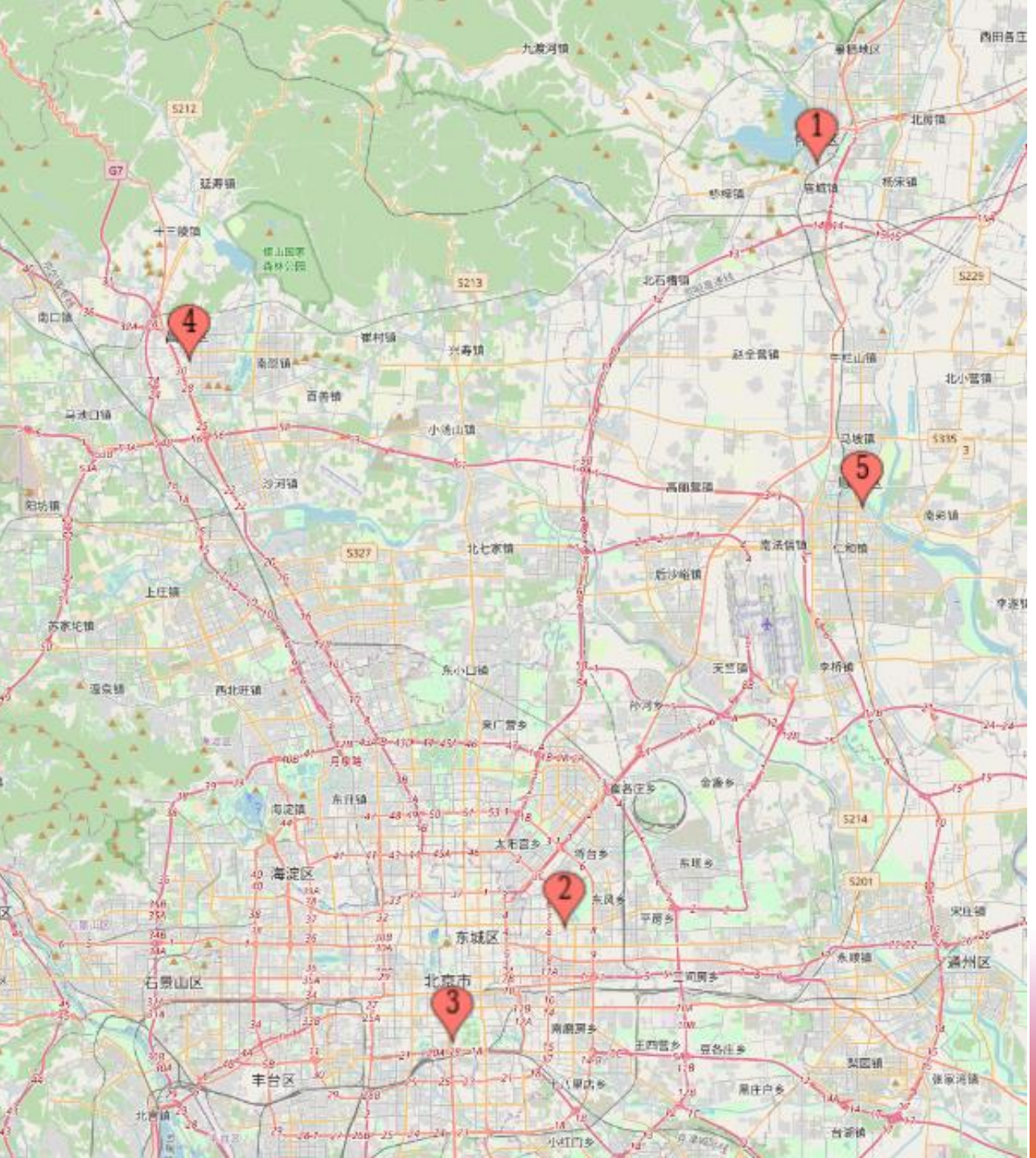
Predicting the Air Quality Index for Shunyi

Melissa Rajamanuvel



Introduction

- Goal: To predict the Air Quality Index (AQI) for a city in Beijing (Shunyi) using the AQI of four neighboring cities AQI (Changping, Huairou, Nongzhanguan, Tiantan).

No	year	month	day	hour	PM2.5	PM10	SO2	NO2	CO	O3	TEMP	PRES	DEWP	RAIN	wd	WSPM	station
1	2013	3	1	0	7	7	3	2	100	91	-2.3	1020.3	-20.7	0	WNW	3.1	Huairou
2	2013	3	1	1	4	4	3	NA	100	92	-2.7	1020.8	-20.5	0	NNW	1.5	Huairou
3	2013	3	1	2	4	4	NA	NA	100	91	-3.2	1020.6	-21.4	0	NW	1.8	Huairou
4	2013	3	1	3	3	3	3	2	NA	NA	-3.3	1021.3	-23.7	0	NNW	2.4	Huairou
5	2013	3	1	4	3	3	7	NA	300	86	-4.1	1022.1	-22.7	0	NNW	2.2	Huairou
6	2013	3	1	5	4	4	3	3	200	85	-4.2	1022.3	-24.5	0	N	4.3	Huairou
7	2013	3	1	6	3	6	33	7	300	82	-5.9	1023.1	-21.9	0	WNW	0.6	Huairou
8	2013	3	1	7	3	10	13	13	400	71	-2.7	1024.3	-23.2	0	NNE	3.4	Huairou
9	2013	3	1	8	3	13	34	38	800	45	-1.6	1025.2	-23.5	0	NNE	4.6	Huairou
10	2013	3	1	9	17	36	50	28	700	60	-1.1	1025.4	-23.8	0	NE	4.9	Huairou
11	2013	3	1	10	19	32	5	3	300	85	1	1025.1	-23.6	0	ESE	2.7	Huairou
12	2013	3	1	11	10	11	3	NA	300	87	2.4	1024.5	-22.5	0	SE	3.9	Huairou
13	2013	3	1	12	3	6	5	3	300	86	3.9	1023.8	-21.3	0	W	1.9	Huairou
14	2013	3	1	13	3	6	7	3	300	87	4.7	1022.8	-22.4	0	S	3.3	Huairou
15	2013	3	1	14	3	6	9	6	300	87	5.7	1022.3	-22.6	0	NW	1.9	Huairou

Context

- This Dataset is from Kaggle.
- We used the Air Quality Data of 5 cities in Beijing (Changping, Huairou, Nongzhanguan, Tiantan, Shunyi) which has the records for the air quality from 2013 to 2017.
- The original data had 18 columns: **No** (number), **year**, **month**, **day**, **hour**, **$PM_{2.5}$**, **PM_{10}**, **SO2** (sulfur dioxide), **NO2** (nitrogen dioxide), **CO** (carbon monoxide), **O3** (Ozone), **Temperature**, **Pressure**, **Dew Point Temp**, **RAIN**, **wind direction**, **Wind speed** (per minute), and **station** (city).
  - PM = Particulate Matter, which is a mix of solid or liquid particles (such as dirt or dust) in the air

These Breakpoints…							…equal this AQI	…and this category
O_3 (ppm) 8-hour	O_3 (ppm) 1-hour[1]	$PM_{2.5}$ ($\mu g/m^3$) 24-hour	PM_{10} ($\mu g/m^3$) 24-hour	CO (ppm) 8-hour	SO_2 (ppb) 1-hour	NO_2 (ppb) 1-hour	AQI	
0.000 - 0.054	-	0.0 – 12.0	0 - 54	0.0 - 4.4	0 - 35	0 - 53	0 - 50	Good
0.055 - 0.070	-	12.1 – 35.4	55 - 154	4.5 - 9.4	36 - 75	54 - 100	51 - 100	Moderate
0.071 - 0.085	0.125 - 0.164	35.5 – 55.4	155 - 254	9.5 - 12.4	76 - 185	101 - 360	101 - 150	Unhealthy for Sensitive Groups
0.086 - 0.105	0.165 - 0.204	(55.5 - 150.4)[3]	255 - 354	12.5 - 15.4	(186 - 304)[4]	361 - 649	151 - 200	Unhealthy
0.106 - 0.200	0.205 - 0.404	(150.5 - (250.4)[3]	355 - 424	15.5 - 30.4	(305 - 604)[4]	650 - 1249	201 - 300	Very unhealthy
([2])	0.405 - 0.504	(250.5 - (350.4)[3]	425 - 504	30.5 - 40.4	(605 - 804)[4]	1250 - 1649	301 - 400	Hazardous
([2])	0.505 - 0.604	(350.5 - 500.4)[3]	505 - 604	40.5 - 50.4	(805 - 1004)[4]	1650 - 2049	401 - 500	Hazardous

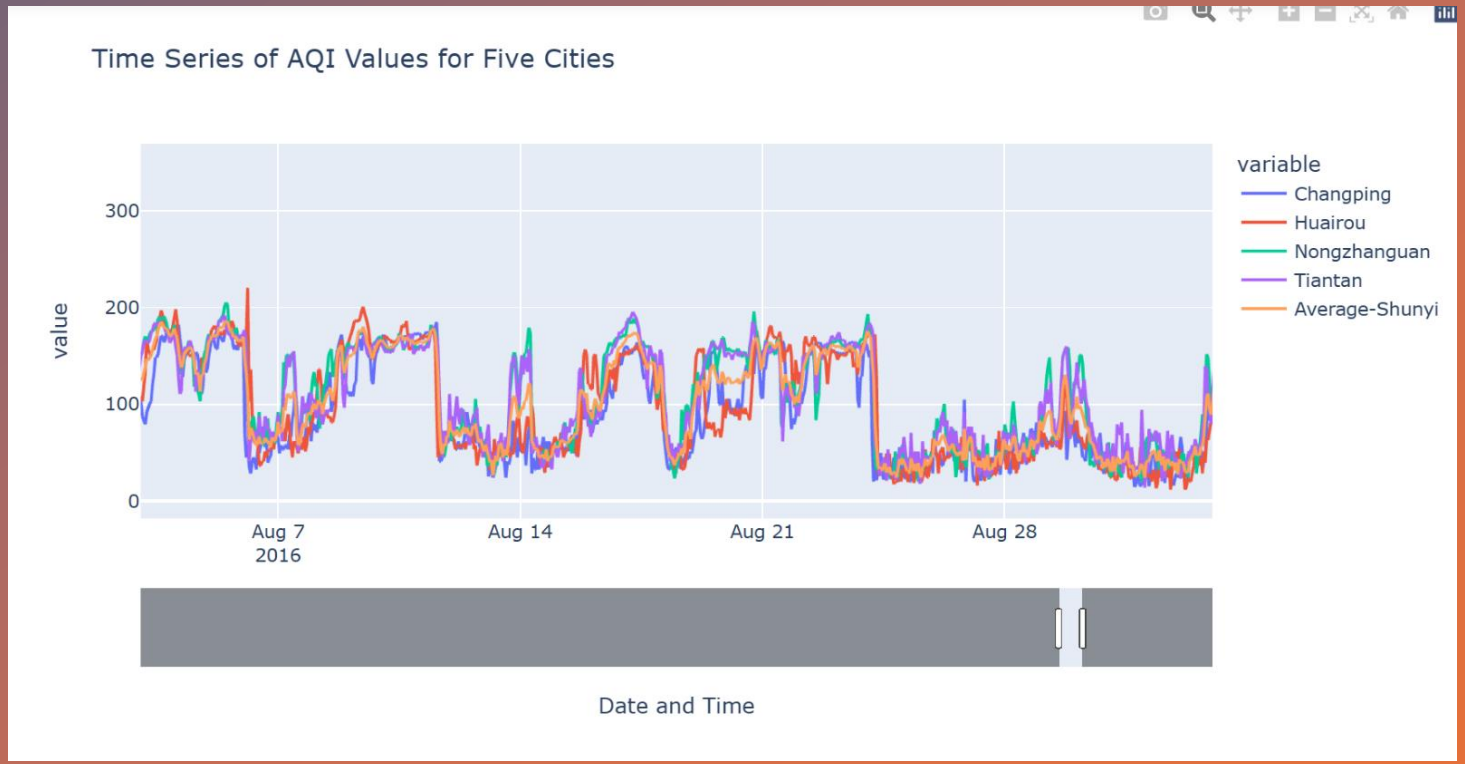
Data Cleaning

- We had a lot of cleaning of our data:
  - Finding the AQI index values using the pollutants in our data.
  - Altered the Units of the Pollutants so that it matches the formula while calculating the AQI Index.
  - Combining the year, month, day, hour, variables into one Date and Time variable.
  - Filling the Missing values using Moving Average
  - Combining the the data of all the five cities into one dataset.
- The first step in our data cleaning process was to have a column of the air quality index. We then calculated the air quality index by using the standard way used in the airnow.gov website.

Model 1: Average of Time Series

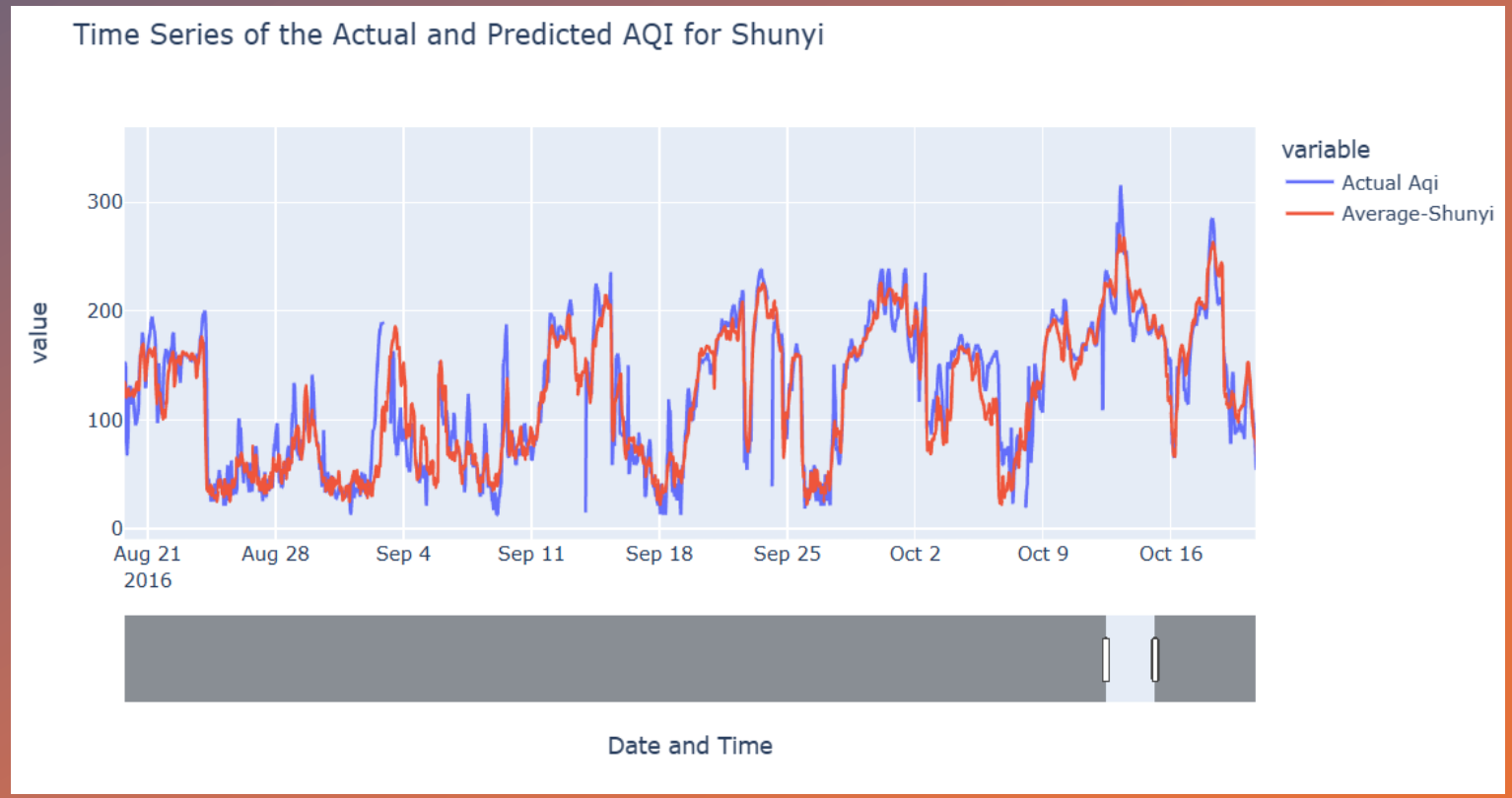
- Using the **Date and Time** and the AQI index for the four cities, we created a Time Series plot which showed us similar patterns of the AQI index for Huairou, Nongzhanguan, Tiantan and Changping.

- This showed us visually that the neighboring cities could have similar AQI indexes, so we took the Average of the four cites to predict the AQI of the Shunyi city.



Model 1: Average of Time Series

- The graph here shows the result of our prediction.
- This model did a really good job in predicting the AQI of Shunyi with:
  - **RMSE (Root Mean Square Error) = 28.22**
  - **R-squared value = 83.33%**



Model 2: Multiple Regression Analysis

- Using Multiple Regression Analysis, we tried out a many predictor subsets of how to get a better prediction of the Shunyi AQI.

- **Subset 1**: Huairou, Nongzhanguan, Tiantan and Changping
  - **R-squared = 82.23%**

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
29.1749	82.24%	82.23%	82.23%

- **Subset 2**: Huairou, Nongzhanguan, Tiantan, Changping, Predicted Shunyi (using Average)
  - **R-squared = 84.71%**

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
27.0620	84.72%	84.71%	84.71%

- **Subset 3**: Huairou, Predicted Shunyi (using Average)
  - **R-squared = 84.33%**

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
27.3849	84.33%	84.33%	84.32%

Response is Actual-Shunyi AQI

34250 cases used, 745 cases contain missing values

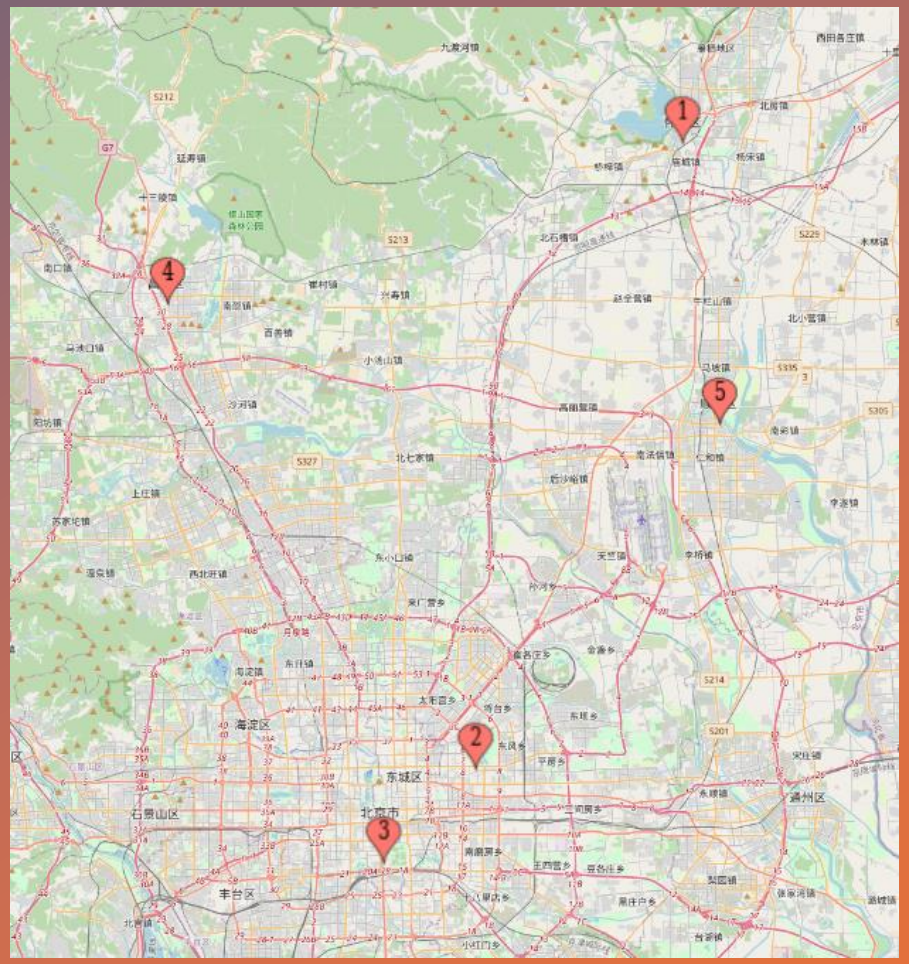
Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Changping-AQI	Huairou-AQI	Nongzhanguan-AQI	Tiantan-AQI	Predicted-Shunyi-AQI
1	83.4	83.4	83.4	2954.7	28.204					X
1	75.2	75.2	75.1	21430.7	34.505		X			
2	84.3	84.3	84.3	849.7	27.394		X			X
2	83.6	83.6	83.6	2563.3	28.055			X		X
3	84.6	84.6	84.6	345.9	27.197	X	X			X
3	84.5	84.5	84.5	435.0	27.232		X	X		X
4	84.7	84.7	84.7	17.1	27.067	X	X	X		X
4	84.6	84.6	84.6	168.9	27.127	X	X		X	X
5	84.7	84.7	84.7	6.0	27.062	X	X	X	X	X

Model 2: Multiple Regression Analysis Using Wind Direction

We then filtered out the data using the WD of the city that points to Shunyi.

Variable: wd (Wind Direction)

- **NA** - Not applicable (NA is not a standard wind direction)
- **E** - East
- **ENE** - East-Northeast
- **ESE**- East-Southeast
- **N** - North
- **NE** - Northeast
- **NNE** - North-Northeast
- **NNW** - North-Northwest
- **NW** - Northwest
- **S** - South
- **SE** - Southeast
- **SSE** - South-Southeast
- **SSW** - South-Southwest
- **SW** - Southwest
- **W** - West
- **WNW** - West-Northwest
- **WSW** - West-Southwest



City 1: Huairou, **City 2**: Nongzhanguan, **City 3**: Tiantan, **City 4**: Changping, **City 5**: Shunyi

Model 2: Multiple Regression Analysis Using Wind Direction

- **Subset 1**: Huairou, Nongzhanguan, Tiantan, Changping, Predicted Shunyi (using Average)
  - **R-squared = 88.51%**

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
24.1949	88.51%	87.58%	83.69%

Response is Actual-Shunyi AQI

68 cases used, 4 cases contain missing values

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Changping-AQI	Huairou-AQI	Nongzhanguan-AQI	Tiantan-AQI	Predicted-Shunyi AQI
1	85.2	85.0	84.3	15.9	26.621					X
1	84.0	83.8	83.1	22.3	27.668			X		
2	87.7	87.4	85.9	4.2	24.410			X		X
2	86.4	86.0	84.7	11.4	25.702		X	X		
3	87.9	87.4	83.9	5.0	24.392		X	X		X
3	87.9	87.4	85.7	5.1	24.408	X		X		X
4	88.4	87.6	84.0	4.7	24.134	X	X	X		X
4	88.1	87.4	85.7	6.1	24.415	X		X	X	X
5	88.5	87.6	83.7	6.0	24.195	X	X	X	X	X