Shanghai Smog

Major cities in China have been infamous for having high air pollution. Since major events such as the Olympics, and the World Expo, China has been implementing more changes to reduce the pollution. The data is from the US Department of State tracks hourly air samples from Shanghai.

Introduction/Motivation

I saw the "beautiful" graph shown on Reddit[1] but felt that the visualization did not clearly illustrate that there was a decrease in smog, only a negative change in yearly smog particle concentration peaks. To test the hypothesis, I plotted the same data[2], ran a trendline and listed the results below.

Exploration and Results

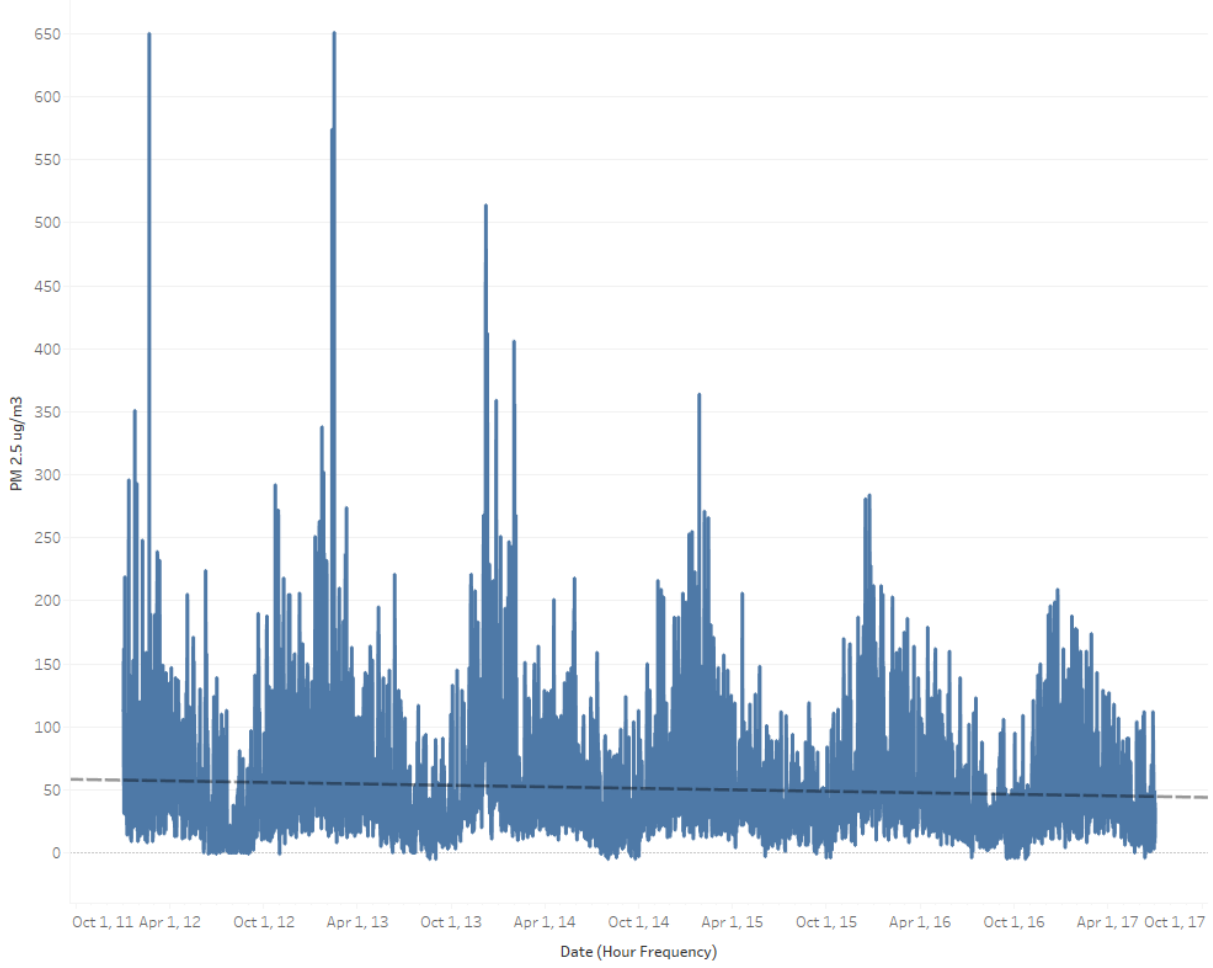


Figure 1 Shanghai PM 2.5 Levels by Hour Frequency

[1] https://www.reddit.com/r/dataisbeautiful/comments/7o1i0v/decrease_in_shanghai_air_pollution_over_the_past/

[2] http://www.stateair.net/web/historical/1/1.html

The trendline has a coefficient decline of -0.0064767 as shown in Table 1. Given the large absolute t-value of -20.4228, the coefficient is significant with more than a 99% confidence. This shows that the Redditor's statement was correct that the pollution in Shanghai is improving. Since the term is in hour of date, it is unclear as to what the trend represents so I repeat the results comparing the averages and maximums by week, month, and quarters.

In addition to the linear regression, I look at the exponential regression for the same trends because I expect a decreasing efficiency for the cleaning of air. The minimum regression values are calculated but are not analyzed because they provide no insights for change. The minimum values are included to visually illustrate 'pinches' where the pollution levels are at an extreme.

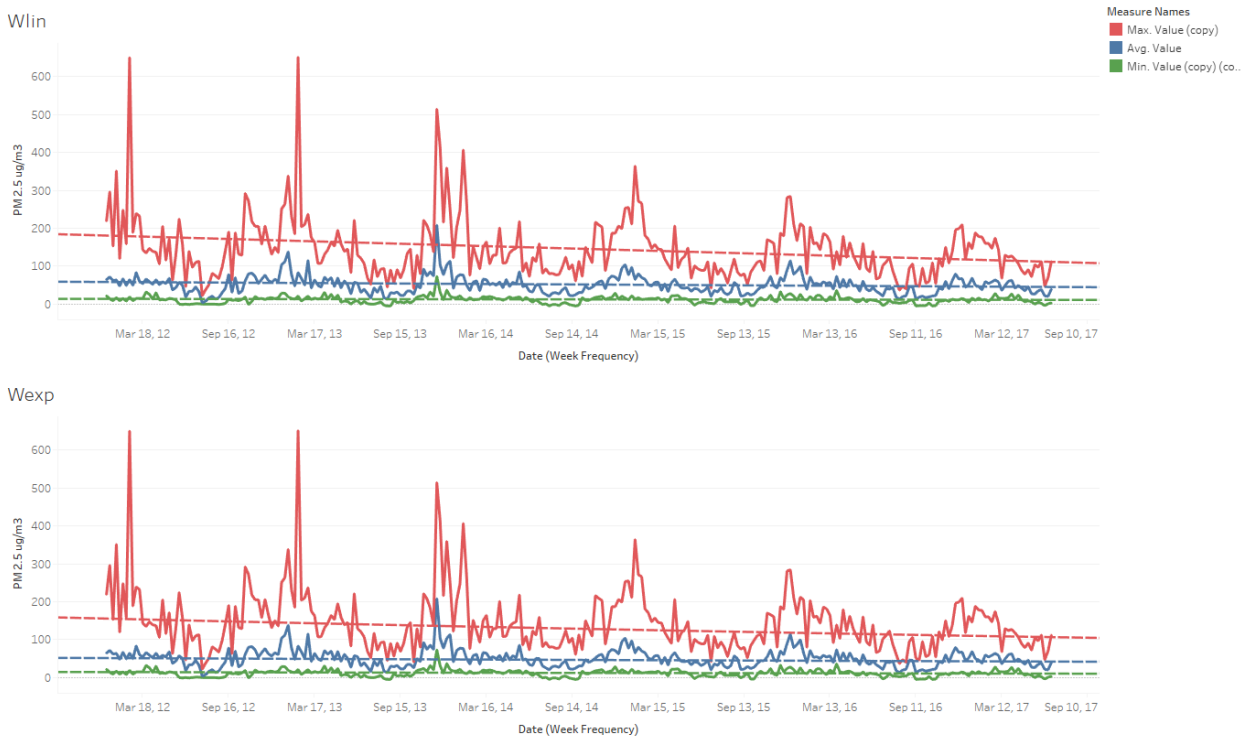


Figure 2 Shanghai PM 2.5 Levels by Week Frequency

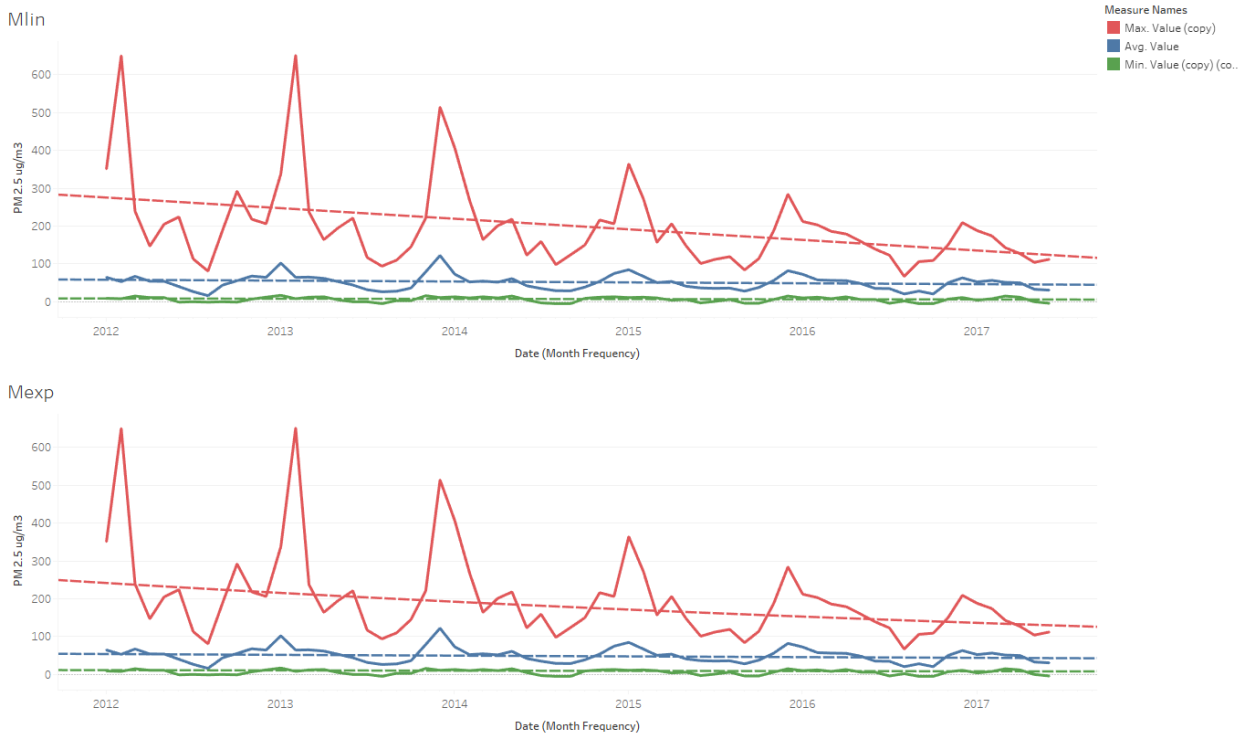


Figure 3 Shanghai PM 2.5 Levels by Month Frequency

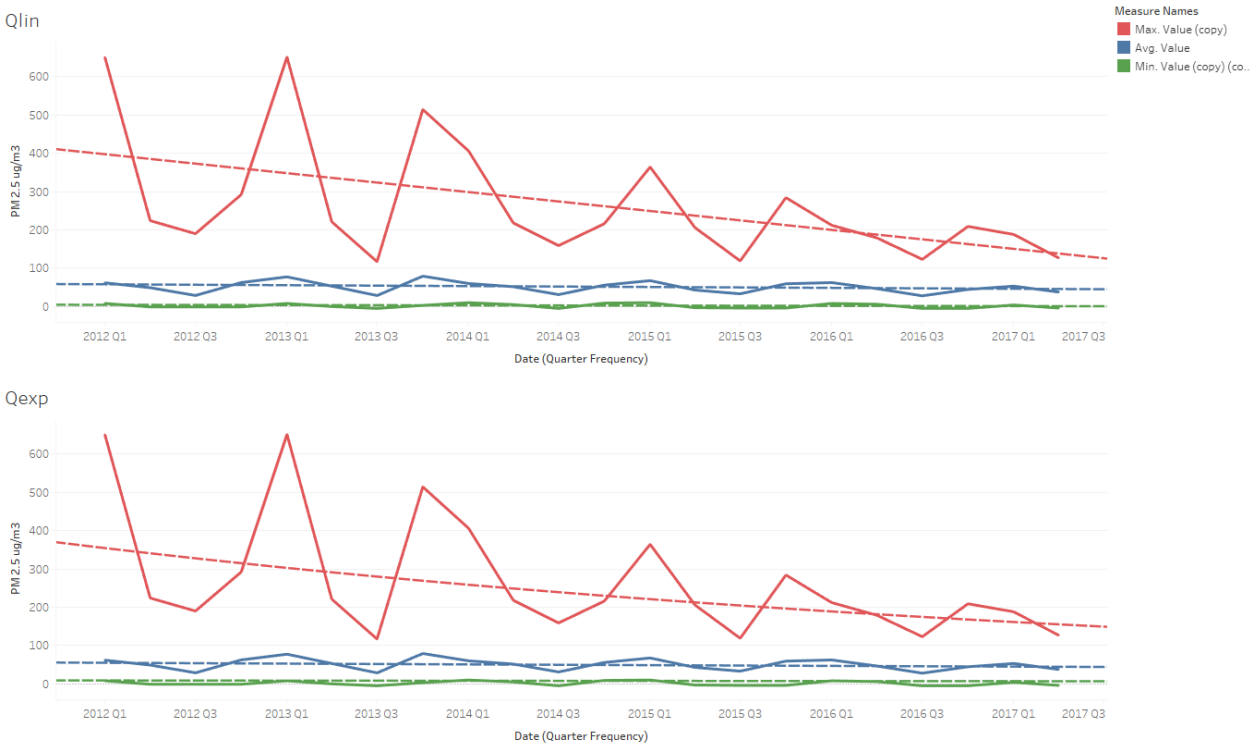


Figure 4 Shanghai PM 2.5 Levels by Quarter Frequency

As shown by the figures above, the trending decrease in air pollution becomes more defined as we view the regression in larger time intervals. The two following tables below provide numerical evidence of the decreasing trends as denoted by the negative coefficients for the PM 2.5 Values.

Type	Term	Value	StdErr	t-value	p-value	Significant
	Hour of Date	-0.0064767	0.0003171	-20.4228	< 0.0001	99.9%
Avg	Week of Date	-0.0065373	0.0023682	-2.76041	0.0061473	99%
Max	Week of Date	-0.0346079	0.0080522	-4.29794	< 0.0001	99.9%
Avg	Month of Date	-0.0063956	0.0040748	-1.56955	0.121452	80%
Max	Month of Date	-0.0765785	0.0225836	-3.39088	0.0011981	99.8%
Avg	Quarter of Date	-0.0063921	0.0055636	-1.14892	0.264143	70%
Max	Quarter of Date	-0.135168	0.0510845	-2.64596	0.0155018	98%

Table 1 Linear Regression Results

Type	Term	Value	StdErr	t-value	p-value	Significant
Avg	Week of Date	-0.0001013	4.996e-05	-2.02653	0.043641	95%
Max	Week of Date	-0.0001904	5.01e-05	-3.79946	0.0001773	99.9%
Avg	Month of Date	-0.0001154	8.273e-05	-1.39518	0.167785	85%
Max	Month of Date	-0.0003154	9.21e-05	-3.42418	0.0010808	99.5%
Avg	Quarter of Date	-0.0001121	0.0001185	-0.946262	0.355308	60%
Max	Quarter of Date	-0.0004325	0.0001647	-2.62528	0.0162141	98%

Table 2 Exponential Regression Results

The T-values will be compared as two-tails since the air pollution concentration could have increased or decreased. From the tables above, several of the average concentration regressions were not significant to the 95% level of confidence apart from weekly and hourly frequencies. This should be because I did not adjust the data for serial correlation, which is expected because seasons and associated weathers affect air quality. But overall the results still show evidence to support that the air quality in Shanghai is improving, and that the post by the Redditor has merit.

Comparison

Now to compare Shanghai to the assertions the publication[3] made that initiated the reddit post. Assertions made in publication (Only Beijing):

- [Air] was the best since pollution control measures were implemented five years ago [in 2012].
- The average PM2.5 level in 2017 was 58 micrograms per cubic meter — down 20.5 percent from the year before.
- While the number of heavy pollution days decreased from 58 in 2013 to 23, Beijing also enjoyed 226 "good air" days last year — 28 more than in 2016, the bureau said.
- In November, PM2.5 levels dropped 54 percent year-on-year.

Let's see how Shanghai compared

Shanghai Yearly Particle Averages					
Year	2012	2013	2014	2015	2016
Average ug/m3	50.54	59.86	49.64	50.73	45.38
Percent Change		18.43%	-17.08%	2.20%	-10.53%
Shanghai November Particle Averages					
Year	2012	2013	2014	2015	2016
Average ug/m3	67.82	79.12	53.61	56.00	50.02
Percent Change		16.67%	-32.24%	4.46%	-10.68%

Table 3 Shanghai PM 2.5 Concentration Averages by Year and for November

From the results in Table 3, Shanghai does have a lower particle count in 2016 than 2012 with a yearly average of 45.38 ug/m^3 and 50.54 ug/m^3 respectively. The 2016 from 2015 particle count dropped 10.53% which is an improvement in air quality. Since there is no metric for "heavy pollution" days, this was not compared. Comparing November changes, Shanghai did not get the same 54% decreases. At most Shanghai had an improvement of 32.24% from November of 2013 to November of 2014. The results differ randomly from Beijing and does not illustrate to the same magnitude of pollution reduction. What is the same is the decrease in air pollution concentration.

Conclusion

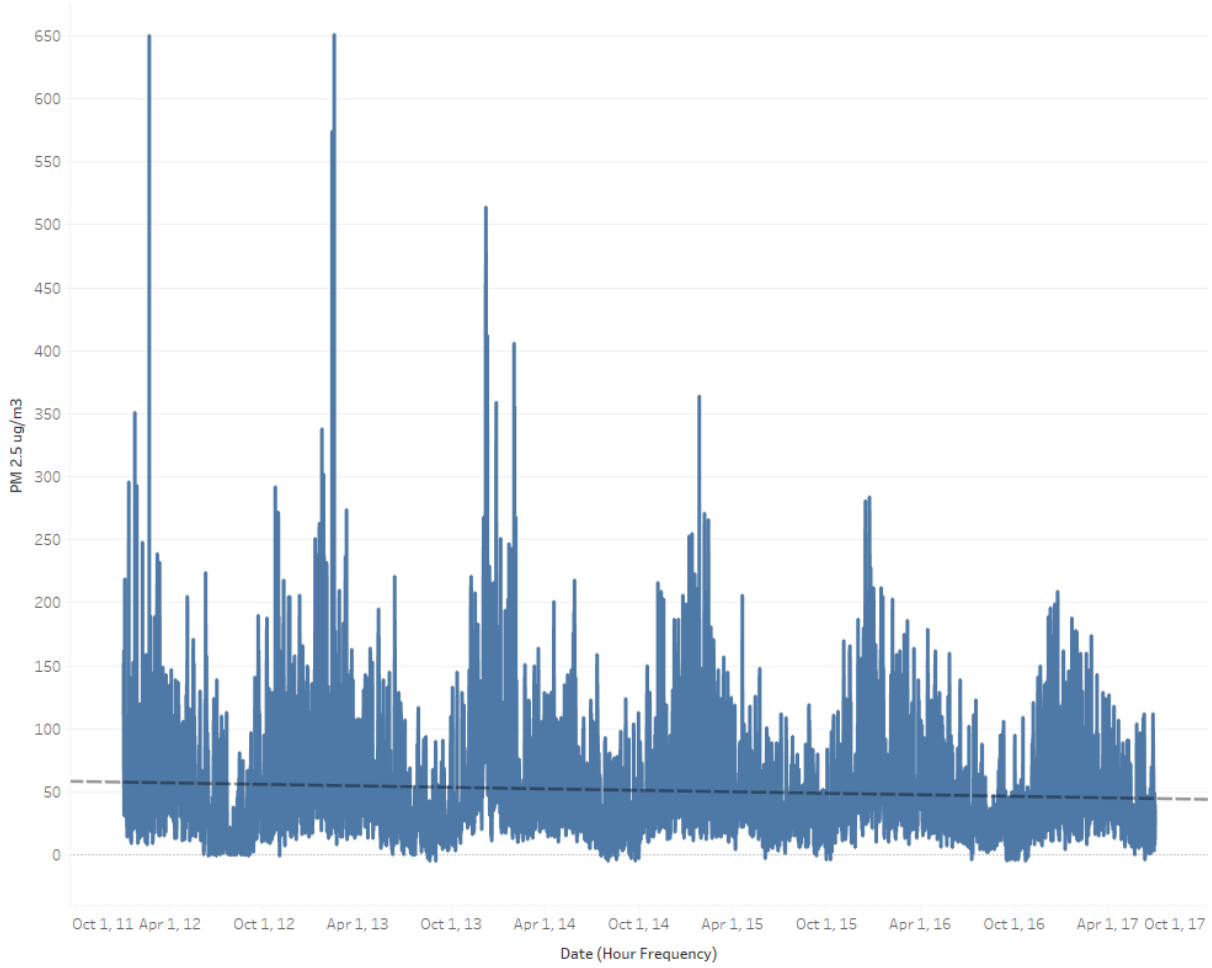
Shanghai did not achieve as great of success as illustrated in the publication for Beijing, but Shanghai still has been on a decreasing trend for their level of pollution. The Redditor who uploaded the "beautiful" data is correct that there is a decreasing in smog for Shanghai, and this project put forth statistical evidence to support that claim.

This paper only shows the statistical evidence to support the claim but does not explain why there is a decreased. Further research that could be built off of this paper could include background for extrema, for instance the trough as a result of Hangzhou's G20 summit meeting, or the week-long smog shroud during December of 2013.

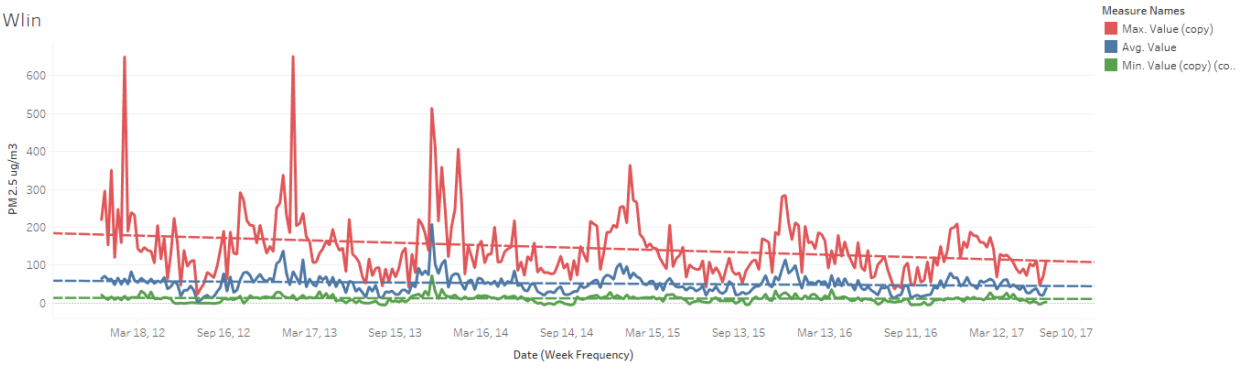
[3] https://www.hongkongfp.com/2018/01/04/beijing-records-best-air-quality-five-years-2017/

Appendix

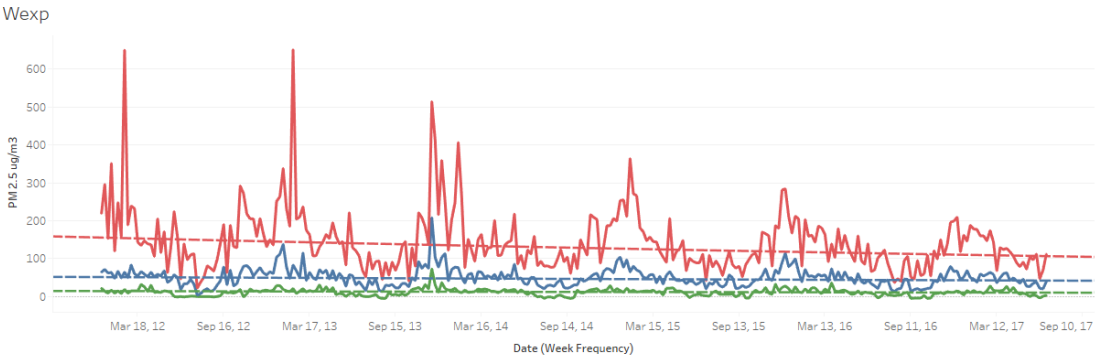
Graphs



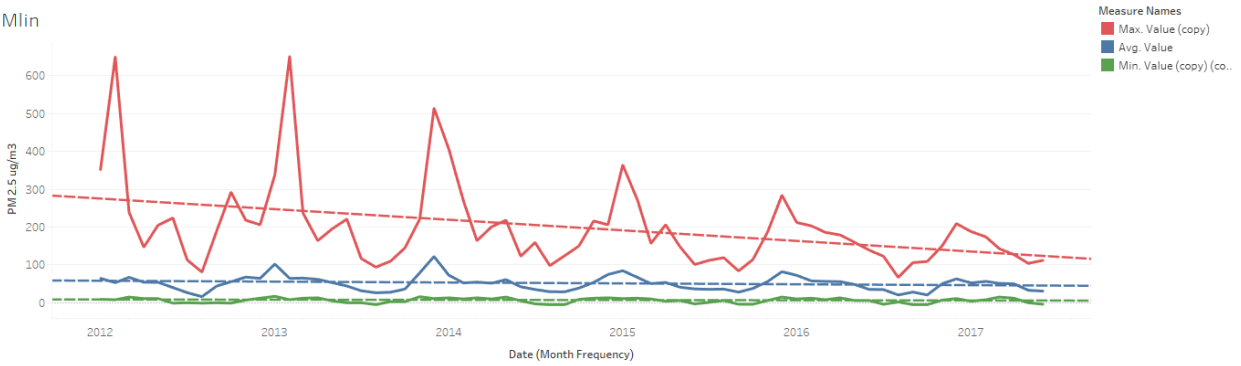
Wlin



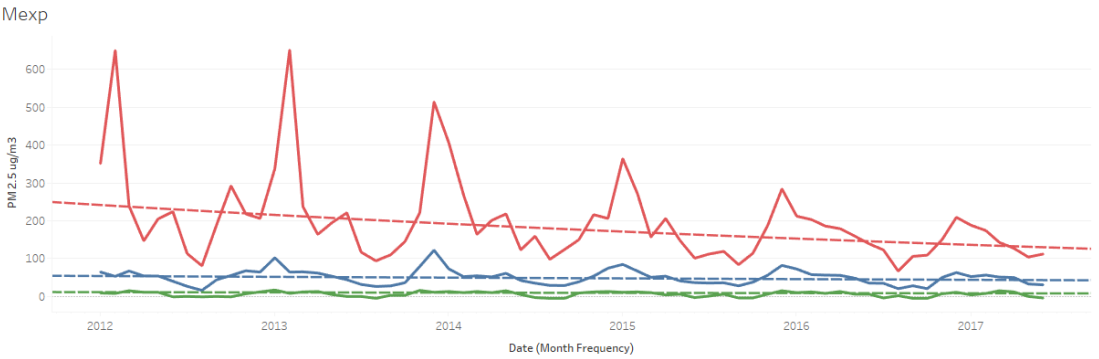
Wexp

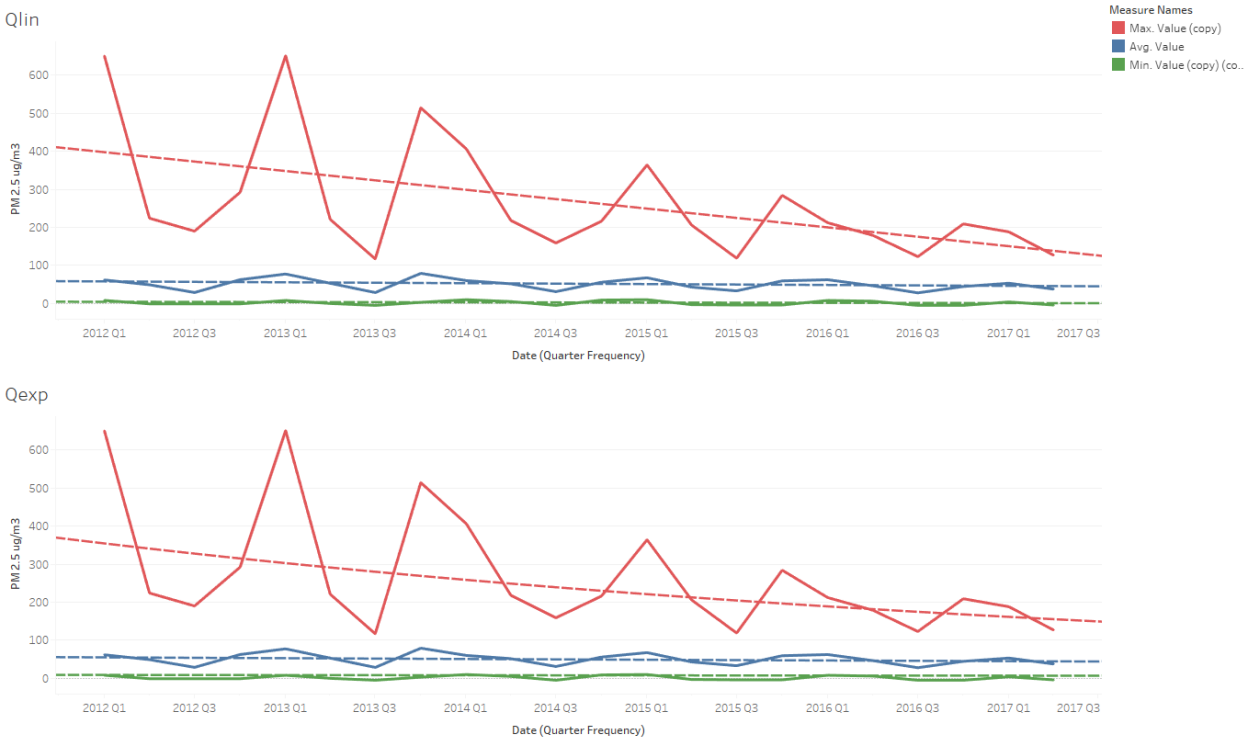


Mlin



Mexp





Regressions

Linear

Hour

Number of modeled observations: 46945
Number of filtered observations: 0
Model degrees of freedom: 2
Residual degrees of freedom (DF): 46943
SSE (sum squared error): 7.43619e+07
MSE (mean squared error): 1584.09
R-Squared: 0.0088068
Standard error: 39.8006
p-value (significance): < 0.0001

Individual trend lines:

Panes		Line		Coefficients				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Value	Hour of Date	< 0.0001	46943	Hour of Date	-0.0064767	0.0003171	-20.4228	< 0.0001
				intercept	322.172	13.2931	24.2359	< 0.0001

Week

Number of modeled observations: 861
Number of filtered observations: 0
Model degrees of freedom: 6
Residual degrees of freedom (DF): 855
SSE (sum squared error): 1.96307e+06

MSE (mean squared error): 2295.99
R-Squared: 0.59024
Standard error: 47.9165
p-value (significance): < 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Measure Names	4	2769768.1	692442	301.588	< 0.0001

Individual trend lines:

Color	Line		Coefficients				
Measure Names	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Min. Value	0.177997	285	Week of Date	-0.0012883	0.0009541	-1.35028	0.177997
			intercept	65.2524	39.9914	1.63166	0.103856
Avg. Value	0.0061473	285	Week of Date	-0.0065373	0.0023682	-2.76041	0.0061473
			intercept	324.58	99.2624	3.26992	0.0012078
Max. Value	< 0.0001	285	Week of Date	-0.0346079	0.0080522	-4.29794	< 0.0001
			intercept	1595.33	337.5	4.72689	< 0.0001

Month

Number of modeled observations: 198
Number of filtered observations: 0
Model degrees of freedom: 6
Residual degrees of freedom (DF): 192
SSE (sum squared error): 750554
MSE (mean squared error): 3909.14
R-Squared: 0.661934
Standard error: 62.5231
p-value (significance): < 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Measure Names	4	1417003.5	354251	90.6212	< 0.0001

Individual trend lines:

Color	Line		Coefficients				
Measure Names	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Min. Value	0.336925	64	Month of Date	-0.0013473	0.0013926	-0.967523	0.336925
			intercept	62.451	58.3515	1.07026	0.288524
Avg. Value	0.121452	64	Month of Date	-0.0063956	0.0040748	-1.56955	0.121452
			intercept	318.587	170.741	1.8659	0.0666386
Max. Value	0.0011981	64	Month of Date	-0.0765785	0.0225836	-3.39088	0.0011981
			intercept	3406.87	946.298	3.60021	0.0006208

Quarter

Number of modeled observations: 66
Number of filtered observations: 0
Model degrees of freedom: 6
Residual degrees of freedom (DF): 60
SSE (sum squared error): 390526
MSE (mean squared error): 6508.76

R-Squared: 0.721454
Standard error: 80.6769
p-value (significance): < 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Measure Names	4	960884.51	240221	36.9073	< 0.0001

Individual trend lines:

Color	Line		Coefficients				
Measure Names	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Min. Value	0.388124	20	Quarter of Date	-0.0018431	0.0020891	-0.882248	0.388124
			intercept	78.6676	87.4754	0.899311	0.379187
Avg. Value	0.264143	20	Quarter of Date	-0.0063921	0.0055636	-1.14892	0.264143
			intercept	318.274	232.954	1.36625	0.187025
Max. Value	0.0155018	20	Quarter of Date	-0.135168	0.0510845	-2.64596	0.0155018
			intercept	5925.9	2138.98	2.77043	0.0118031

Exponential

Week

Number of modeled observations: 819
Number of filtered observations: 42
Model degrees of freedom: 6
Residual degrees of freedom (DF): 813
SSE (sum squared error): 264.978
MSE (mean squared error): 0.325926
R-Squared: 0.751846
Standard error: 0.570899
p-value (significance): < 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Measure Names	4	793.55404	198.389	608.691	< 0.0001

Individual trend lines:

Color	Line		Coefficients				
Measure Names	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Min. Value (copy) (copy)	0.0221011	243	Week of Date	-0.0001853	8.046e-05	-2.3034	0.0221011
			intercept	10.1693	3.37424	3.01382	0.002853
Avg. Value	0.043641	285	Week of Date	-0.0001013	4.996e-05	-2.02653	0.043641
			intercept	8.05791	2.09413	3.84785	0.0001471
Max. Value (copy)	0.0001773	285	Week of Date	-0.0001904	5.01e-05	-3.79946	0.0001773
			intercept	12.827	2.10005	6.10792	< 0.0001

Month

Number of modeled observations: 179
Number of filtered observations: 19
Model degrees of freedom: 6
Residual degrees of freedom (DF): 173
SSE (sum squared error): 36.6674
MSE (mean squared error): 0.21195

R-Squared: 0.875576
Standard error: 0.460381
p-value (significance): < 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Measure Names	4	255.20264	63.8007	301.017	< 0.0001

Individual trend lines:

Color	Line		Coefficients				
Measure Names	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Min. Value (copy) (copy)	0.202522	45	Month of Date	-0.0001932	0.0001494	-1.29326	0.202522
			intercept	10.2187	6.25994	1.6324	0.109574
Avg. Value	0.167785	64	Month of Date	-0.0001154	8.273e-05	-1.39518	0.167785
			intercept	8.68756	3.46646	2.50617	0.0147571
Max. Value (copy)	0.0010808	64	Month of Date	-0.0003154	9.21e-05	-3.42418	0.0010808
			intercept	18.384	3.85919	4.7637	< 0.0001

Quarter

Number of modeled observations: 54
Number of filtered observations: 12
Model degrees of freedom: 6
Residual degrees of freedom (DF): 48
SSE (sum squared error): 7.52678
MSE (mean squared error): 0.156808
R-Squared: 0.923959
Standard error: 0.39599
p-value (significance): < 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Measure Names	4	90.43005	22.6075	144.173	< 0.0001

Individual trend lines:

Color	Line		Coefficients				
Measure Names	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Min. Value (copy) (copy)	0.567496	8	Quarter of Date	-0.0001509	0.0002531	-0.596244	0.567496
			intercept	8.21052	10.5969	0.774803	0.460741
Avg. Value	0.355308	20	Quarter of Date	-0.0001121	0.0001185	-0.946262	0.355308
			intercept	8.57284	4.96177	1.72778	0.0994391
Max. Value (copy)	0.0162141	20	Quarter of Date	-0.0004325	0.0001647	-2.62528	0.0162141
			intercept	23.5595	6.89772	3.41555	0.0027412