

2024 Seattle PM_{2.5} Prediction

Team 2

Yi-Kai Chen,

Neha S. Rachapalli,

Teresa Tseng,

Bridget Bills,

Chris Chen



Seattle PM2.5 Exploration

Project Overview

- PM2.5 is a crucial air quality indicator as it consists of fine particles that pose potential health risks. Understanding and predicting PM2.5 levels in Seattle is significant for *public health, environment and urban planning.*

Motivation

- Contribute to a healthier and more sustainable urban environment, promoting data-driven decision-making for both public health and urban planning initiatives.

Intended Audiences

- Citizens of Seattle
- Officials of forest management

Data Overview

Source

We gathered data from “*United States Environmental Protection Agency*”

Timespan

The data contains five years time frame from

2019 to 2023

which is well maintain and meaningful records for our research questions.

Scope

Providing a broad scope from the entire United States and narrow down into cities in each state. Listing specific features related to air quality indicators such as monitor stations, method name, and capture time. Enabling us to analyze PM 2.5 in a larger area and dive into certain places simultaneously with the rich information.

Dataset Details

The dataset consists of
29 columns with
2,655,219 entries
from **2019** to **2023**

#	Column	Dtype
0	State Code	int64
1	County Code	int64
2	Site Num	int64
3	Parameter Code	int64
4	POC	int64
5	Latitude	float64
6	Longitude	float64
7	Datum	object
8	Parameter Name	object
9	Sample Duration	object
10	Pollutant Standard	object
11	Date Local	object
12	Units of Measure	object
13	Event Type	object
14	Observation Count	int64
15	Observation Percent	float64
16	Arithmetic Mean	float64
17	1st Max Value	float64
18	1st Max Hour	int64
19	AQI	float64
20	Method Code	float64
21	Method Name	object
22	Local Site Name	object
23	Address	object
24	State Name	object
25	County Name	object
26	City Name	object
27	CBSA Name	object
28	Date of Last Change	object

Dataset Details



The dataset consists of **29** columns with **2,655,219** entries from **2019** to **2023**

Key Variables:

- **Date**
 - Samples are collected daily, sometimes multiple times per day depending on the site.
- **Daily Mean of PM2.5**
 - This measures the average level of air pollution observed over the course of the day from each site.
- **City name**
 - For our evaluation, we focused on Seattle, Philadelphia, and Tampa

Variable Exploration



Total Variables: 29

- **Location (US Only)**
 - State, County, City, Address, Latitude, Longitude
- **Sample Collection**
 - Date, Measure Duration, Unit of Measure, Mean, PM 2.5 Level, Max Value, AQI



Few insights of descriptive statistics:

- The mean arithmetic mean (average pollutant concentration) of **Seattle** is **7.26**, with a standard deviation of **10.5**.
- The mean arithmetic mean (average pollutant concentration) of **Tampa** is **8.34** with a standard deviation of **2.92**.
- The mean arithmetic mean (average pollutant concentration) of **Philadelphia** is **9.39** with a standard deviation of **8.48**.



Descriptive statistics

Data Preparation

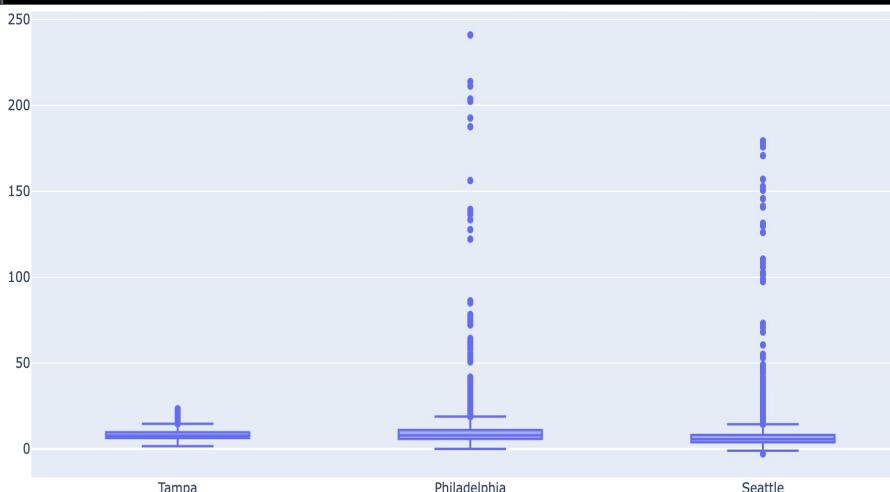
1. Data Processing

- Data type conversion: Change to Date data type (Object to datetime64)

2. Missing Data

- Drop irrelevant missing columns (AQI, Pollutant Standard, Method Code)

3. Keep Outliers



State Code	0
County Code	0
Site Num	0
Parameter Code	0
POC	0
Latitude	0
Longitude	0
Datum	0
Parameter Name	0
Sample Duration	0
Pollutant Standard	15582
Date Local	0
Units of Measure	0
Event Type	0
Observation Count	0
Observation Percent	0
Arithmetic Mean	0
1st Max Value	0
1st Max Hour	0
AQI	15582
Method Code	3119
Method Name	0
Local Site Name	0
Address	0
State Name	0
County Name	0
City Name	0
CBSA Name	0
Date of Last Change	0
dtype: int64	

A List of Missing Values

Data Preparation

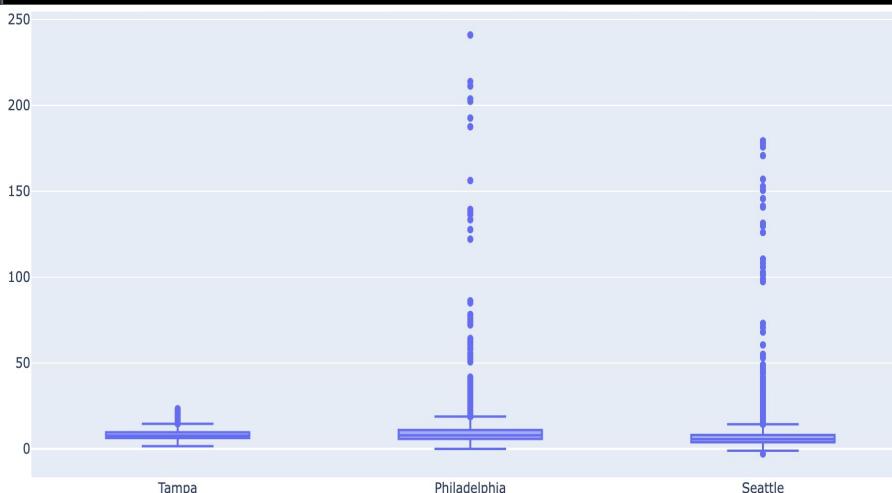
1. Data Processing

- Data type conversion: Change to Date data type (Object to datetime64)

2. Missing Data

- Drop irrelevant missing columns (AQI, Pollutant Standard, Method Code)

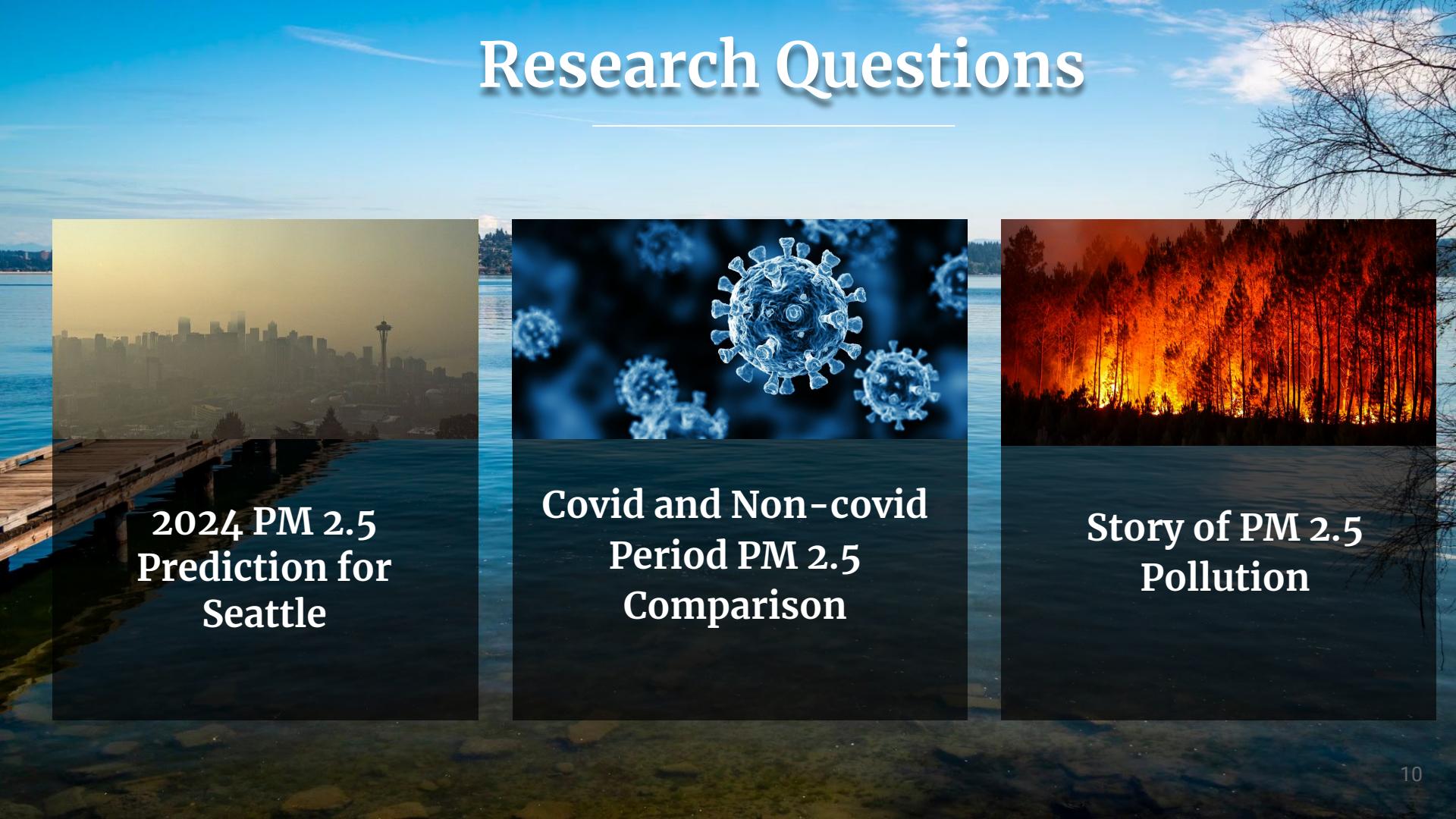
3. Keep Outliers



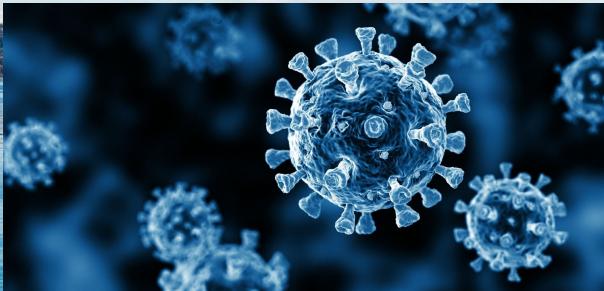
Key Missing Values:

- Pollutant Standard: 15,000 missing records
- Air Quality Index (AQI): 15,000 missing records
 - This informed our decision to use the PM2.5 daily mean instead of AQI. Though this measurement is more well-known, we did not feel we had sufficient records to use this measurement.
- Method Code: 3,000 missing records
- Other than these variables, we had very little missing data from the rest of the file.

Research Questions

A photograph of the Seattle skyline at sunset, viewed from across a body of water. The city lights are visible through a hazy atmosphere. A wooden pier is in the foreground on the left.

2024 PM 2.5
Prediction for
Seattle

An illustration of several blue COVID-19 viruses against a dark background. The viruses are spherical with prominent spikes.

Covid and Non-covid
Period PM 2.5
Comparison

A photograph of a forest fire. The trees are engulfed in intense orange and red flames, casting a glow on the surrounding area. A bare tree branch is visible in the upper right corner.

Story of PM 2.5
Pollution

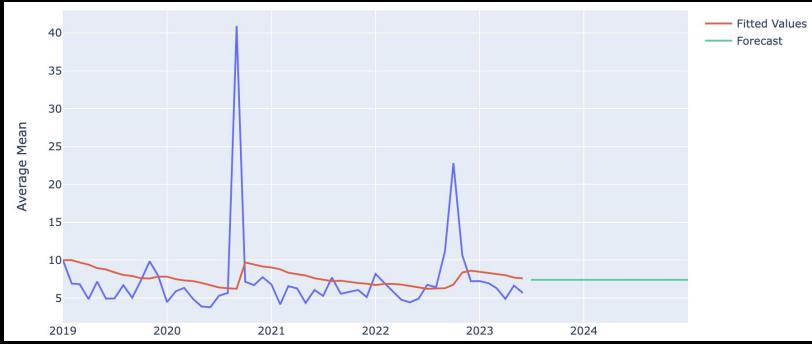
1. 2024 PM 2.5 Prediction for Seattle

Methodology

- ETS Model
- Holt's Model
- Winters' Model
- ARIMA Model
- SARIMA Model

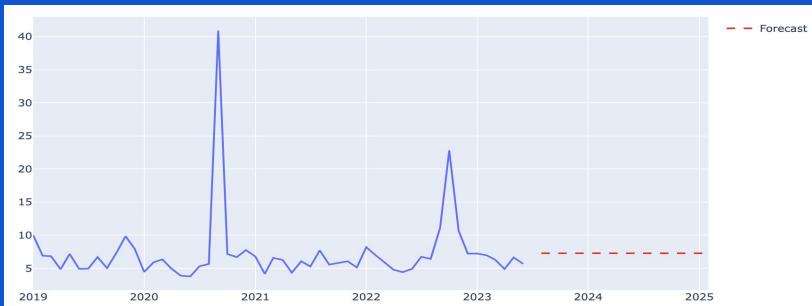
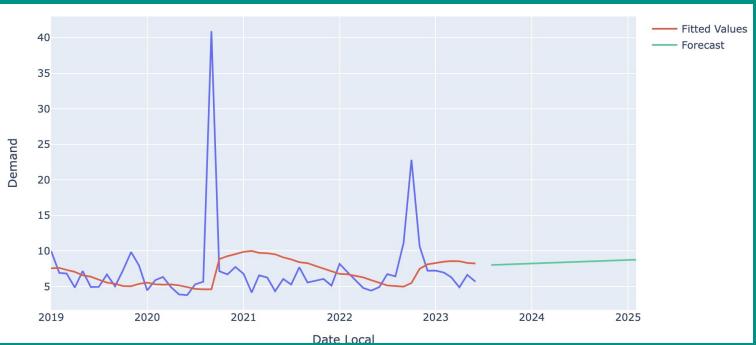
Variables

- Date from 2019 ~ 2023
- Monthly PM 2.5 Average



⚠️ ETS Model

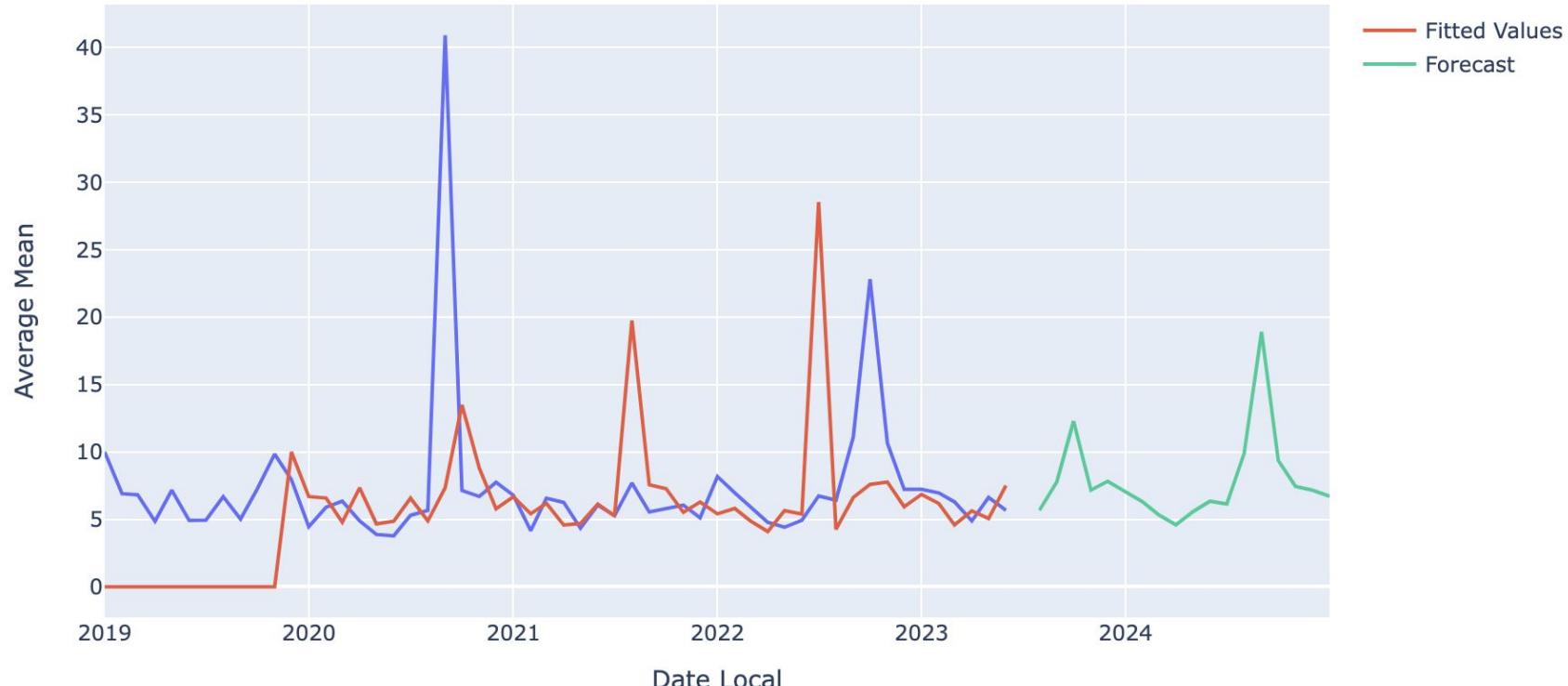
⚠️ Holt's Model



⚠️ ARIMA Model

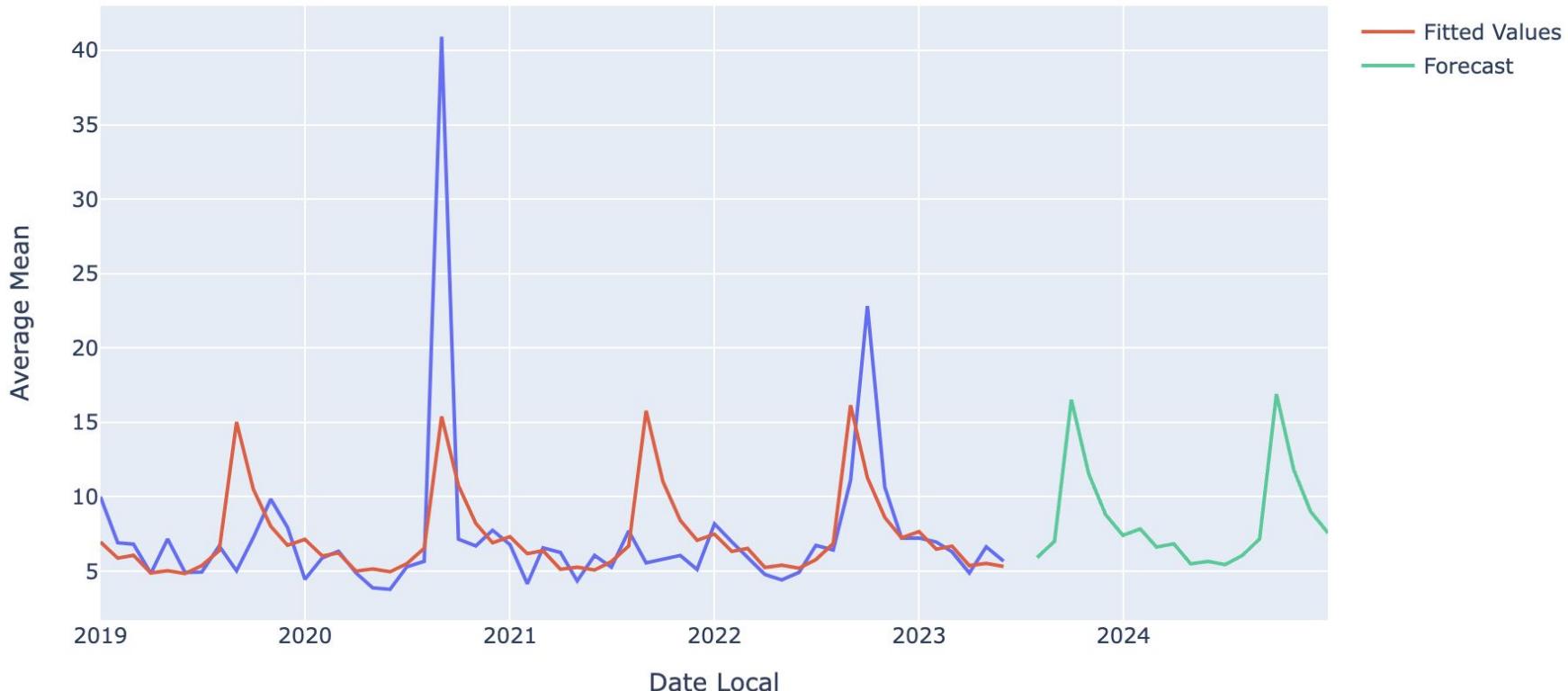
SARIMA

Average Mean Over Time - Seattle



Winters' Model

Average Mean Over Time - Seattle



Summary of Prediction

Statistical Performance

- MAPE of Holt-Winters: **22%**
- MAPE of SARIMA: **26%**

Assumption of Holt-Winters

- Seasonality repeats regularly
- Trend is constant

Things to expect

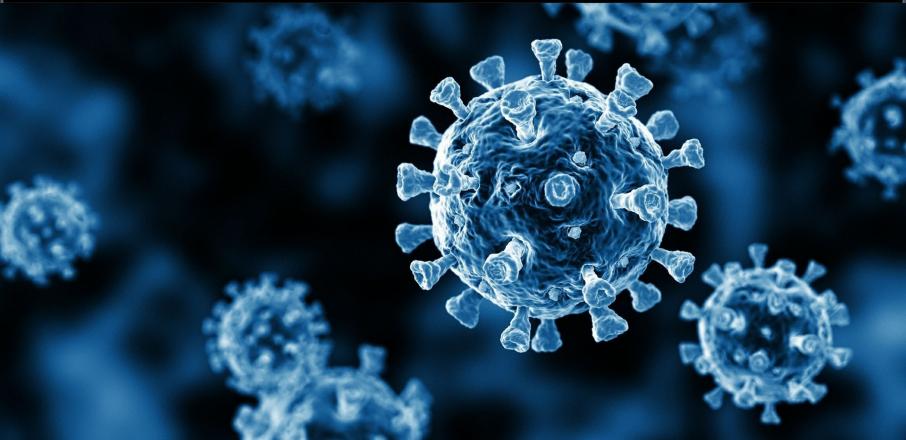
- Peak of average PM 2.5 level is expected to be **16.9** in September 2024

What can we do?

- Buy air purifier
- Wear mask for outdoor activities
- Forest Management

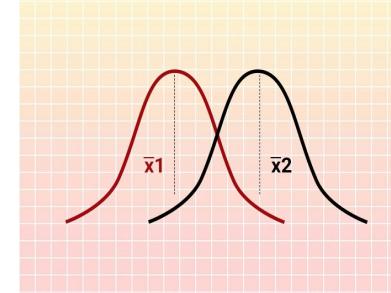
2

Covid Period and Non-Covid Comparison

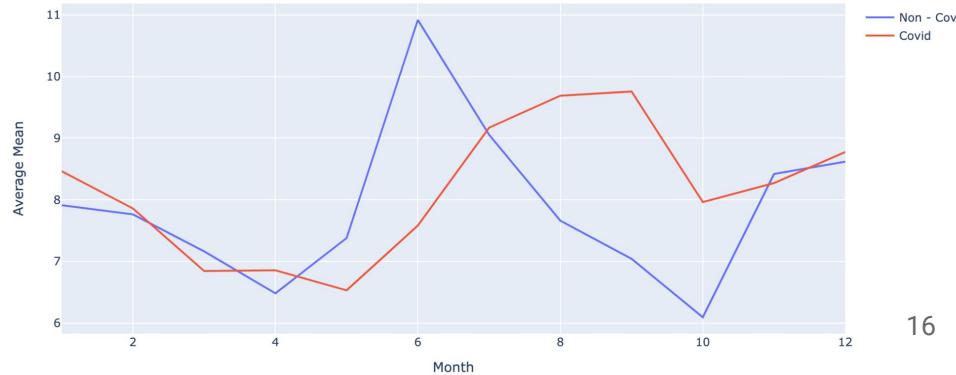


Non- Covid Pandemic
2019, 2023
Covid Pandemic
2020~2022

Mean of the non-COVID group is lower than the mean of the COVID group with
-6.08 of T Statistics

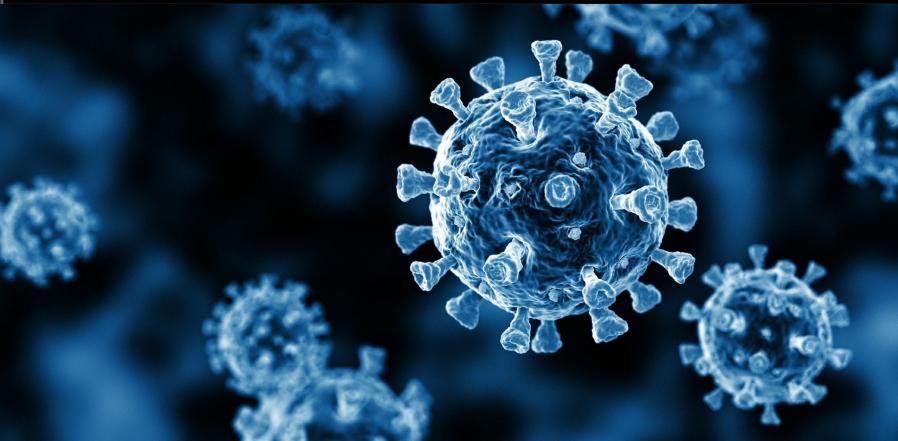


Comparison of Monthly Average Means (Non-Covid) vs (Covid)



2

Covid Period and Non-Covid Comparison

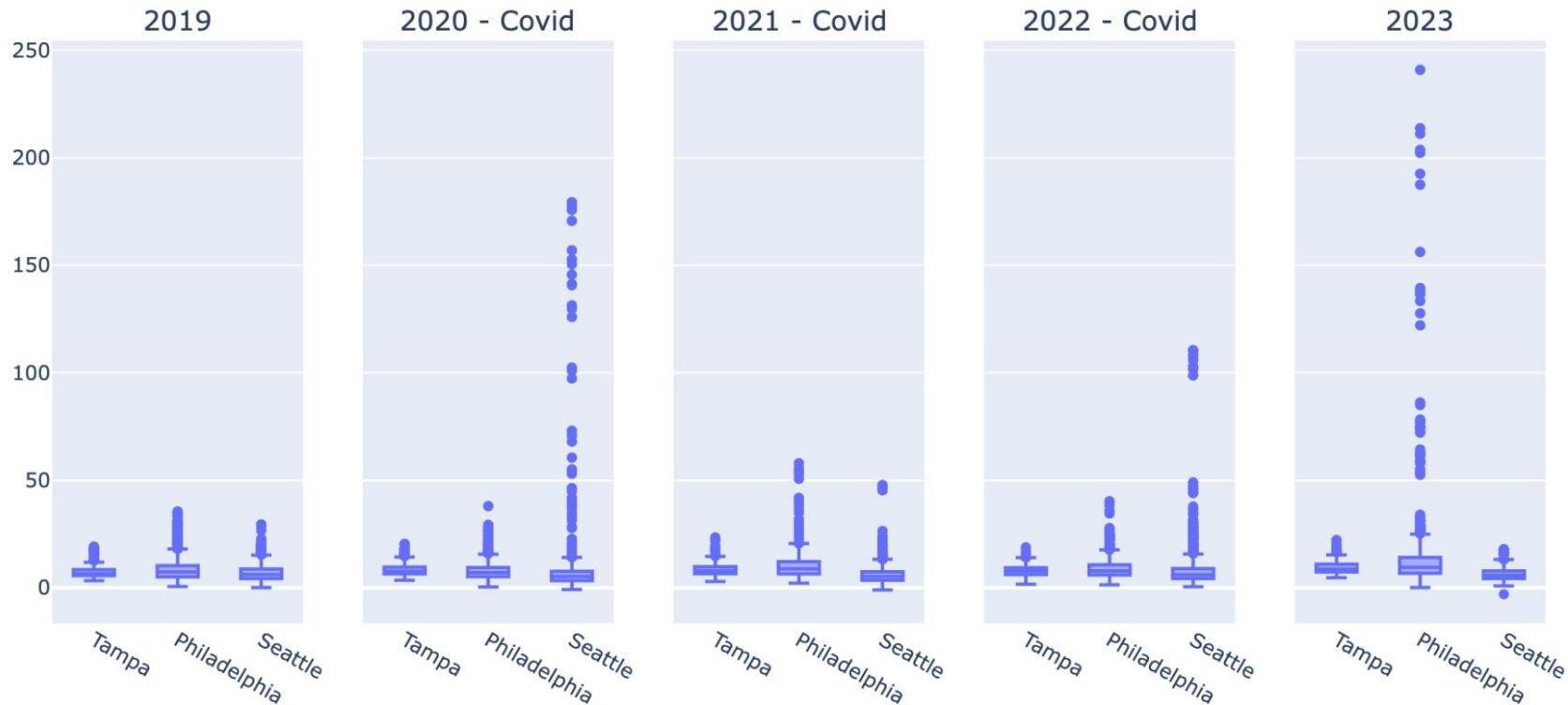


Non- Covid Pandemic
2019, 2023
Covid Pandemic
2020 - 2022

Komolgorov-Smirnoff Distribution Assessment:

- KS Statistic: 0.0231
- P-value: 1.61e-274
- We reject the null hypothesis that air pollution during COVID and Non-COVID periods was the same. The two are significantly different:
 - PM2.5 Non-Covid:
 - PM2.5 Covid:

Box Plots of Non-Covid and Covid Period



3. Story Of PM 2.5 Pollution

Wildfire

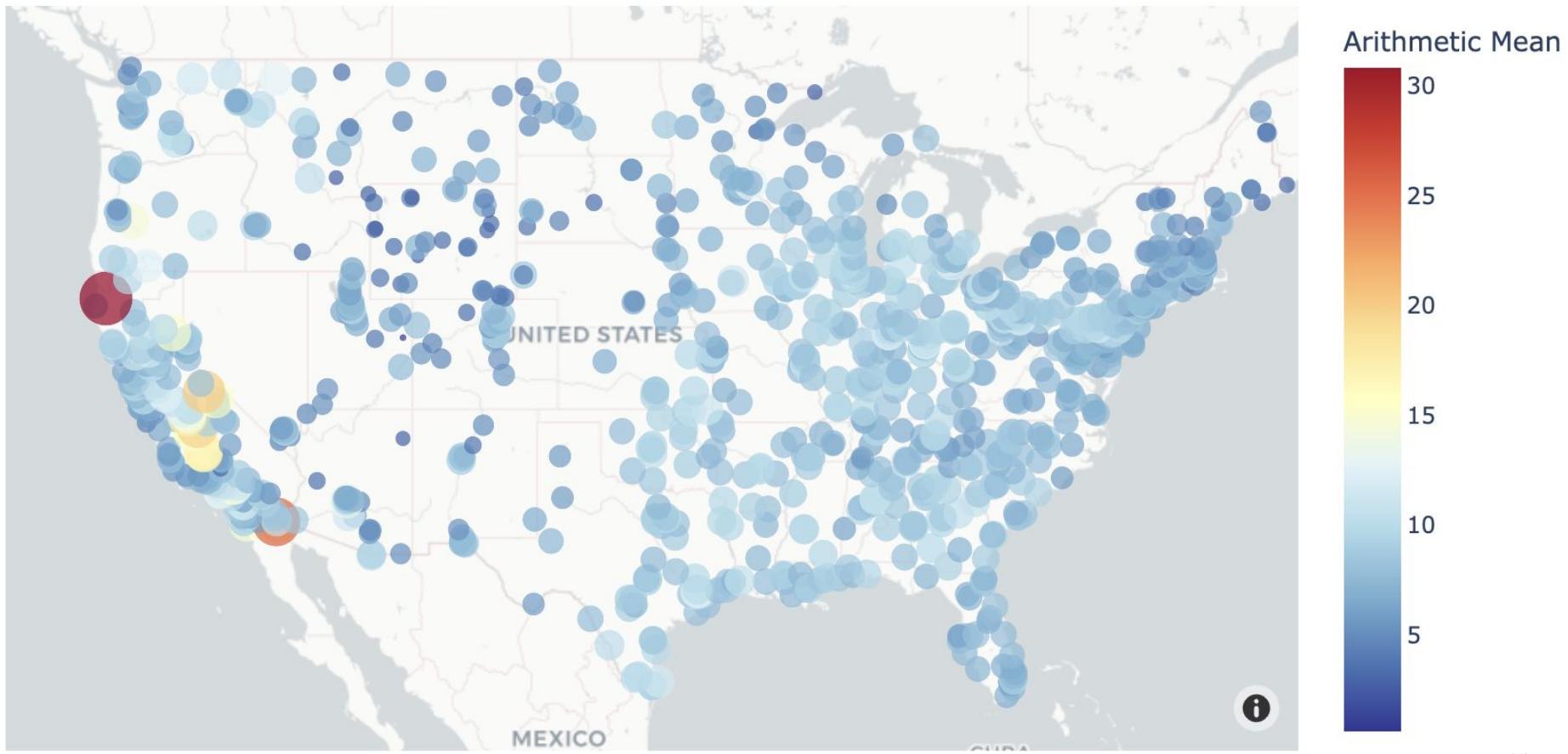
- We strongly suspect that Seattle's main cause of poor air quality. We recommend future research join wildfire data with this dataset to answer the question conclusively.
- Only **4.08%** of days in Seattle are more polluted than WHO target in the past 5 years.
- Annual average PM_{2.5} level: **7.2**
- Average of PM 2.5 when exceeding standard : **36**



Vehicle Emissions

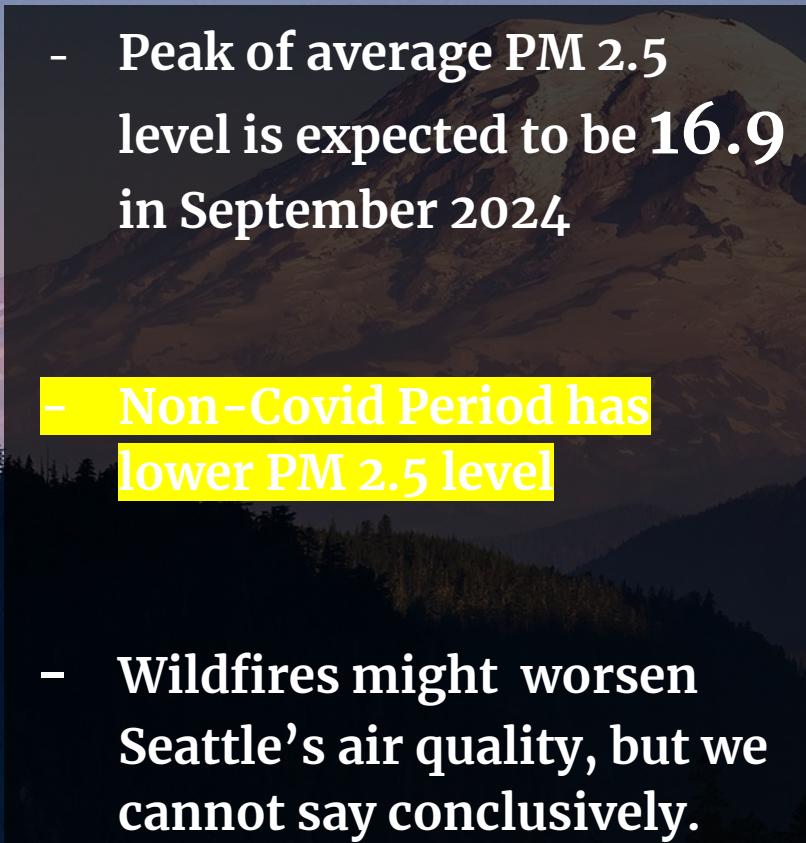
- In Philadelphia, we suspect that emissions from gas and diesel-powered vehicles is the main cause of air pollution.
- Annual average: **9.19**
- **9.07%** of days in Philadelphia are more polluted than WHO target.

Average Pm 2.5 Values by Location



Summary

- Peak of average PM 2.5 level is expected to be 16.9 in September 2024
- Non-Covid Period has lower PM 2.5 level
- Wildfires might worsen Seattle's air quality, but we cannot say conclusively.



Thank You
Q&A



Data Source:

United States Environmental Protection Agency

Link:

<https://www.epa.gov/outdoor-air-quality-data>

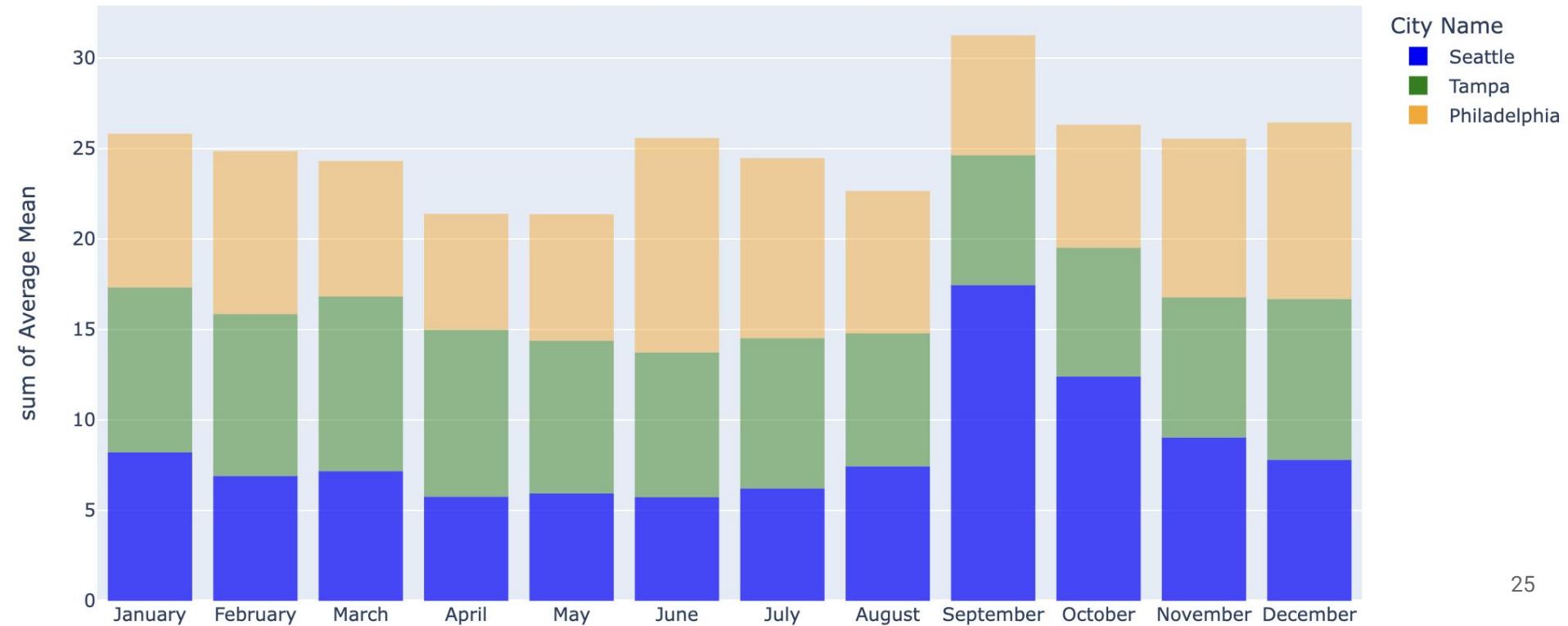


Appendix



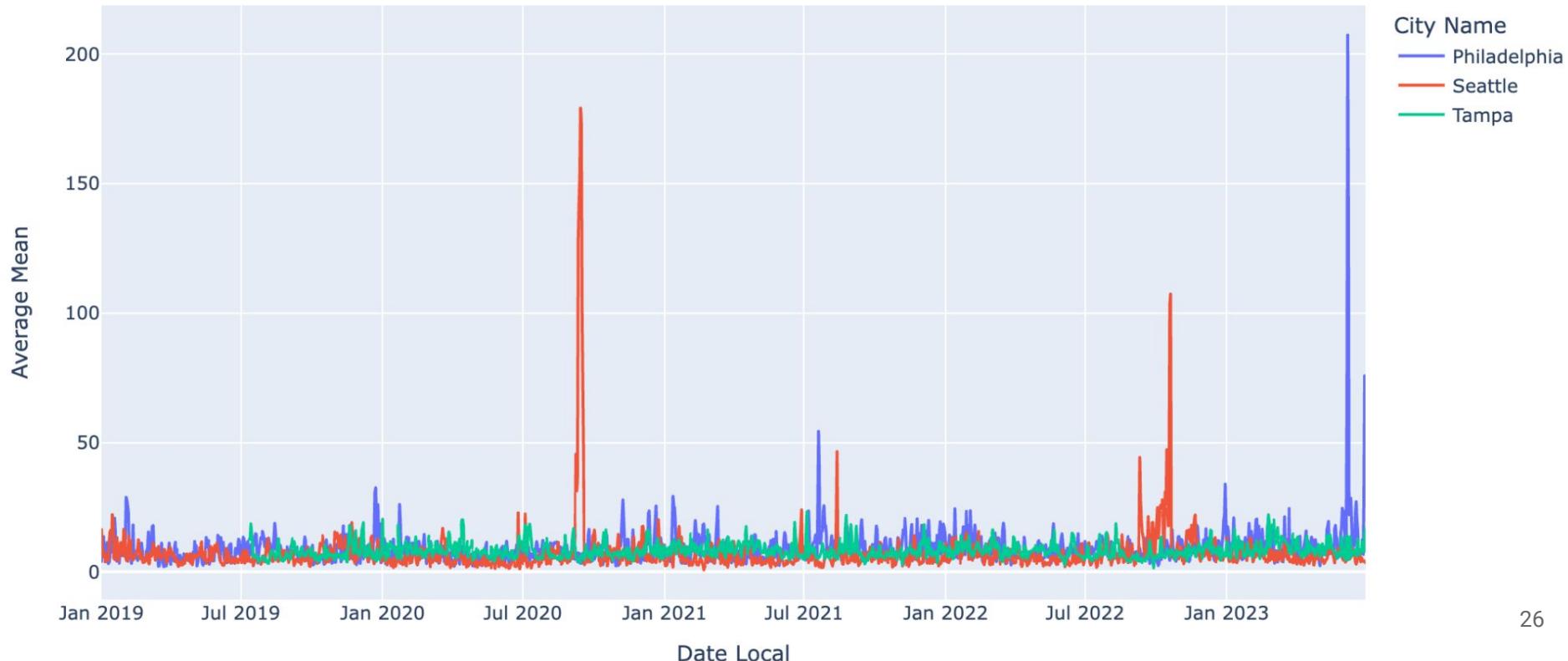
Data Visualization

Distribution of PM_{2.5} per month



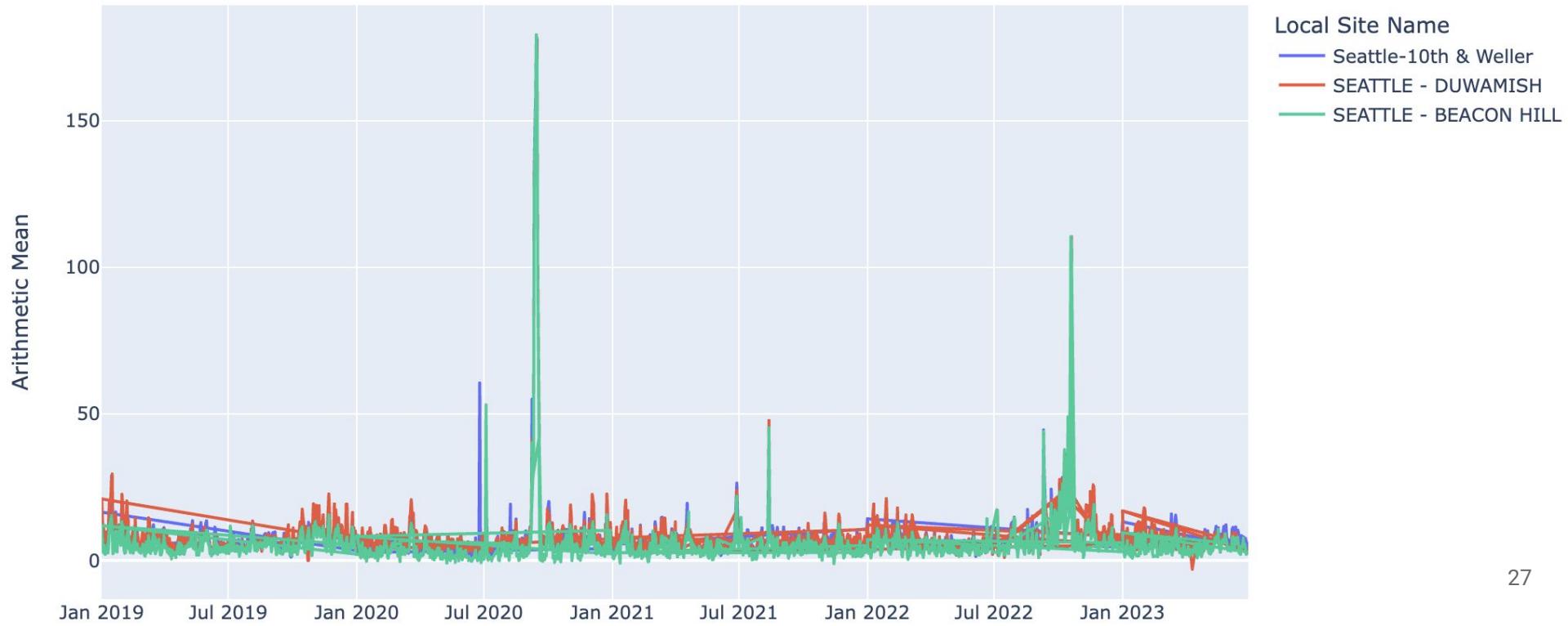
Data Visualization

Change of PM2.5 level over time



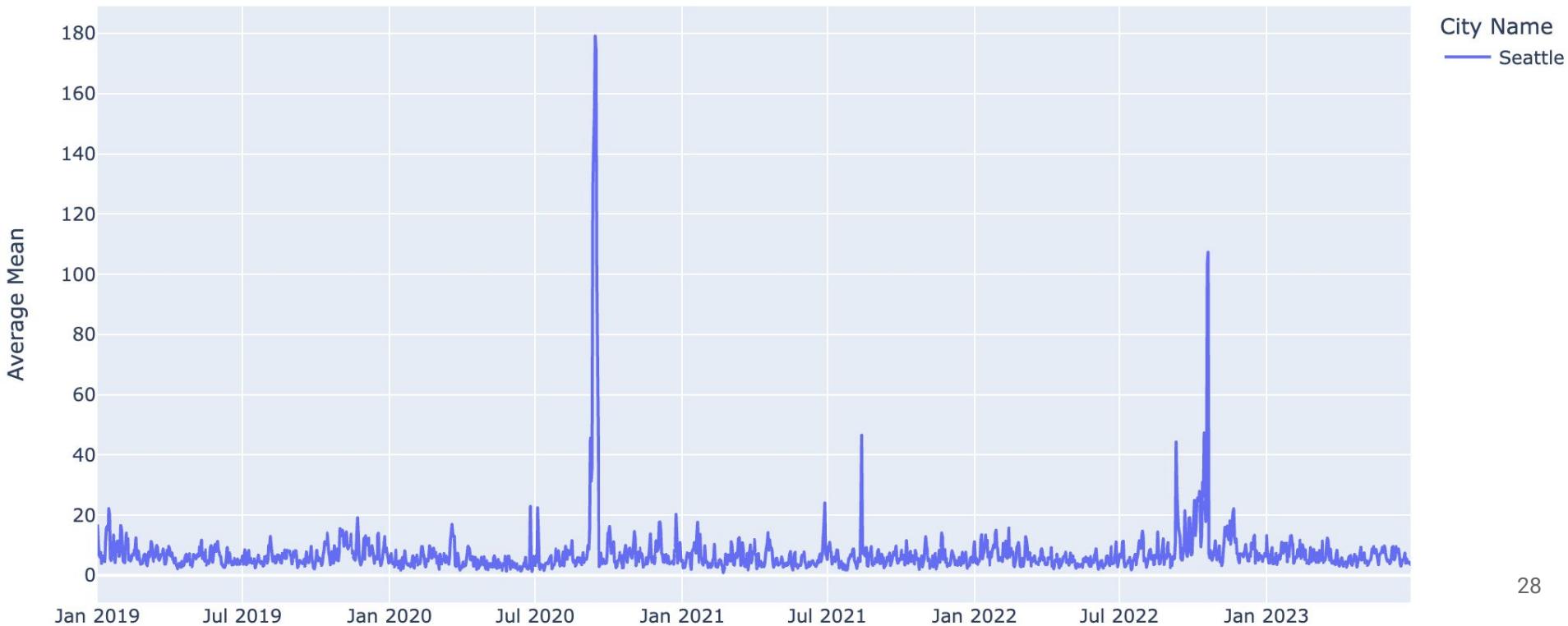
Data Visualization

Change of PM2.5 level over time- 3 Local Sites in Seattle



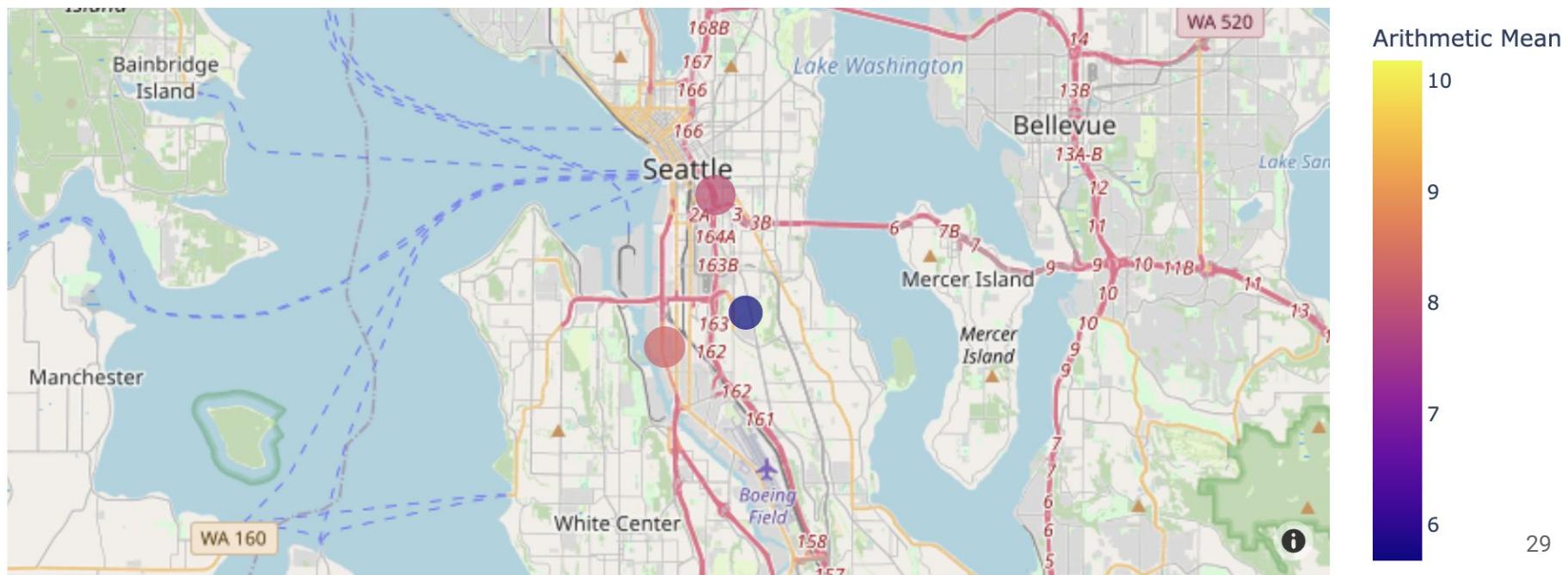
Data Visualization

Change of PM2.5 level over time- Seattle



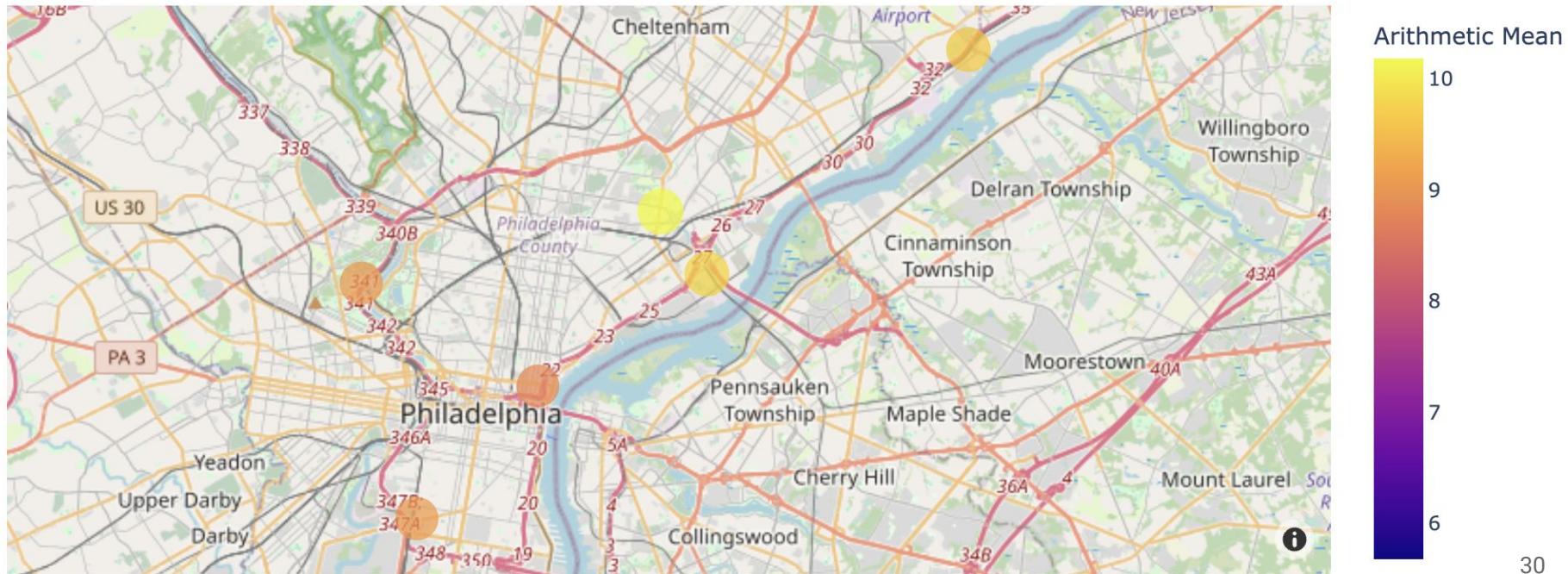
Data Visualization

Location of PM_{2.5} observation station (Seattle)



Data Visualization

Location of PM_{2.5} observation station (Philadelphia)



Data Visualization

Location of PM_{2.5} observation station (Tampa)

