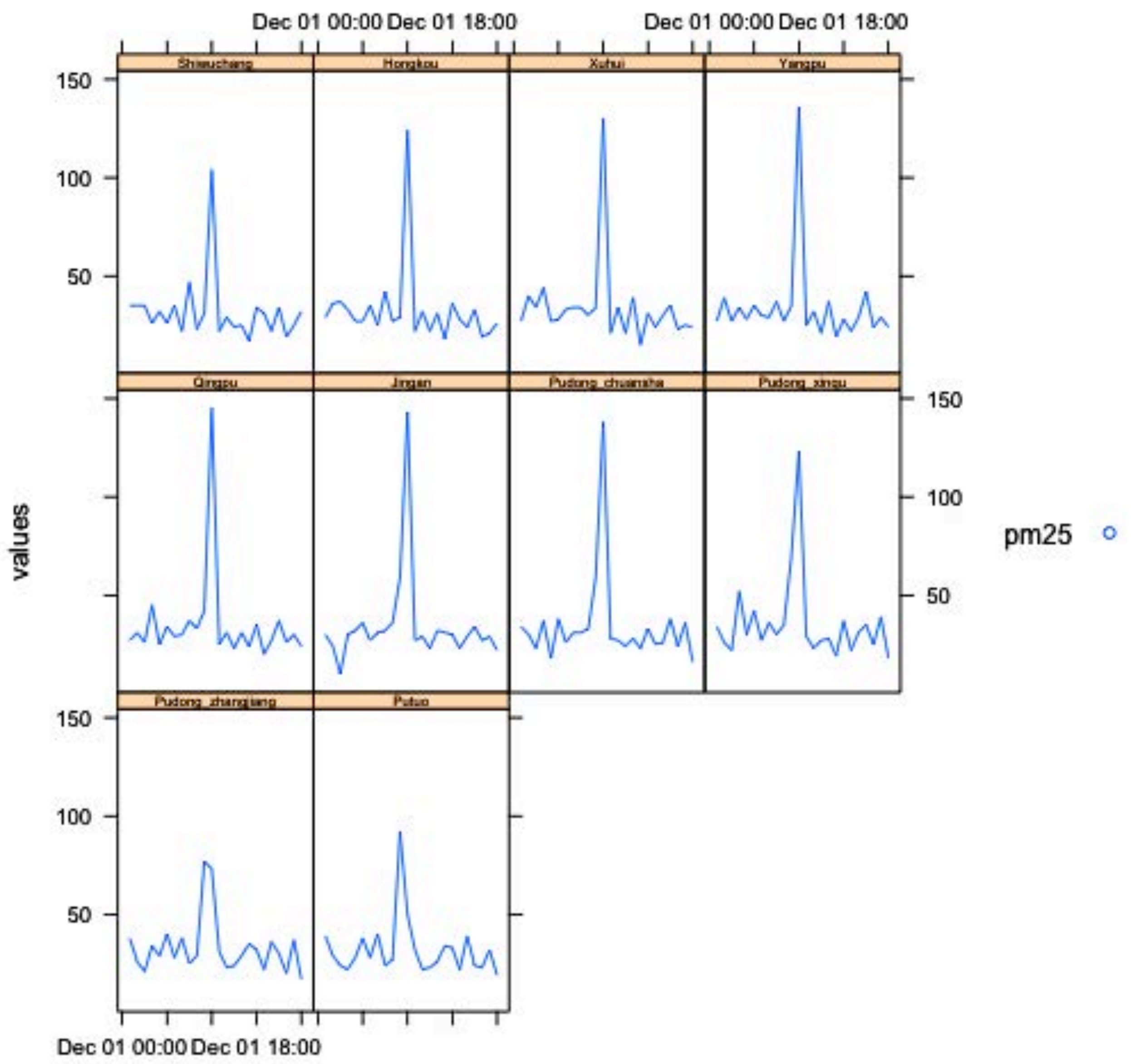
TRENDS EXPLORATION



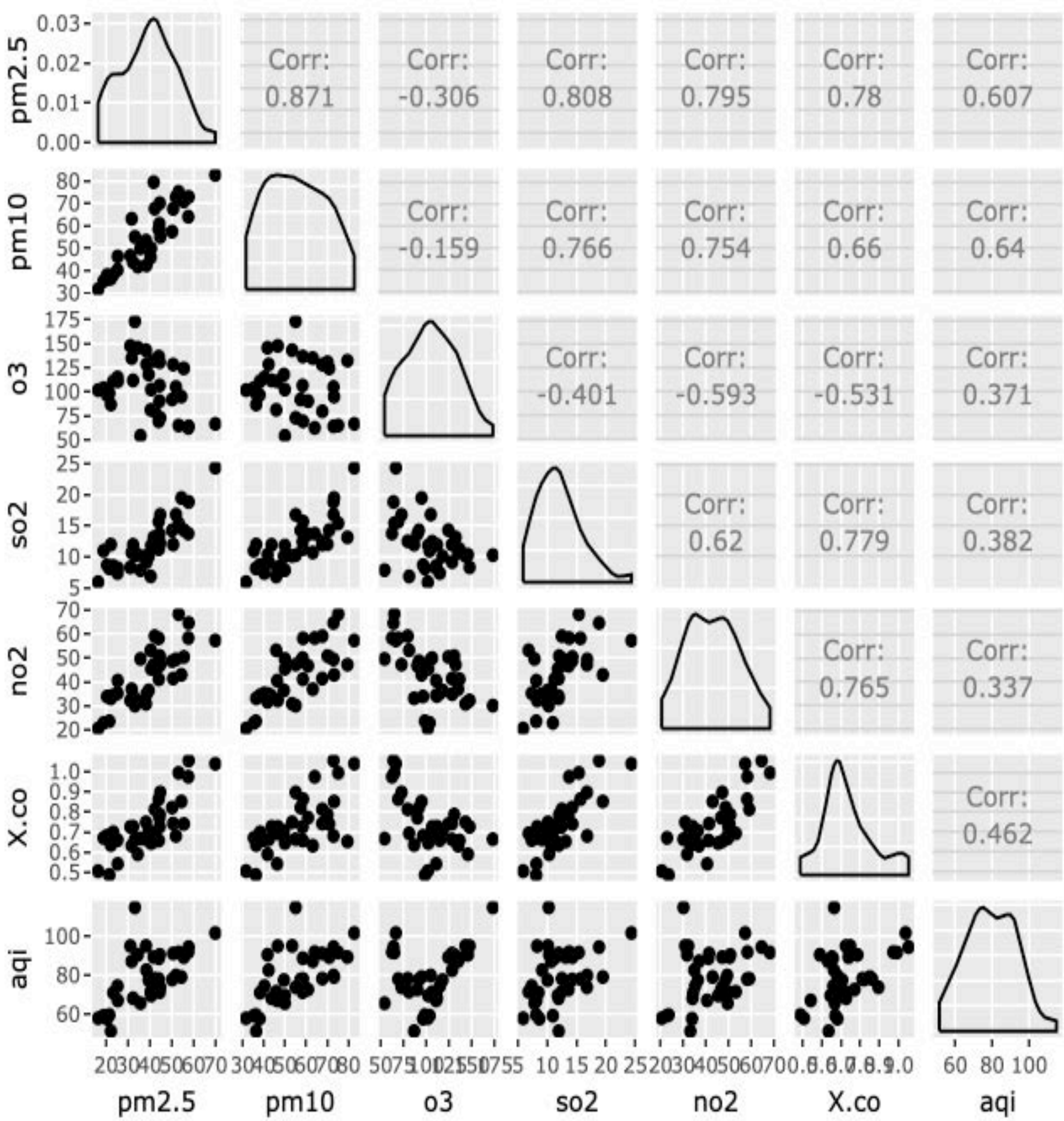


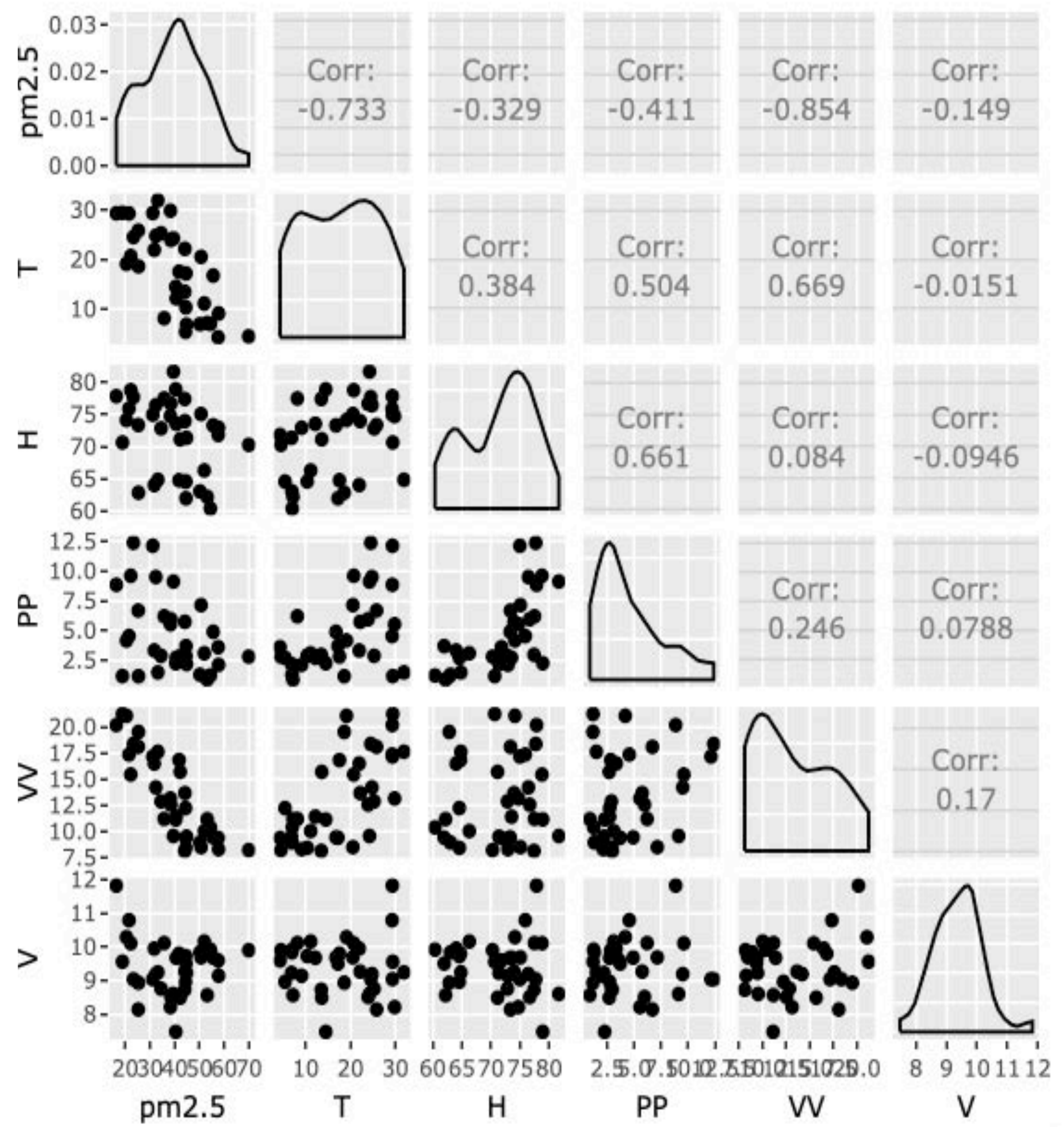
TRENDS EXPLORATION





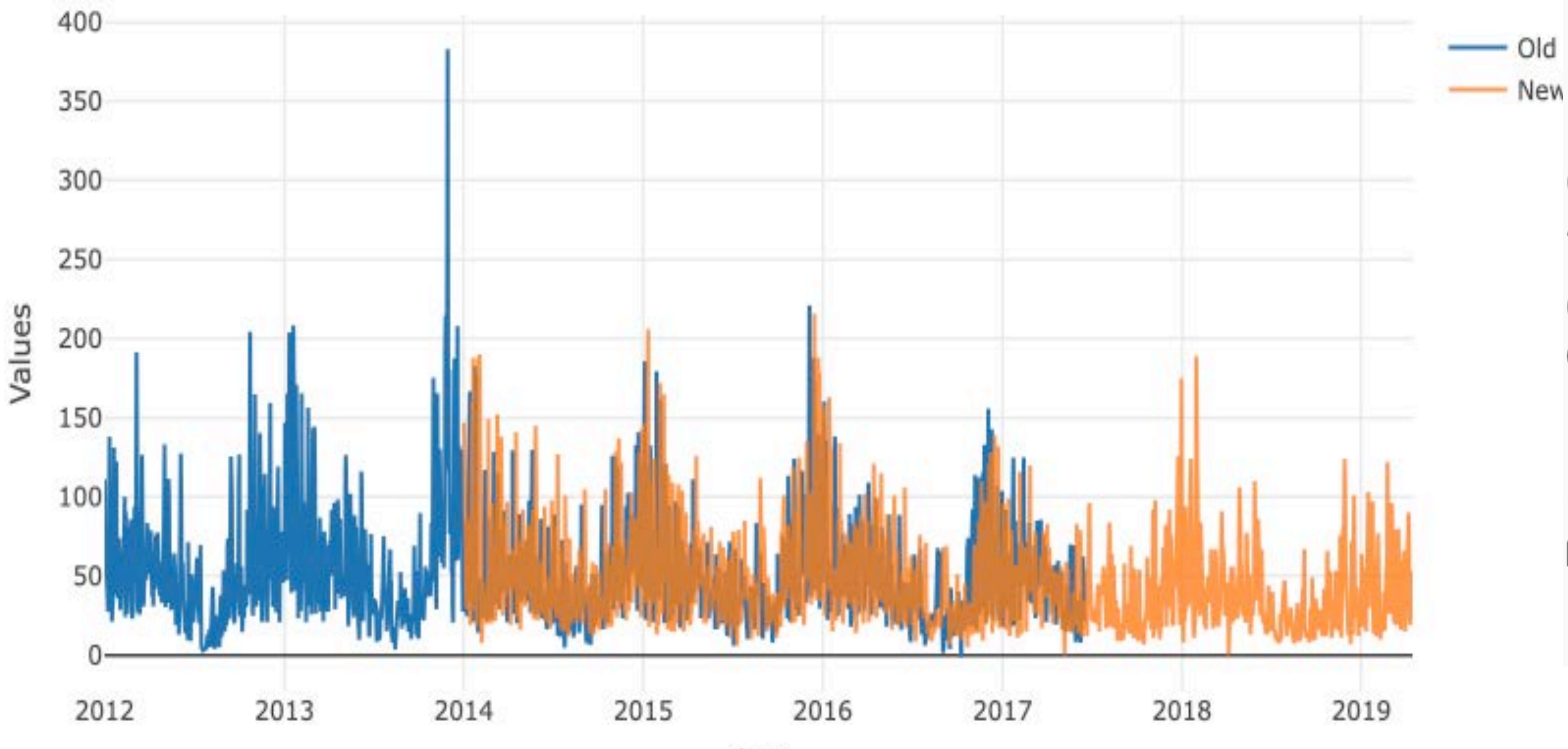
CORRELATIONS



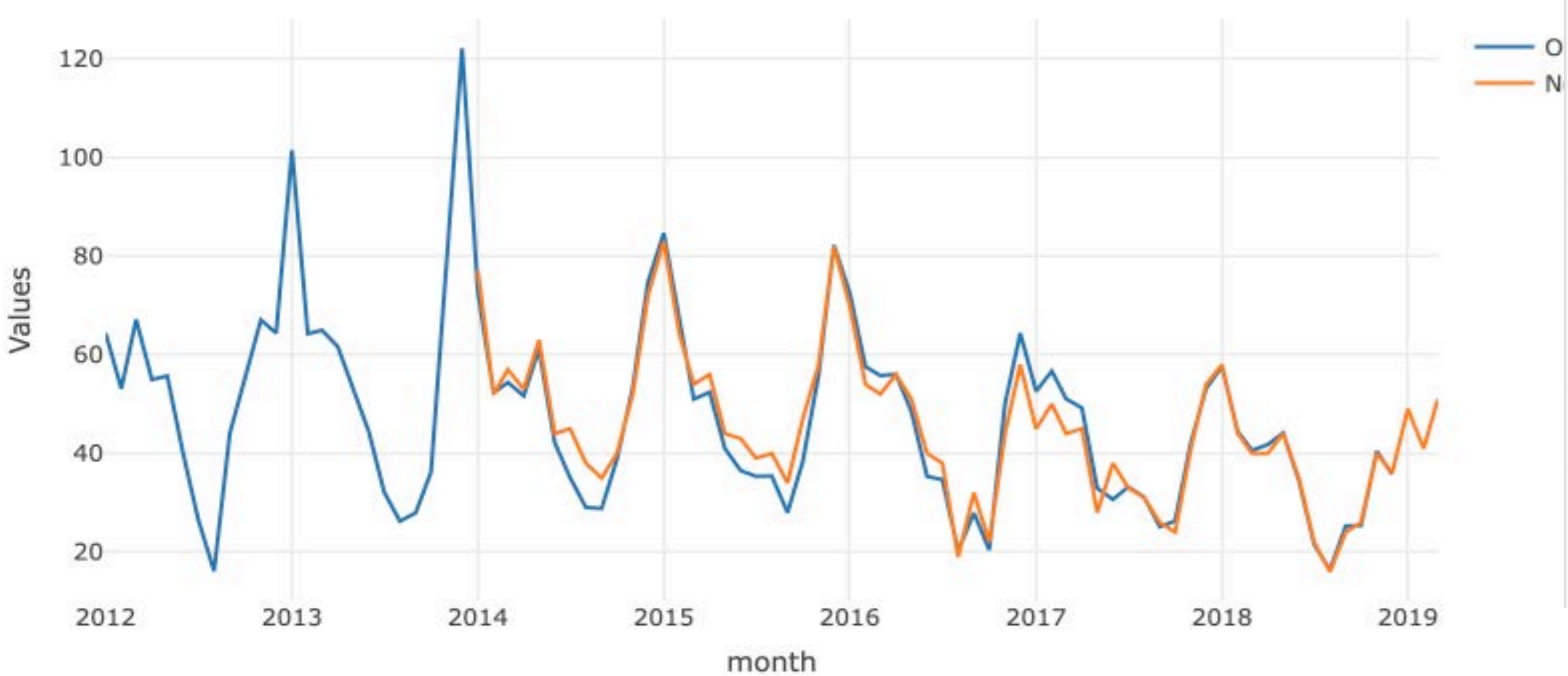



STATISTICAL ANALYSIS

COMBINE DATASET




```
        Welch Two Sample t-test

data:  ppp$pm_2.5 and ppp$pm2.5
t = -0.96686, df = 2516.7, p-value = 0.3337
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.521206  1.195526
sample estimates:
mean of x mean of y
 47.87593  49.03877
```




```
        Welch Two Sample t-test

data:  pppp$pm2.5 and pppp$pm2.5mon
t = 0.21088, df = 117.61, p-value = 0.8333
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.944353  6.122872
sample estimates:
mean of x mean of y
 44.86667  44.27741
```

TRENDS ANALYSIS

- **Linear Model with Time Series Components**
  - $y_t = S_t + T_t + R_t$
  - y_t is the data
  - S_t the seasonal component
  - T_t is the trend-cycle component
  - R_t is the remainder component

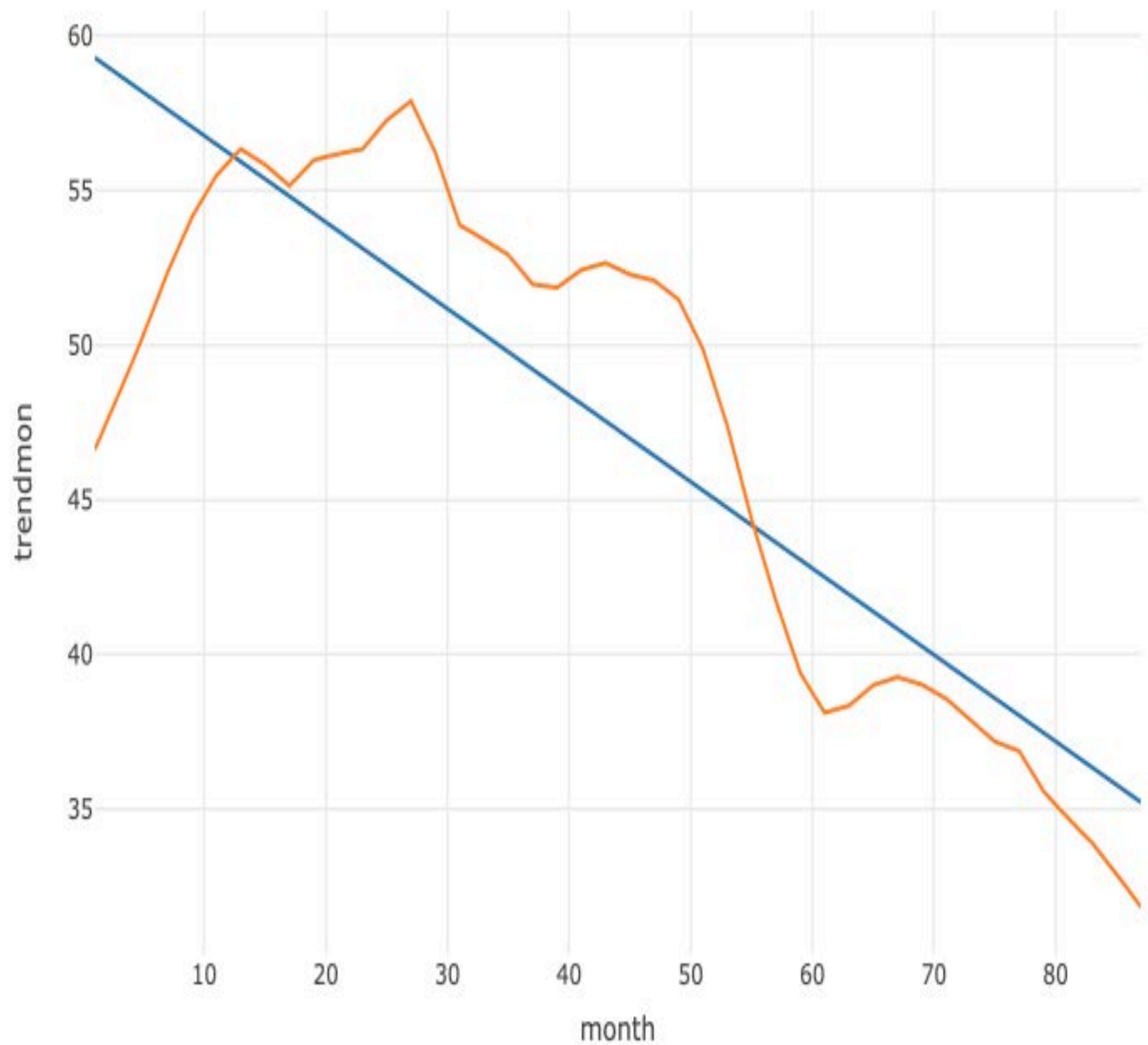
TRENDS ANALYSIS

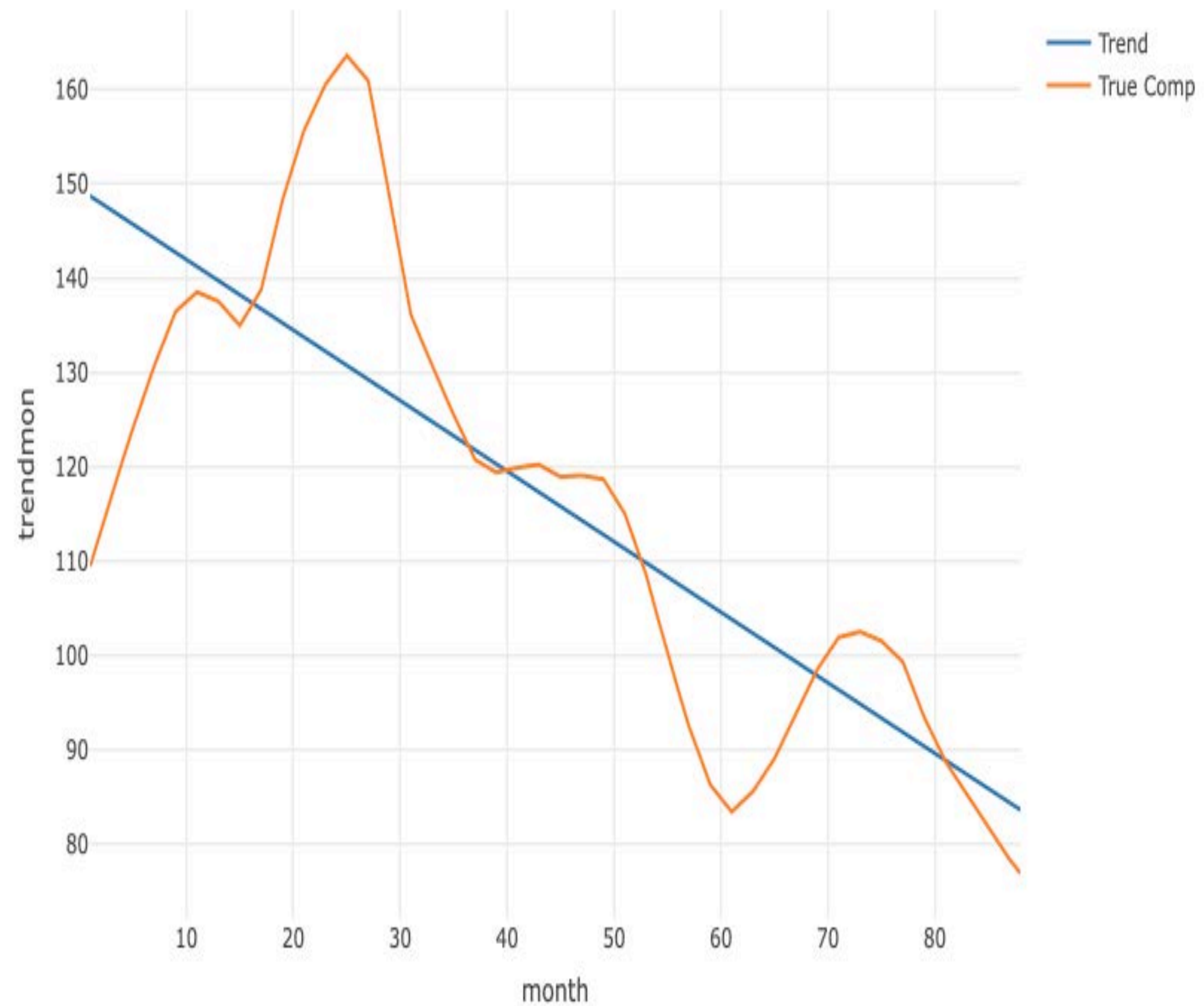
- Monthly Average

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 59.57903    0.89236   66.77   <2e-16 ***
month       -0.27989    0.01761  -15.89   <2e-16 ***
```

- Monthly Maximum

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 149.45557    3.04595   49.07   <2e-16 ***
month        -0.74775    0.05945  -12.58   <2e-16 ***
```





TRENDS ANALYSIS

- Mann-Kendall Trend Test
  - H_0: There does not exist a trend
  - H_a: There exists a downwards trend

- Monthly Average

```
Score =  -2535 , Var(Score) = 74404.34
denominator =  3741
tau = -0.678, 2-sided pvalue =< 2.22e-16
```

- Monthly Maximum

```
Score =  -2398 , Var(Score) = 76985.34
denominator =  3828
tau = -0.626, 2-sided pvalue =< 2.22e-16
```

FITTING MODELS

- MODEL 1: ARIMA
- MODEL 2: Exponential Smoothing Methods
  - Holt-Winter additive model
  - Holt-Winter multiplicative model
- MODEL 3: Generalized Additive Models
  - GAM with different smoothing methods

FITTING MODELS

MONTHLY AVERAGE	RMSE	MAE	AIC	EDF
ARIMA(1,0,0)(1,1,0)[12]	9.813071	6.60298	583.81	
HW additive	9.718209	7.387745	818.2102	
HW multiplicative	9.039098	6.855924	798.7903	
GAM1			888.0311	9.831992
GAM2			898.0336	25.454084
GAM3			896.9211	8.505505
GAM4			893.6073	8.556740
GAM5			892.8290	9.265826

FITTING MODELS

MONTHLY MAXIMUM	RMSE	MAE	AIC	EDF
ARIMA(1,0,0)(1,1,0)[12]	9.813071	6.60298	583.81	
HW additive	36.21408	25.24063	1059.748	
HW multiplicative	35.47848	25.47263	1040.291	
GAM1			1040.291	11.28402
GAM2			653.2566	28.84447
GAM3			652.0351	16.45363
GAM4			645.8053	10.92884
GAM5			647.9323	10.93213

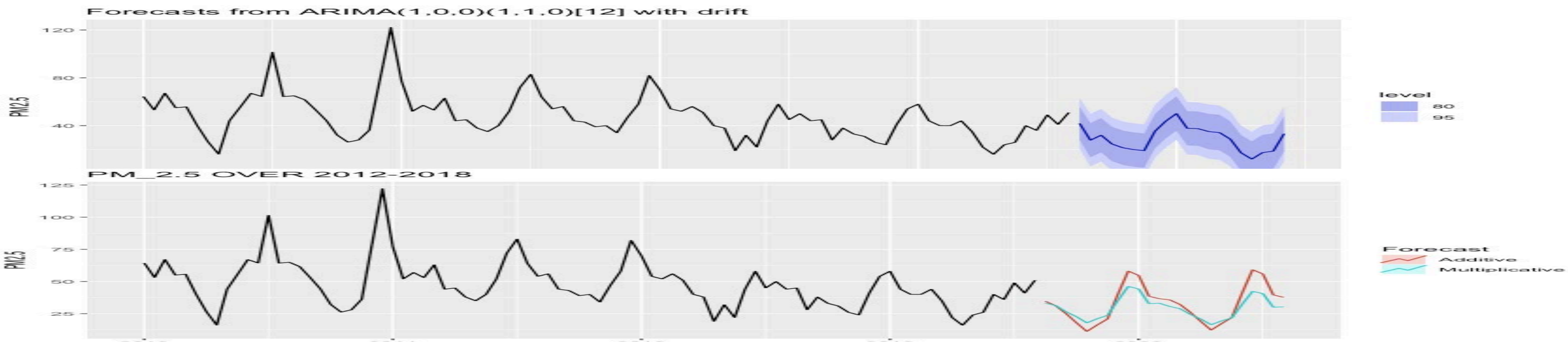
FITTING MODELS

- Daily pm2.5 and other pollutions

```
Approximate significance of smooth terms:
          edf Ref.df       F  p-value    
s(pm10) 2.330  3.020 425.412  < 2e-16 ***
s(o3)   5.551  6.702  13.308 4.77e-16 ***
s(so2)  7.188  8.005   2.508   0.0101 *  
s(no2)  4.852  5.922   6.321 1.41e-06 ***
s(co)   6.528  7.422 134.369  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.

R-sq.(adj) =  0.884   Deviance explained = 88.
GCV = 105.06  Scale est. = 103.56    n = 1931
```

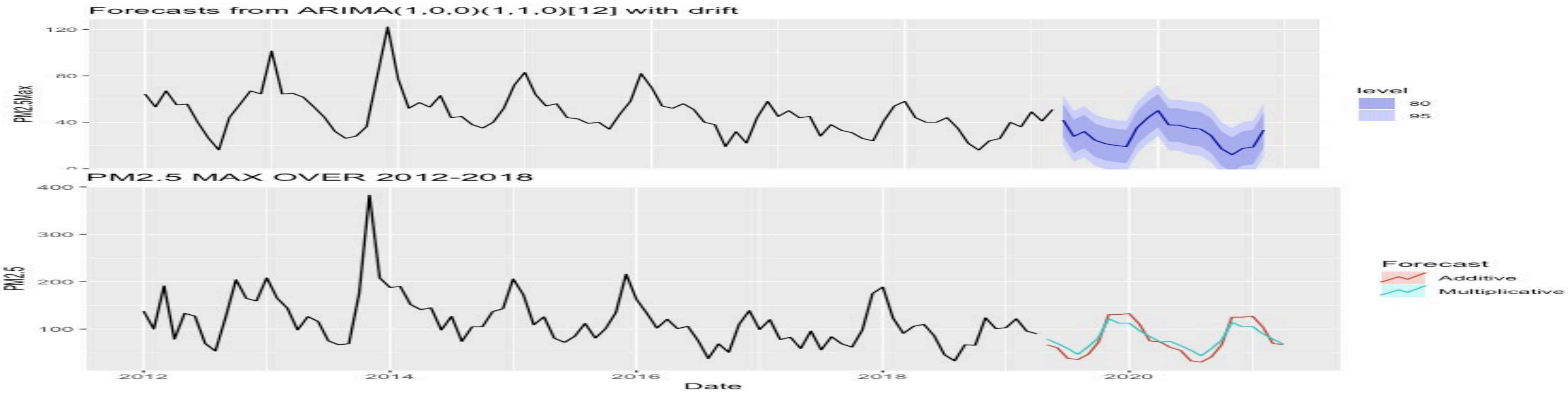
```
Approximate significance of smooth terms:
          edf Ref.df      F  p-value    
s(pm10) 3.486  4.400 81.802  < 2e-16 ***
s(o3)   2.912  3.700 16.973 2.52e-12 ***
s(so2)  5.701  6.745  2.442    0.016 *  
s(no2)  7.288  8.244  5.900 1.43e-07 ***
s(X.co) 5.388  6.514 87.409  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

R-sq.(adj) =  0.855   Deviance explained = 85.8%
GCV = 101.85  Scale est. = 99.459    n = 1096
```

FORECAST

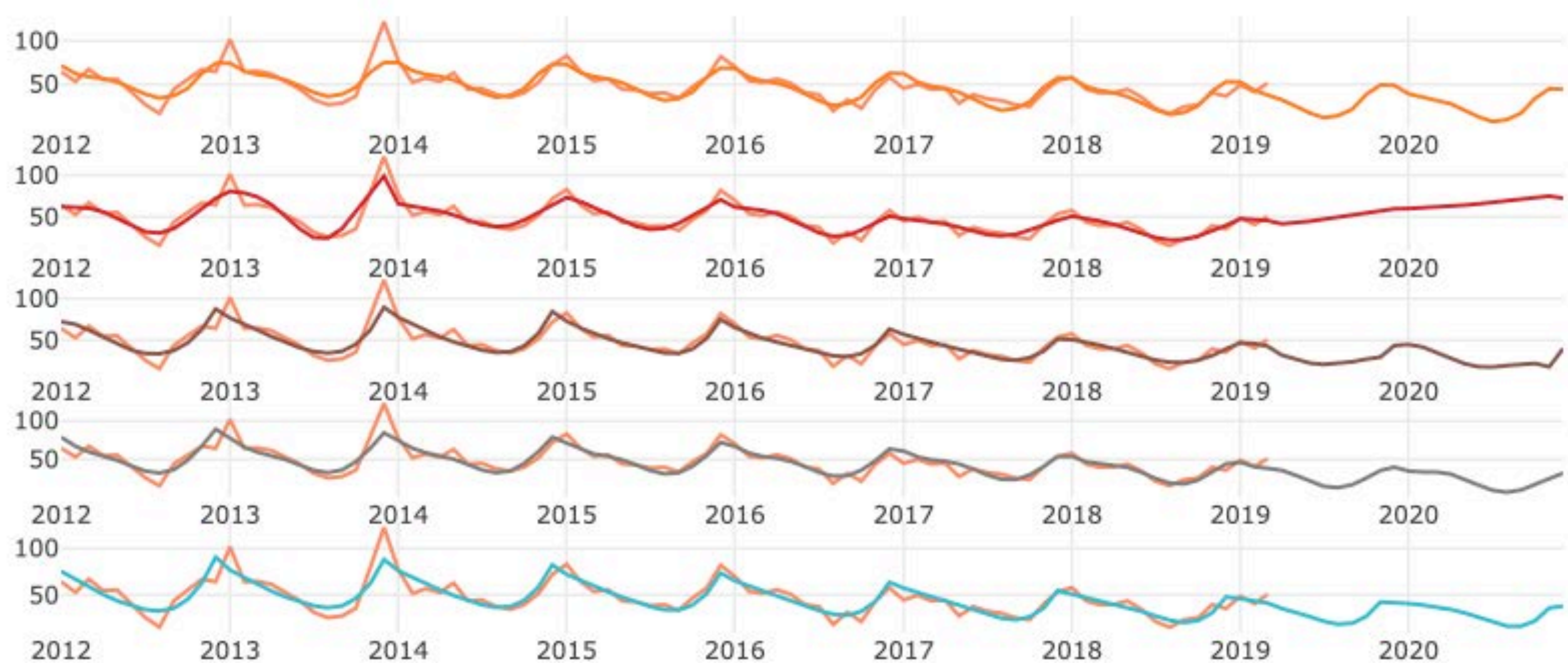

Forecasts from ARIMA(1,0,0)(1,1,0)[12] with drift
120
80
40
PM2.5
level
80
95
PM_2.5 OVER 2012-2018
125
100
75
50
25
PM2.5
Forecast
Additive
Multiplicative


100
50
2012
2013
2014
2015
2016
2017
2018
2019
2020

FORECAST

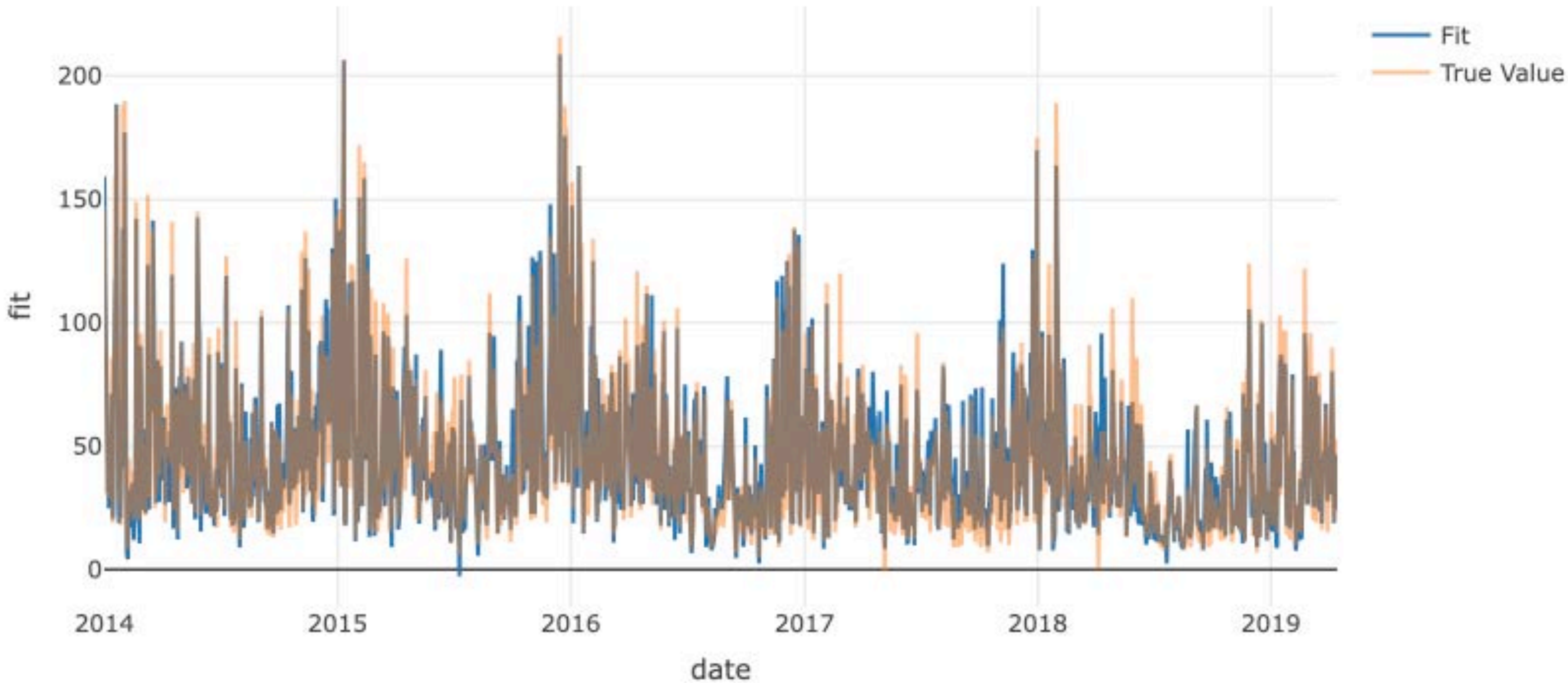


Forecasts from ARIMA(1,0,0)(1,1,0)[12] with drift
PM2.5Max
level
80
95
PM2.5 MAX OVER 2012-2018
PM2.5
Forecast
Additive
Multiplicative
Date
2012
2014
2016
2018
2020



100
50
2012
2013
2014
2015
2016
2017
2018
2019
2020

FORECAST



Fit
True Value
fit
200
150
100
50
0
2014
2015
2016
2017
2018
2019
date

SUGGESTION

REASONS FOR POOR AIR QUALITY IN WINTER

- Shanghai gets wind mainly from South East during summer months, while winds from the North dominate during winter months.
- In winter, the Northern people use more coal-burning boilers for central heating systems, so northern winds carry dirty emissions south.
- Trees density decreases in Shanghai because of losing leaves in late Autumn which can no longer trap dust.
- Straw burning in the provinces nearby during October to December

ACTIONS TAKEN TO IMPROVE SHANGHAI AIR QUALITY

- In late November 2013, Shanghai announced rules for a carbon emissions trading scheme. (industrial emissions)
- In early January 2014, Shanghai launched a joint effort rule with its three closest provinces to tackle air pollution.
- In late January 2014, Shanghai announced a ban on the burning of straw and other bonfires within all of Shanghai. (previous certain areas) (agricultural)
- Comparing the concentrate of PM2.5 between 2013.11-2014.2 and 2014.11-2015.2.

```
        Two Sample t-test

data:  groupdata$pm by groupdata$group
t = 1.9954, df = 238, p-value = 0.02357
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.121682      Inf
sample estimates:
mean in group 1 mean in group 2
       82.32067        70.01900
```

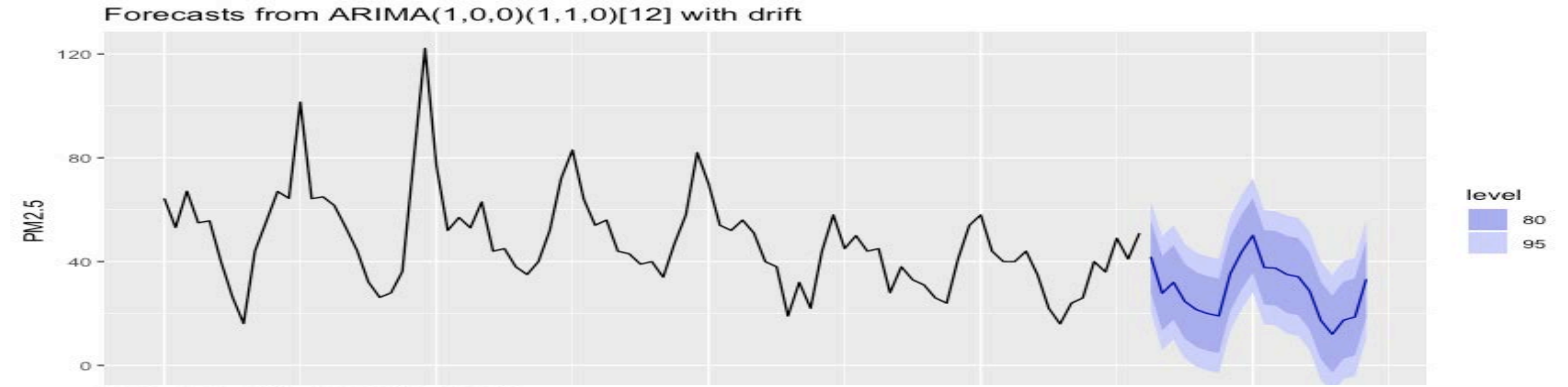
ACTIONS TAKEN TO IMPROVE SHANGHAI AIR QUALITY

- Shanghai government extended subsidies for renewable energy 'green cars'. (Buyer will get a subsidy of RMB 40,000, plus a free Shanghai license plate, worth about RMB 70,000, and also get the central government subsidy of RMB 60,000.) (cars and ships)
- In April 2014, Shanghai adopted the V emission standards for all new vehicles. (cars and ships)
- Shanghai tightened its ban on Yellow Label vehicles from outer ring roads (previous inner ring roads) and a complete ban will come in force in 2015. (cars and ships)

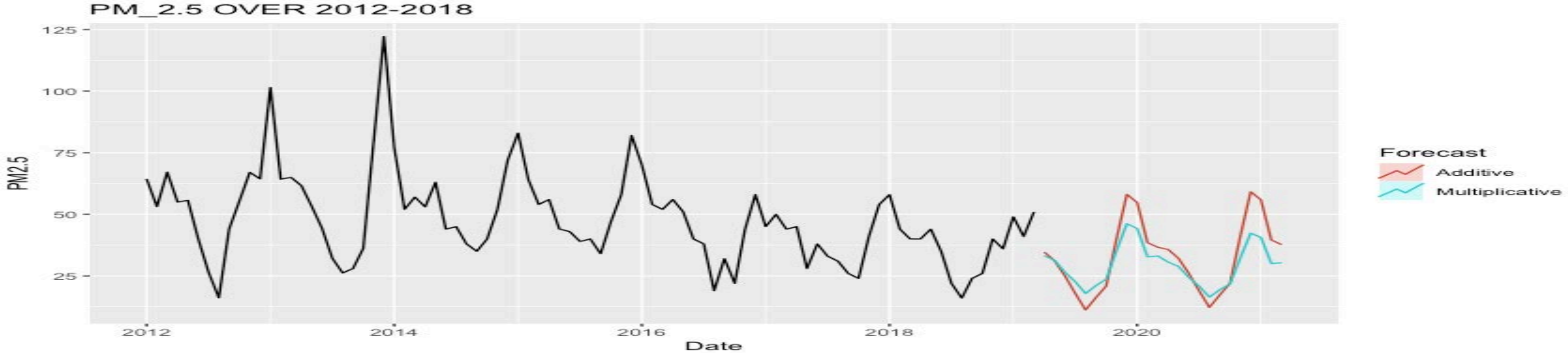
REFERENCE

- Ole Raaschou-Nielsen; et al. (July 10, 2013). "Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)". The Lancet Oncology. 14 (9): 813–22. doi:10.1016/S1470-2045(13)70279-1.
- Ghassan B. Hamra,1 Neela Guha,1 Aaron Cohen; et al. (September 2014). "Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis". Environmental Health Perspectives. 122 (9): 906–11. doi:10.1289/ehp.1408092.
- Hyndman, Rob J. & Athanasopoulos, George. & OTexts.com, issuing body. (2014). Forecasting : principles and practice. [Heathmont?, Victoria] : OTexts.com

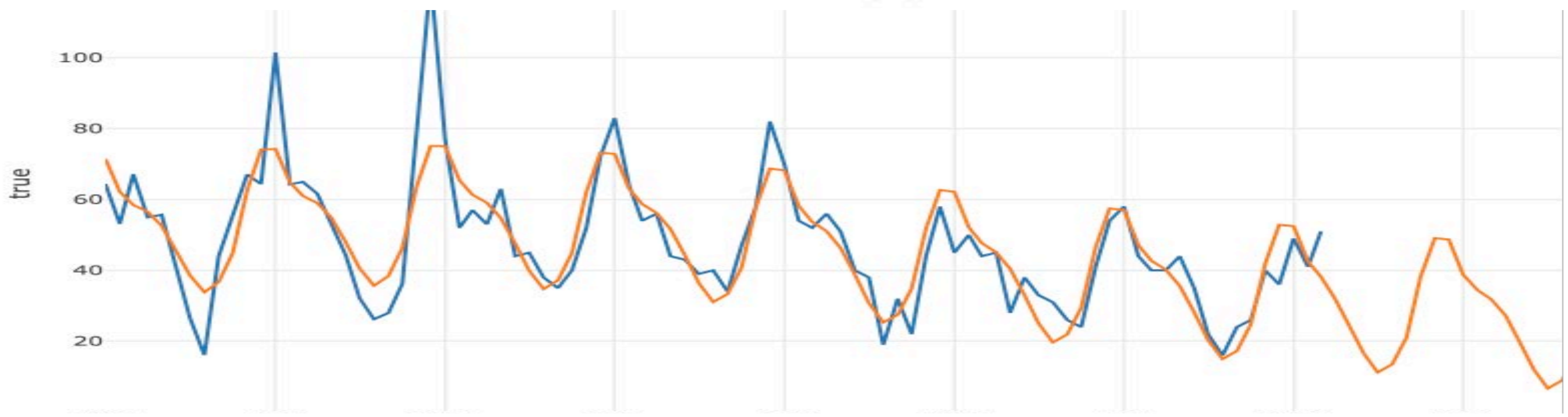
THANK YOU



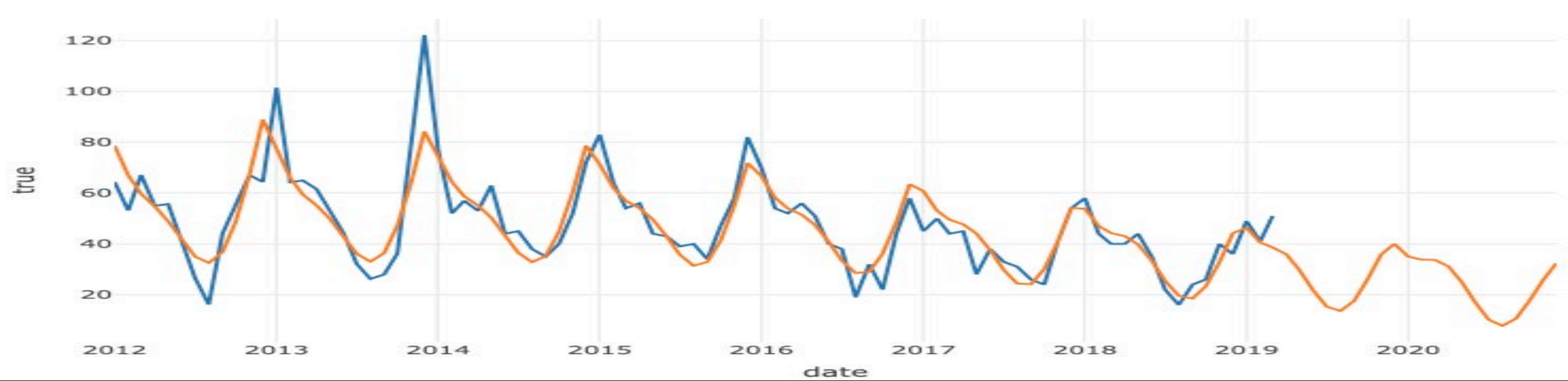
Forecasts from ARIMA(1,0,0)(1,1,0)[12] with drift
PM2.5
level
80
95



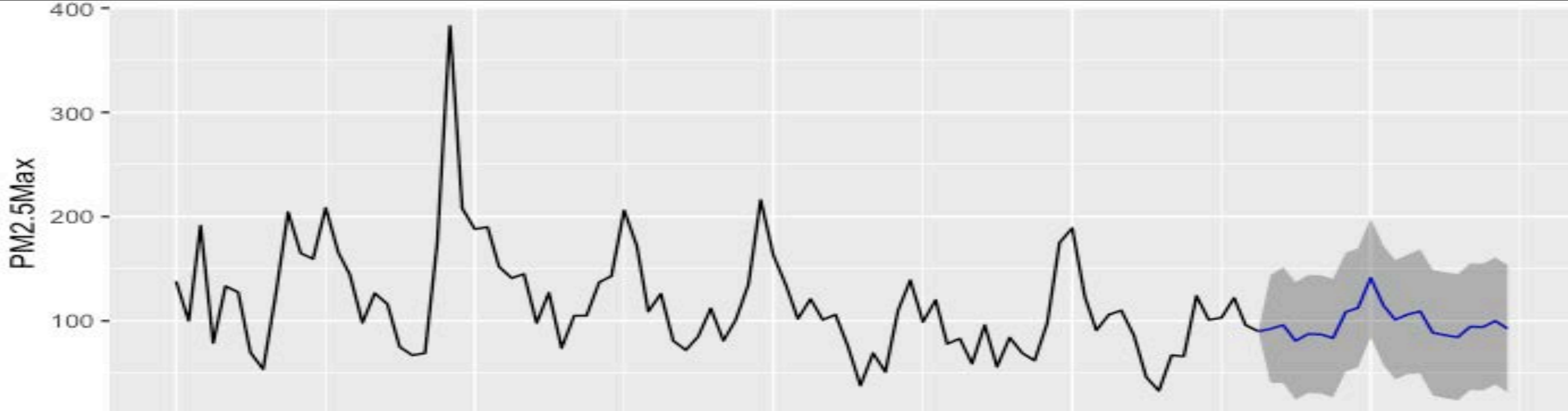
PM_2.5 OVER 2012-2018
PM2.5
Forecast
Additive
Multiplicative
Date



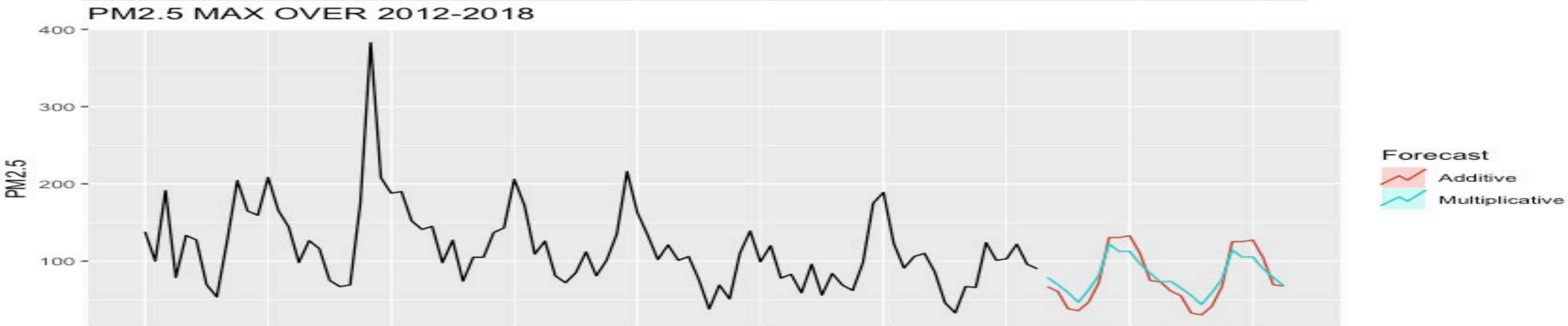
true



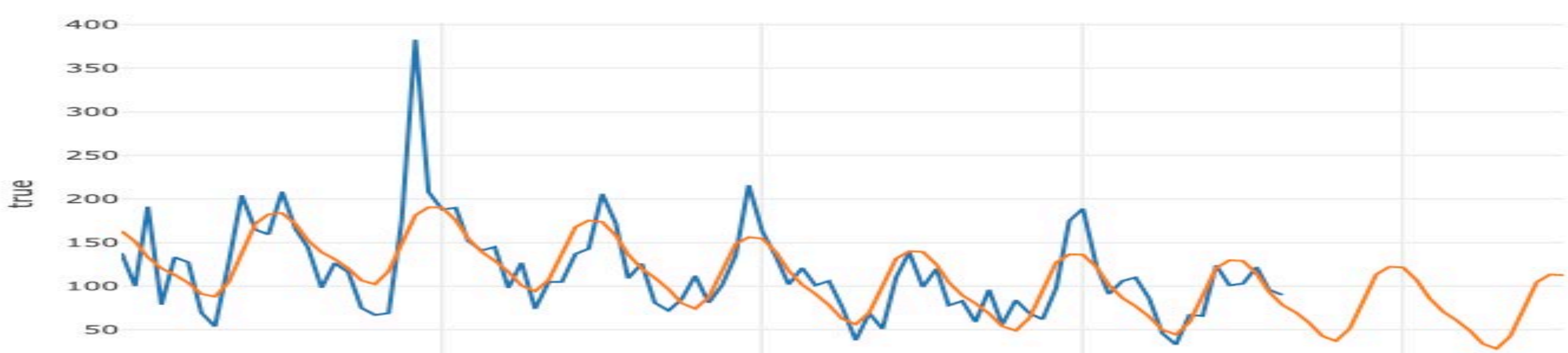
true
date



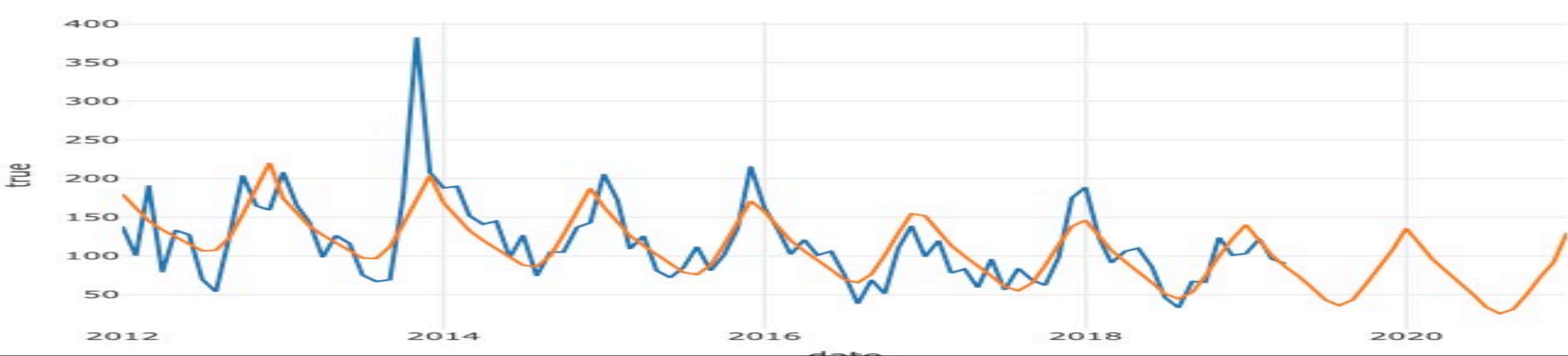
400
300
200
100
PM2.5Max



PM2.5 MAX OVER 2012-2018
400
300
200
100
PM2.5
Forecast
Additive
Multiplicative



400
350
300
250
200
150
100
50
true



400
350
300
250
200
150
100
50
true
2012
2014
2016
2018
2020