Air Quality and COVID in NYC

Justin Williams, DATA 607

Contents

Research Question

Does NYC air quality correlate with COVID case count?

Photo by Hunter Reilly on Unsplash

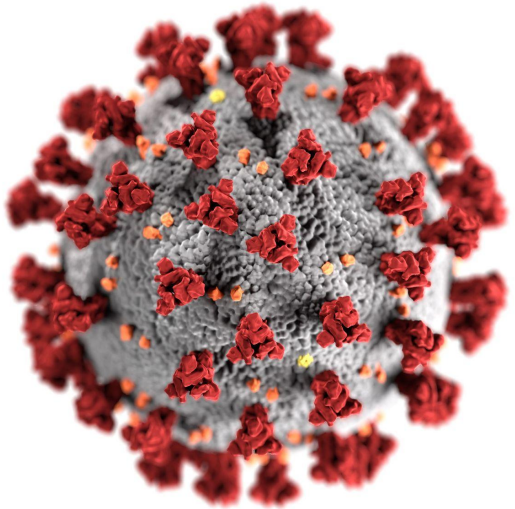


Photo by CDC on Unsplash

What is Particulate Matter 2.5 (PM 2.5)

PM is the term for a mixture of solid particles and liquid droplets found in the air. Some of these include:

- Large: Dust, Dirt, Soot, Smoke
- Small: Others so small need electron microscope

PM 2.5 are fine inhalable particles that have diameter less than 2.5 micrometers. Sources include:

- Some: Construction sites, Unpaved roads, fields, smokestacks, fires
- Most: Power plants, industries and automobiles (namely trucks).

Health & Environmental effects

Health implications:

- Premature death for people with heart or lung disease
- Nonfatal heart attacks
- Irregular heartbeat
- Aggravated asthma
- Decreased lung function

Environmental effects:

- Waterway acidification
- Nutrient balance in riparian environments
- Nutrient depletion in soil
- Damaging forests and farm crops
- Acid rain
- Ecosystem diversity

NYC Dept of Mental Health and Hygiene (DOHMOH)

“Current exposures to the annual average concentrations of PM2.5 above background concentrations cause more than 3,000 premature deaths, more than 2,000 hospitalizations due to respiratory and cardiovascular causes, and approximately 6,000 emergency department visits for asthma (Table 5) in New York City annually.”

Data Sources

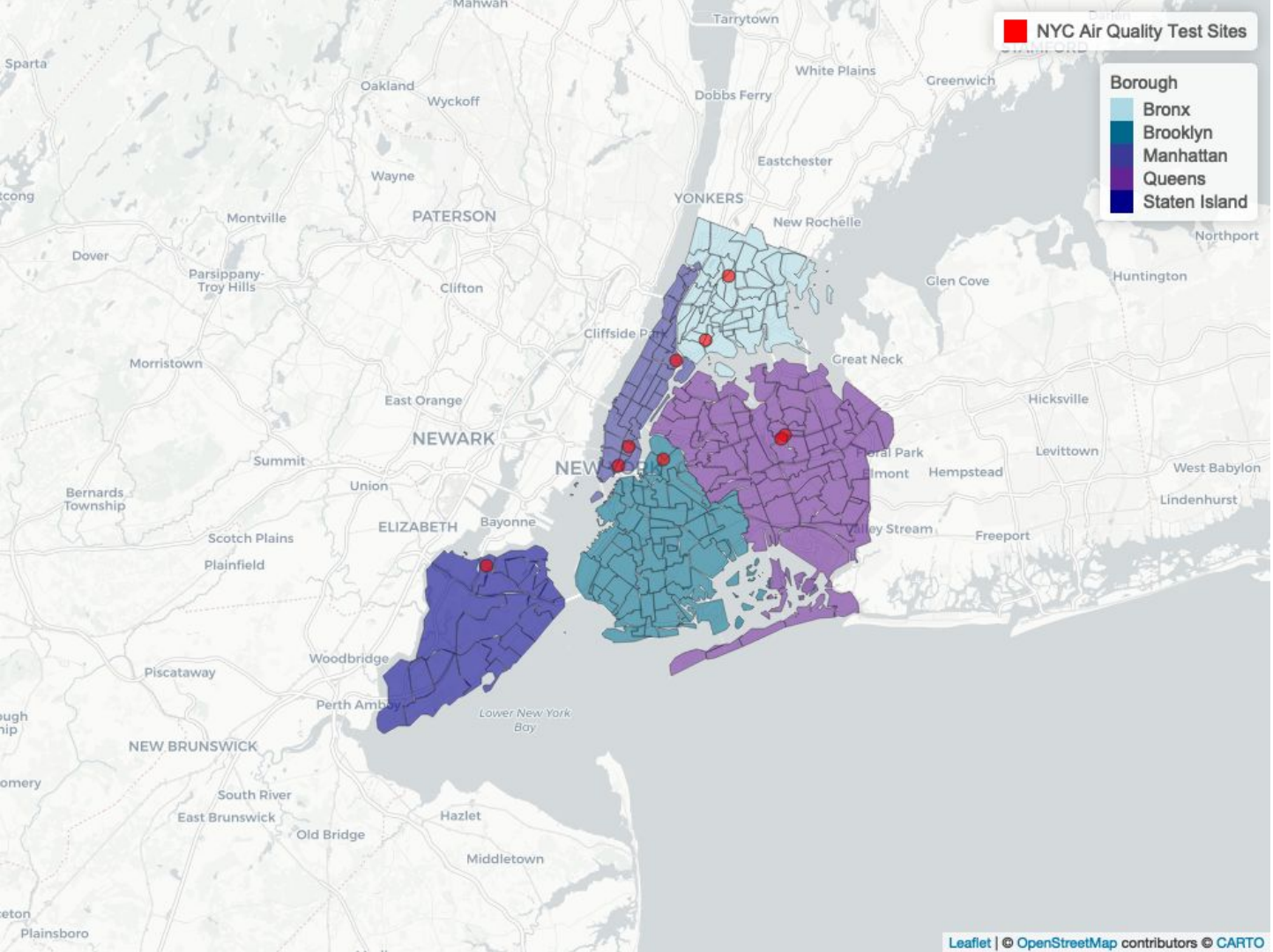
US Environmental Protection Agency (EPA)

- **Air Data** - Air Quality Data Collected at Outdoor Monitors Across the US:
  - 8 Pollutants
    - Particulate Matter 2.5 (PM2.5)
- 2017 - 2021:
  - ~55k rows (daily multiple sites)
  - 32 columns
- **RAQSAPI AQS**

Department of Health & Mental Hygiene (DOHMH)

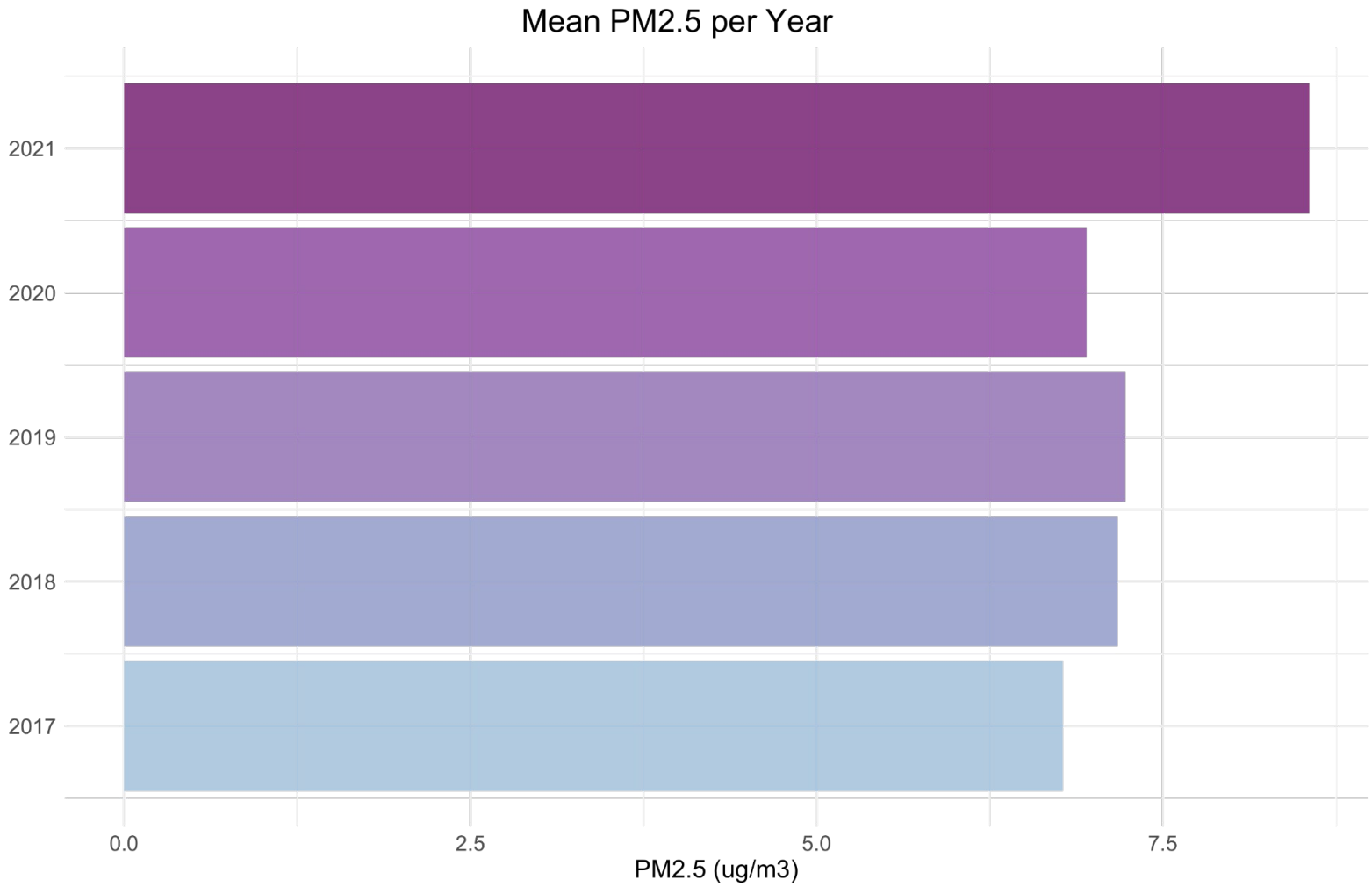
- Daily count of NYC residents who:
  - Tested positive for SARS-CoV-2
  - Hospitalized with COVID-19
  - Deaths among COVID-19 patients.
- March 2020 - 2021
  - 802 rows (daily)
  - 62 columns
- **Socrata API**

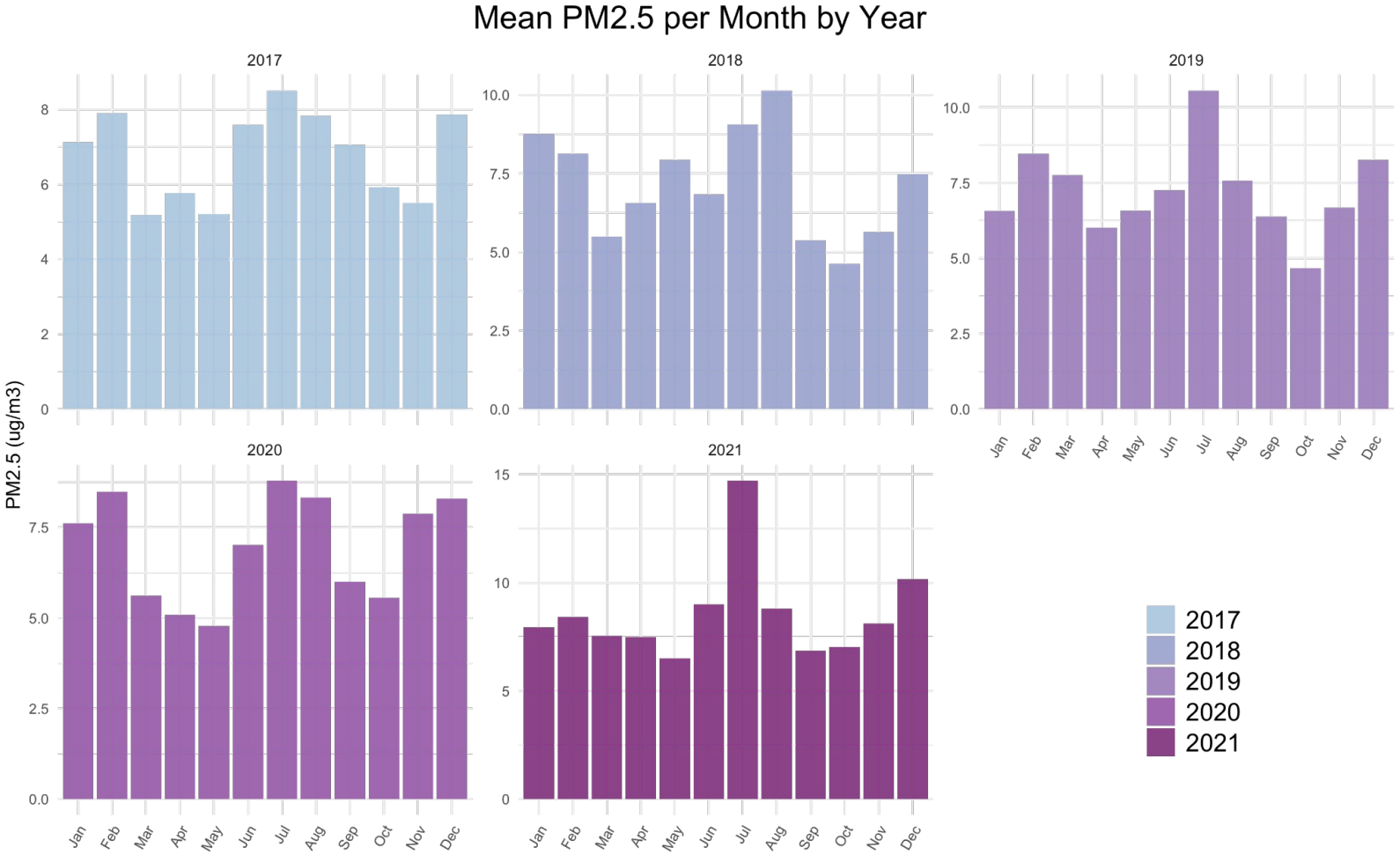
Exploratory Data Analysis

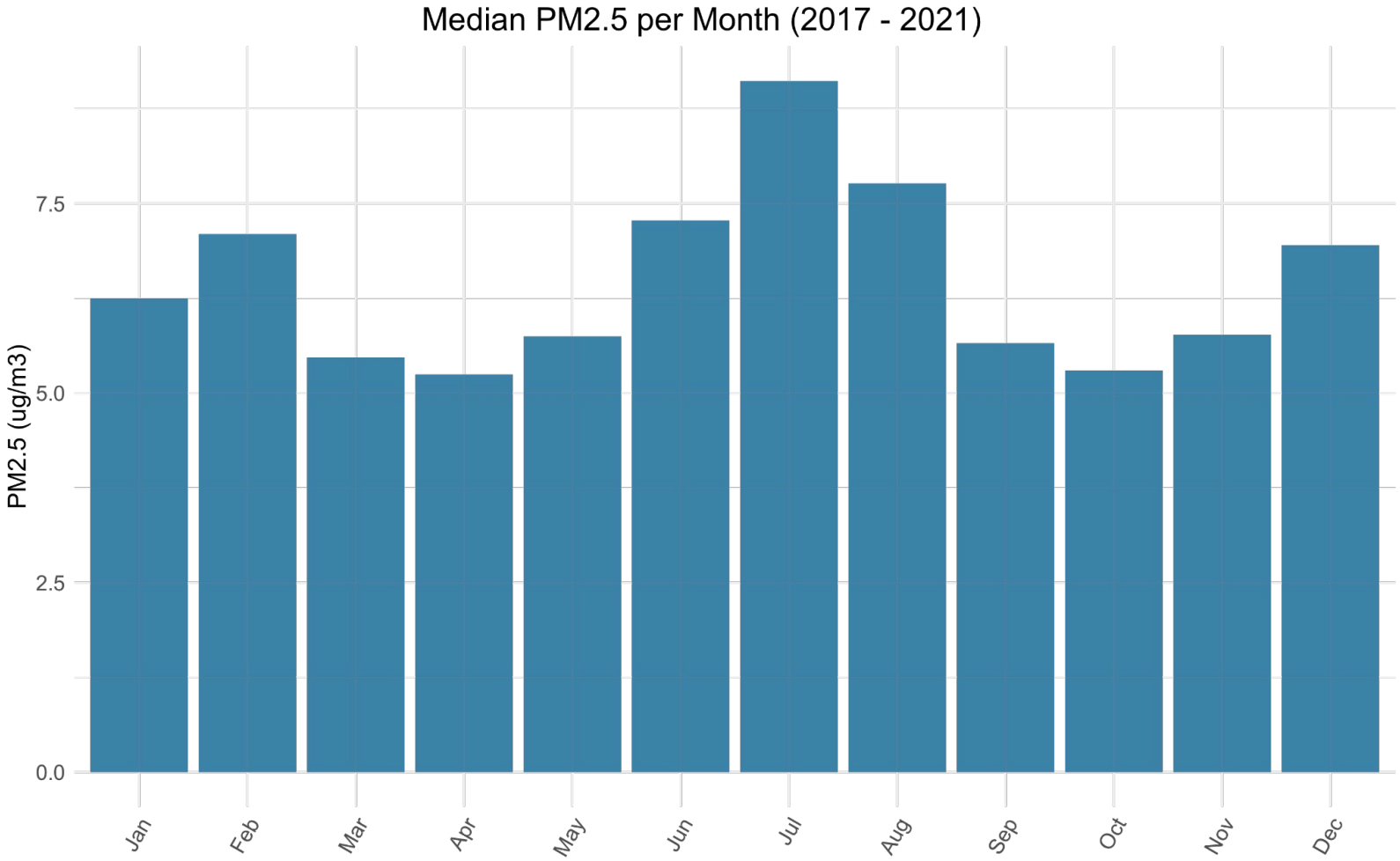


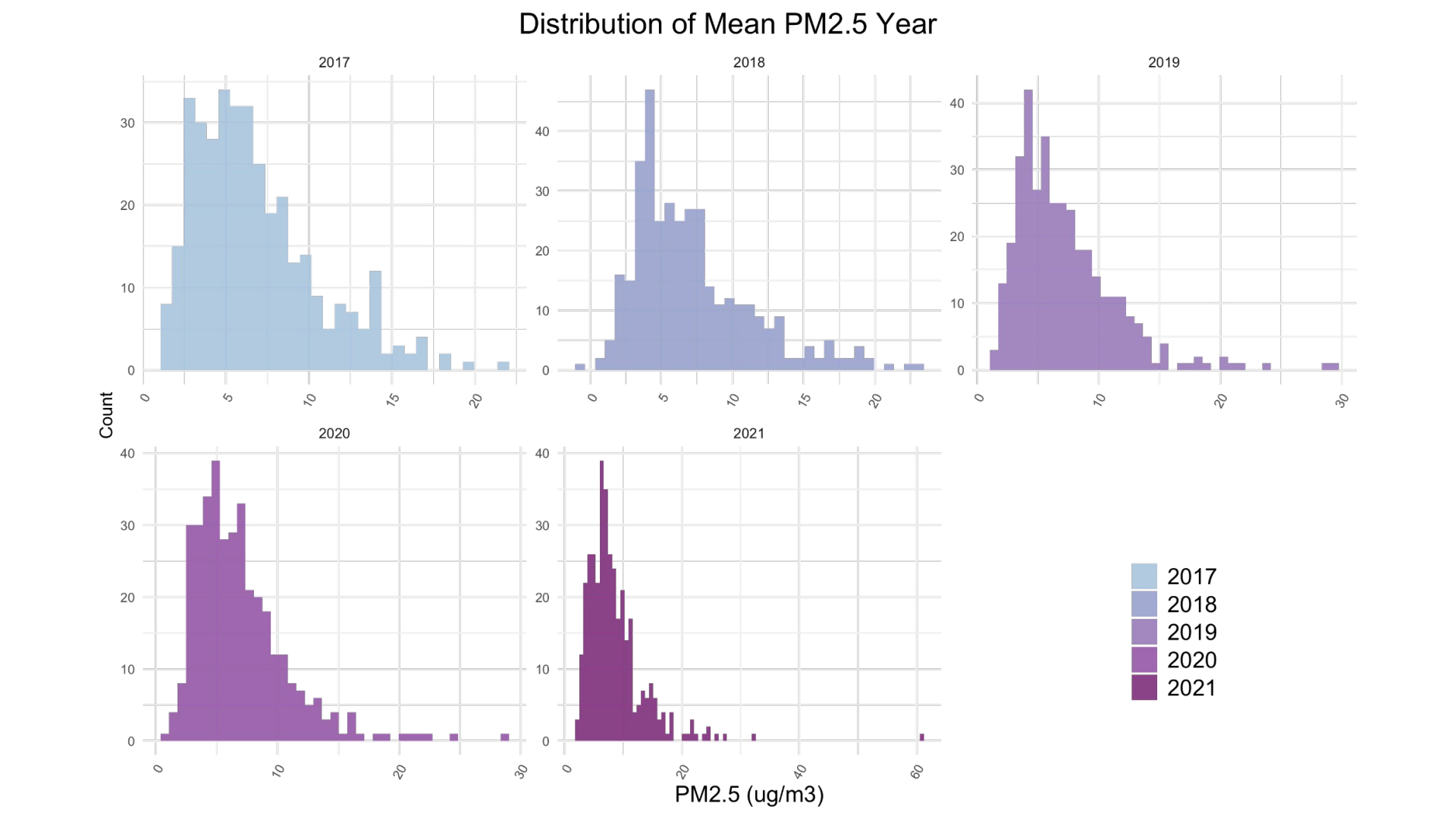
NYC Air Quality Test Sites

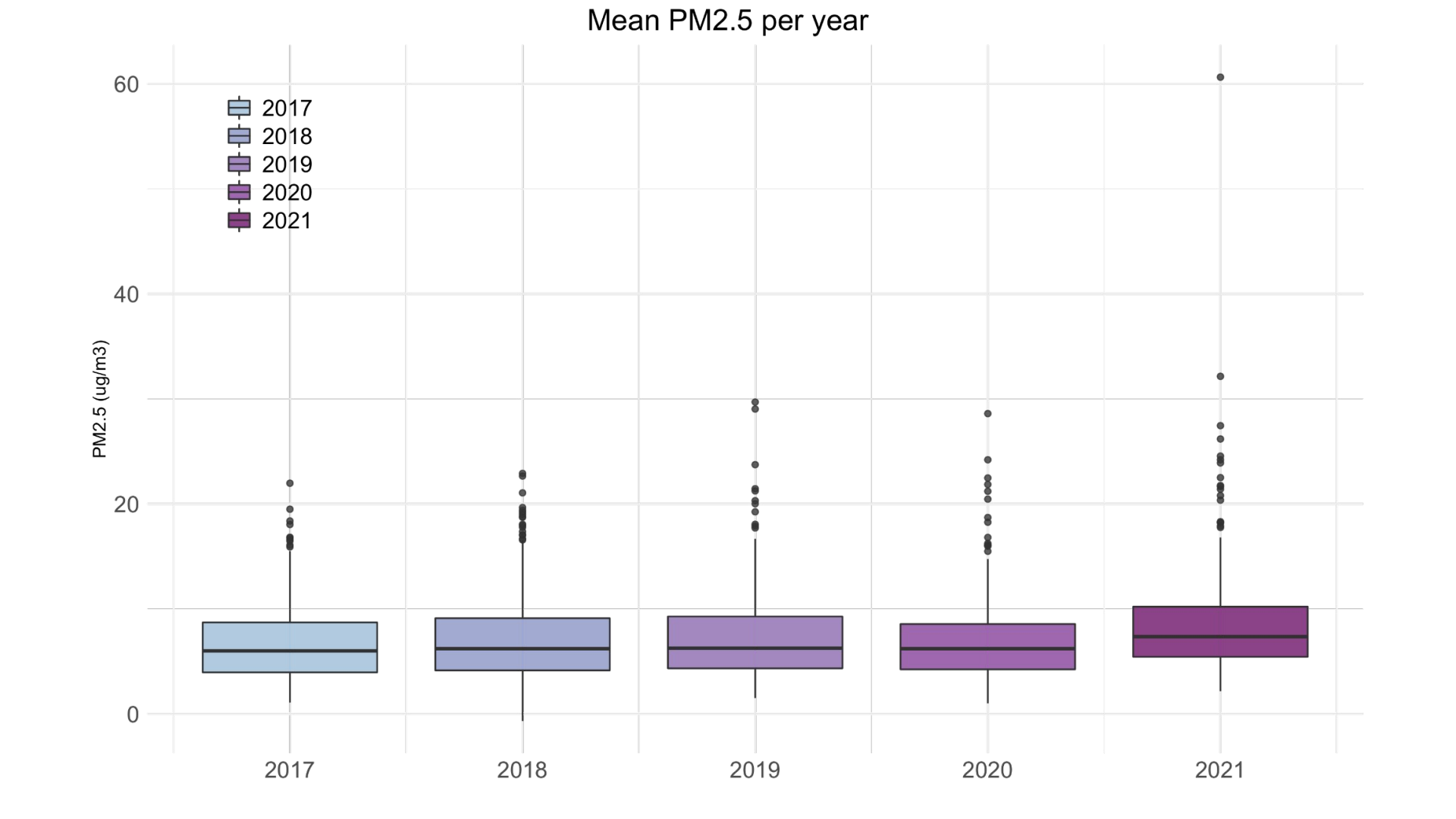
NTA	Borough
Park Cemetery	Bronx
Longwood	Bronx
Greenpoint	Brooklyn
Chinatown	Manhattan
East Village	Manhattan
East Harlem North	Manhattan
Queensboro Hill	Queens
Kew Gardens	Queens
Port Richmond	Staten Island

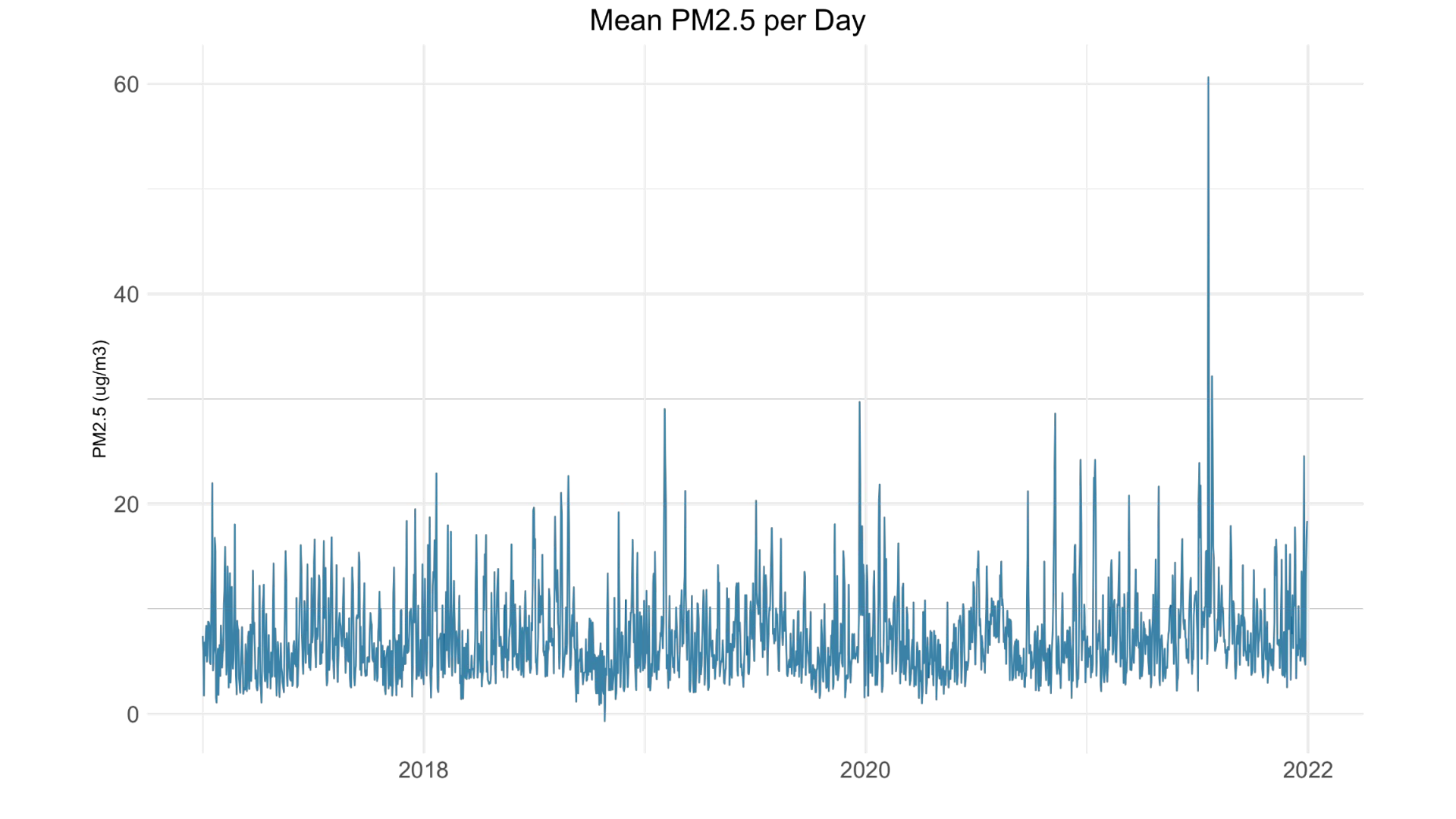

Mean PM2.5 per Year
2021
2020
2019
2018
2017
0.0
2.5
5.0
7.5
PM2.5 (ug/m3)

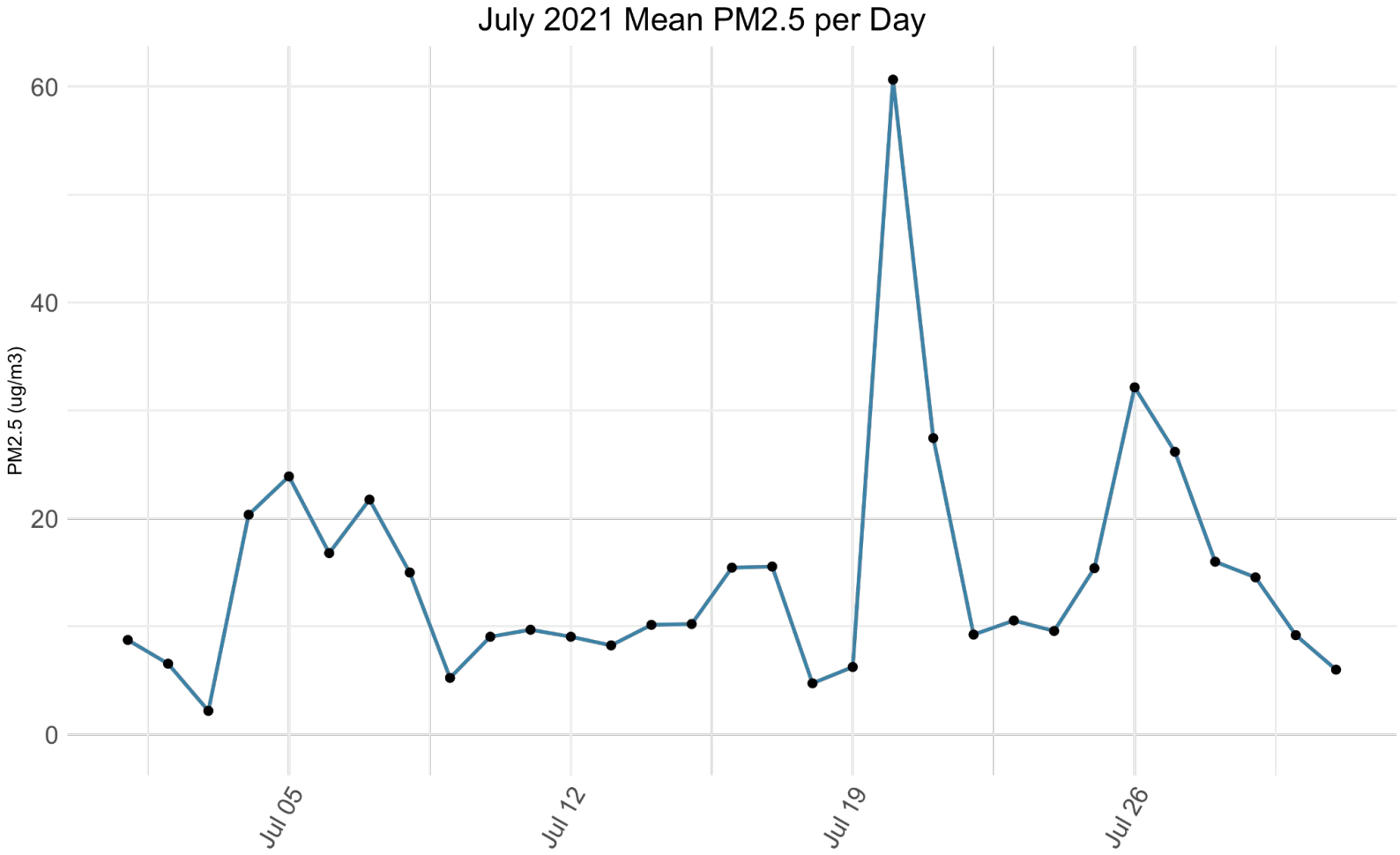

Mean PM2.5 per Month by Year
2017
2018
2019
2020
2021
PM2.5 (ug/m3)
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2017
2018
2019
2020
2021


Median PM2.5 per Month (2017 - 2021)
PM2.5 (ug/m3)
7.5
5.0
2.5
0.0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

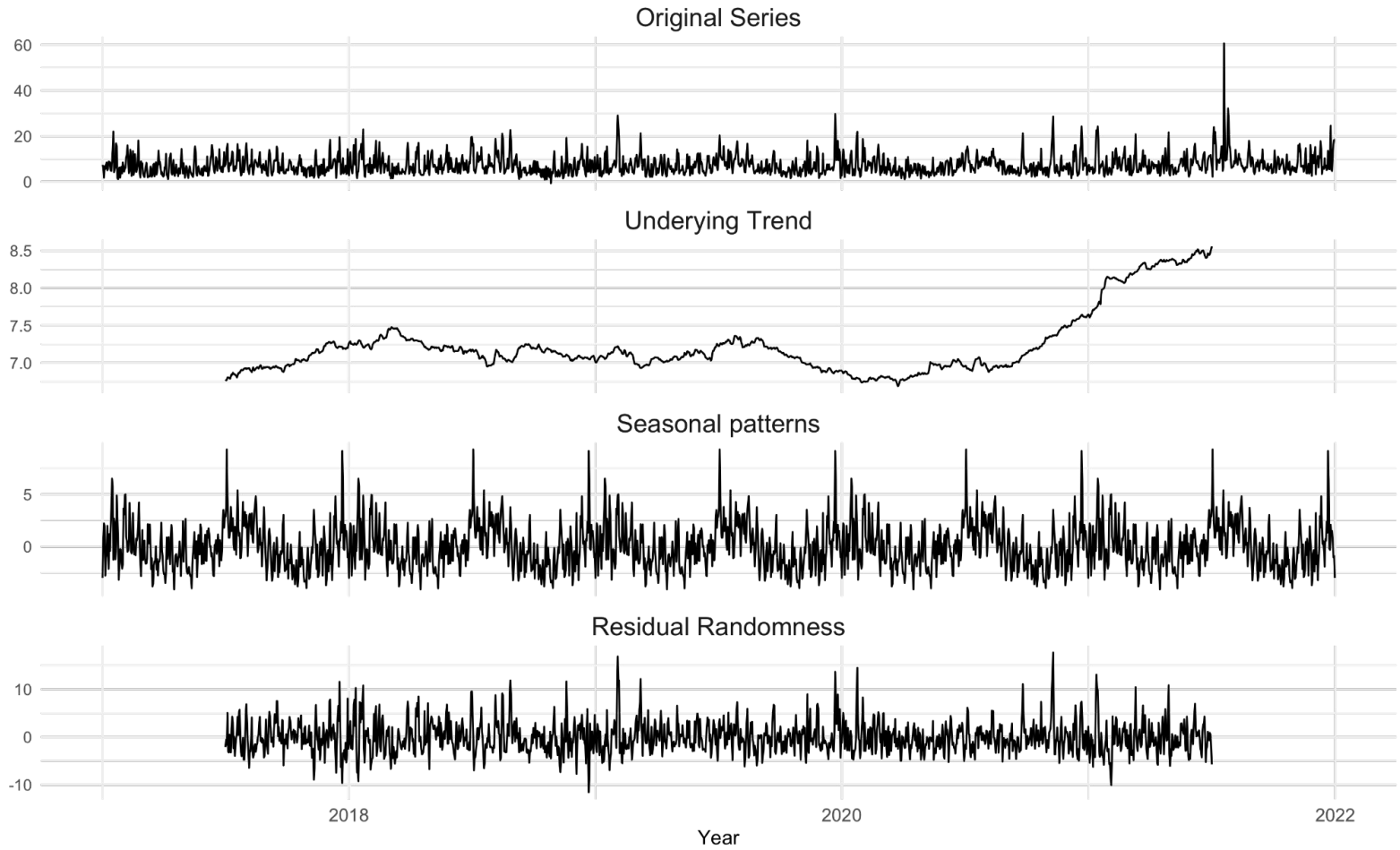

Distribution of Mean PM2.5 Year
2017
2018
2019
2020
2021
Count
PM2.5 (ug/m3)
2017
2018
2019
2020
2021

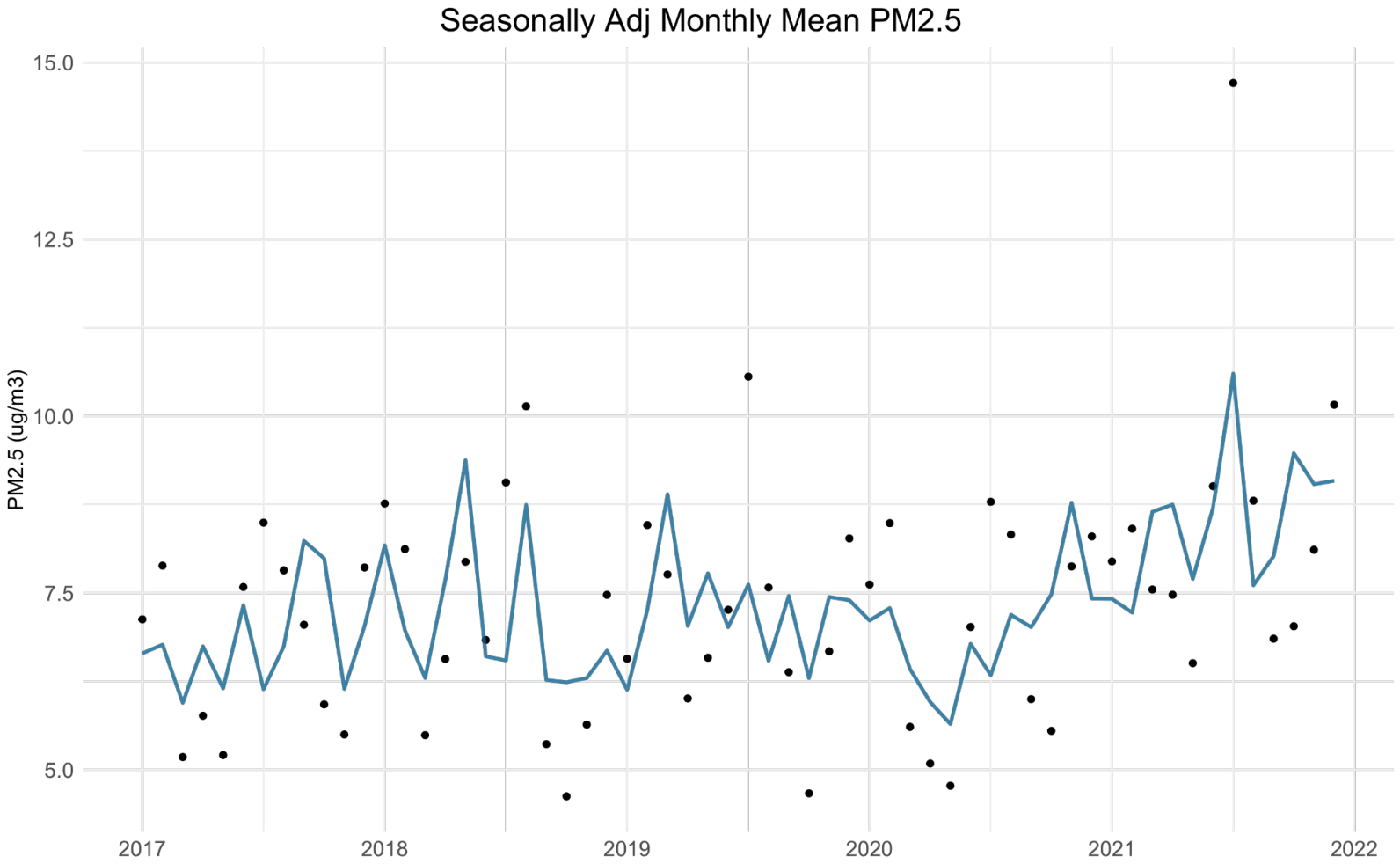

Mean PM2.5 per year
2017
2018
2019
2020
2021
PM2.5 (ug/m3)
60
40
20
0
2017
2018
2019
2020
2021

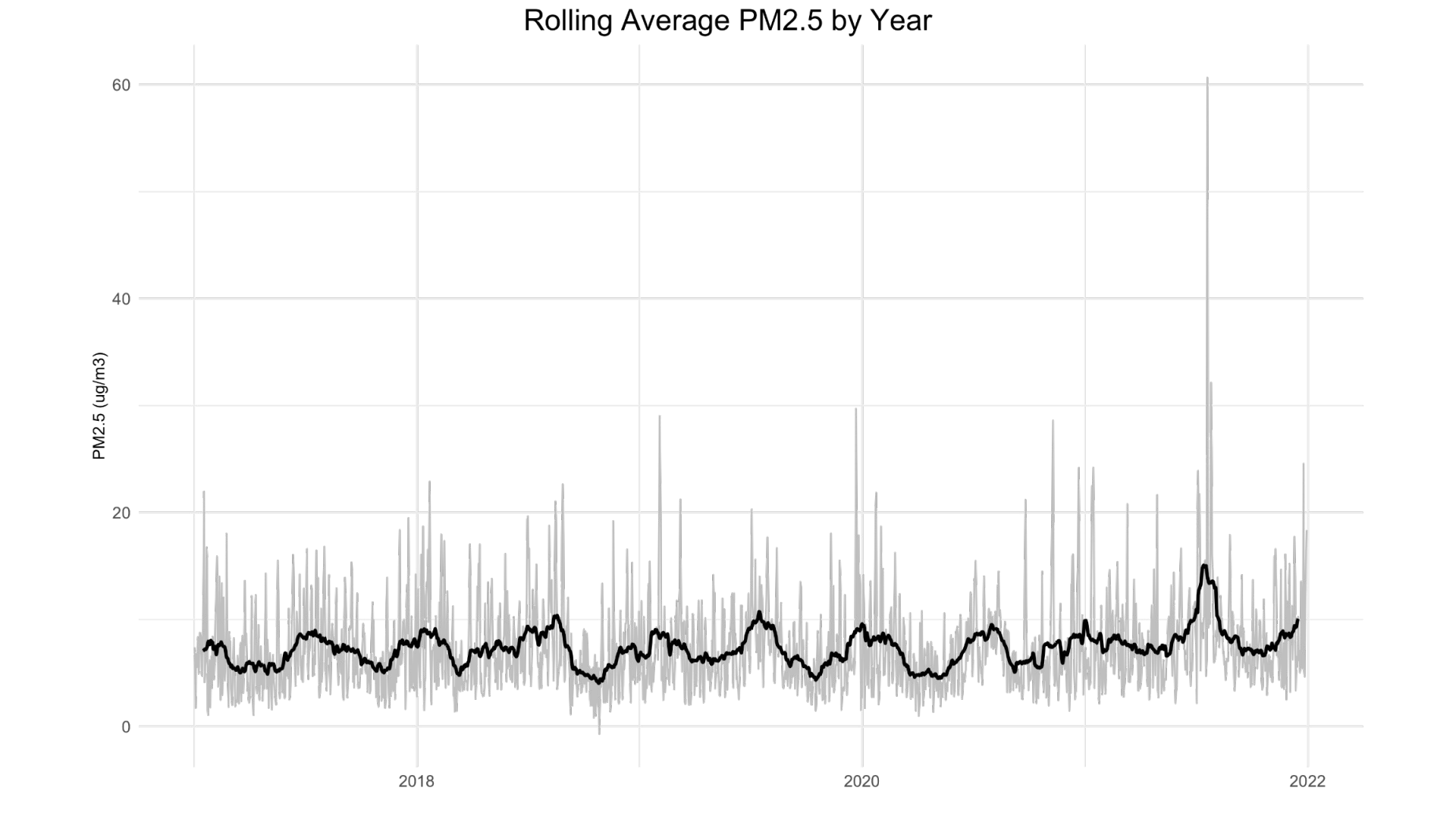

Mean PM2.5 per Day
PM2.5 (ug/m3)
60
40
20
0
2018
2020
2022

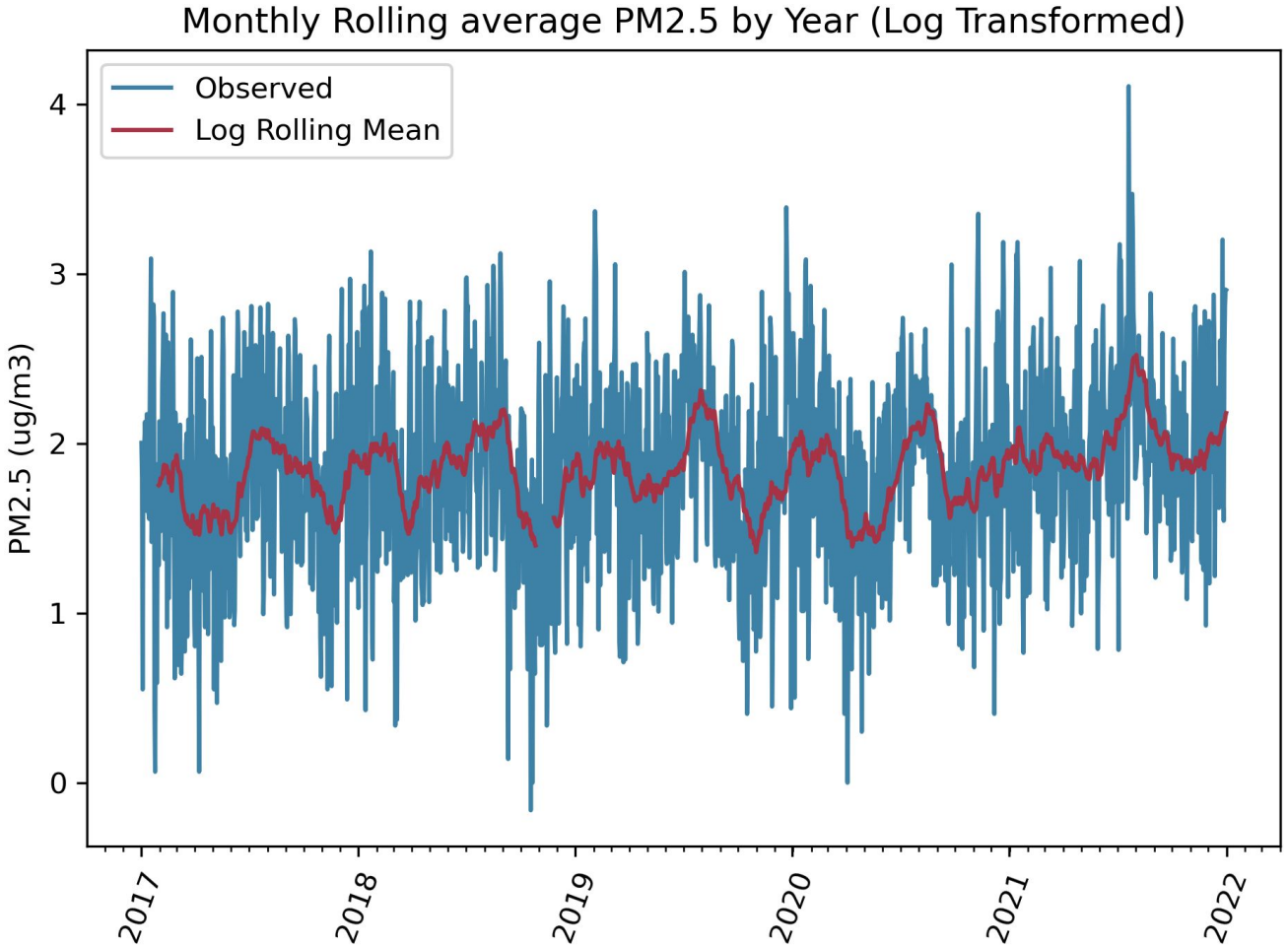

July 2021 Mean PM2.5 per Day
PM2.5 (ug/m3)
60
40
20
0
Jul 05
Jul 12
Jul 19
Jul 26

Time Series Analysis

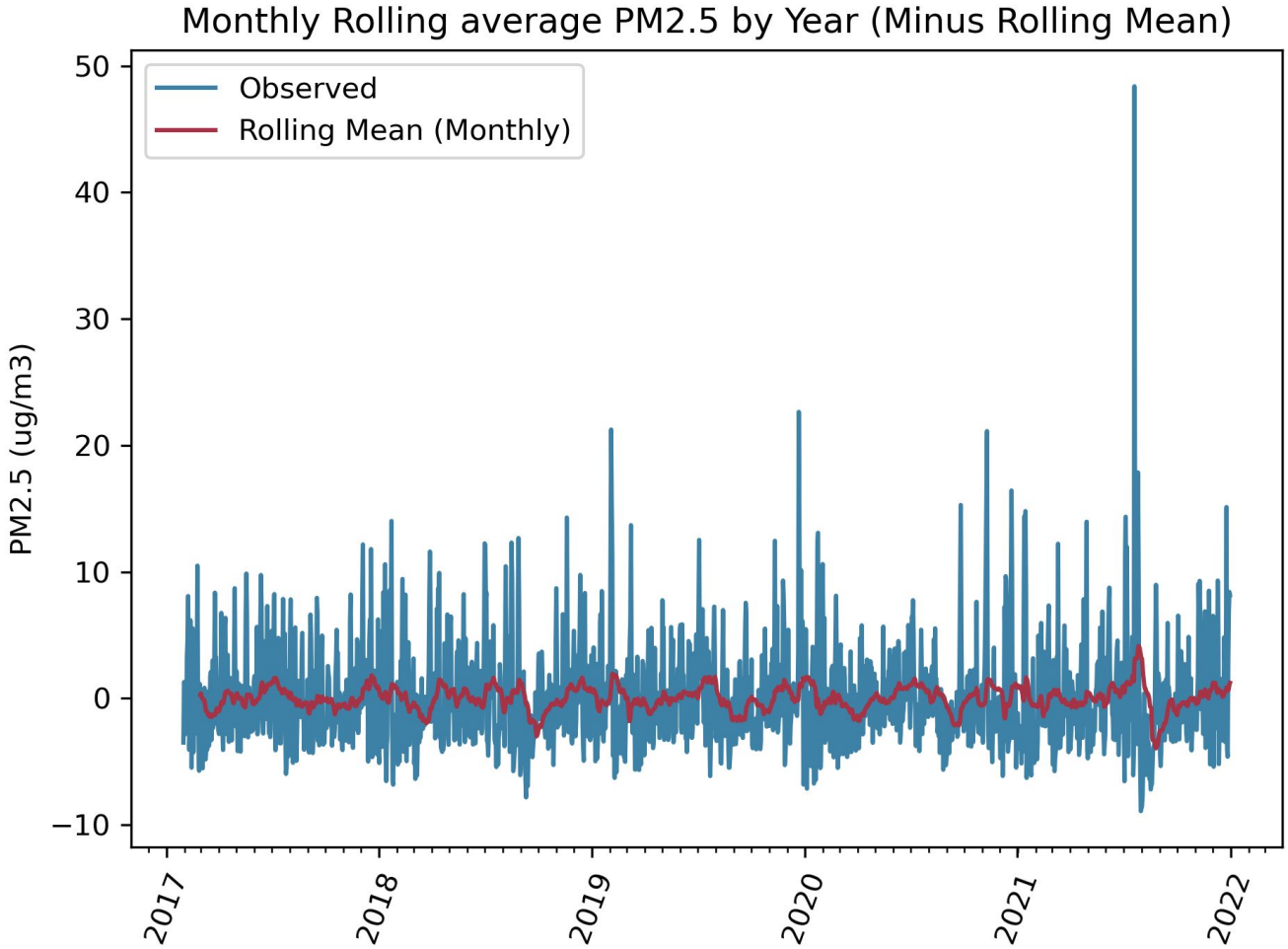

Decomposition of PM2.5 (ug/m3) 2017 - 2021
Original Series
60
40
20
0
Underying Trend
8.5
8.0
7.5
7.0
Seasonal patterns
5
0
Residual Randomness
10
0
-10
2018
2020
2022
Year


Seasonally Adj Monthly Mean PM2.5
PM2.5 (ug/m3)
15.0
12.5
10.0
7.5
5.0
2017
2018
2019
2020
2021
2022

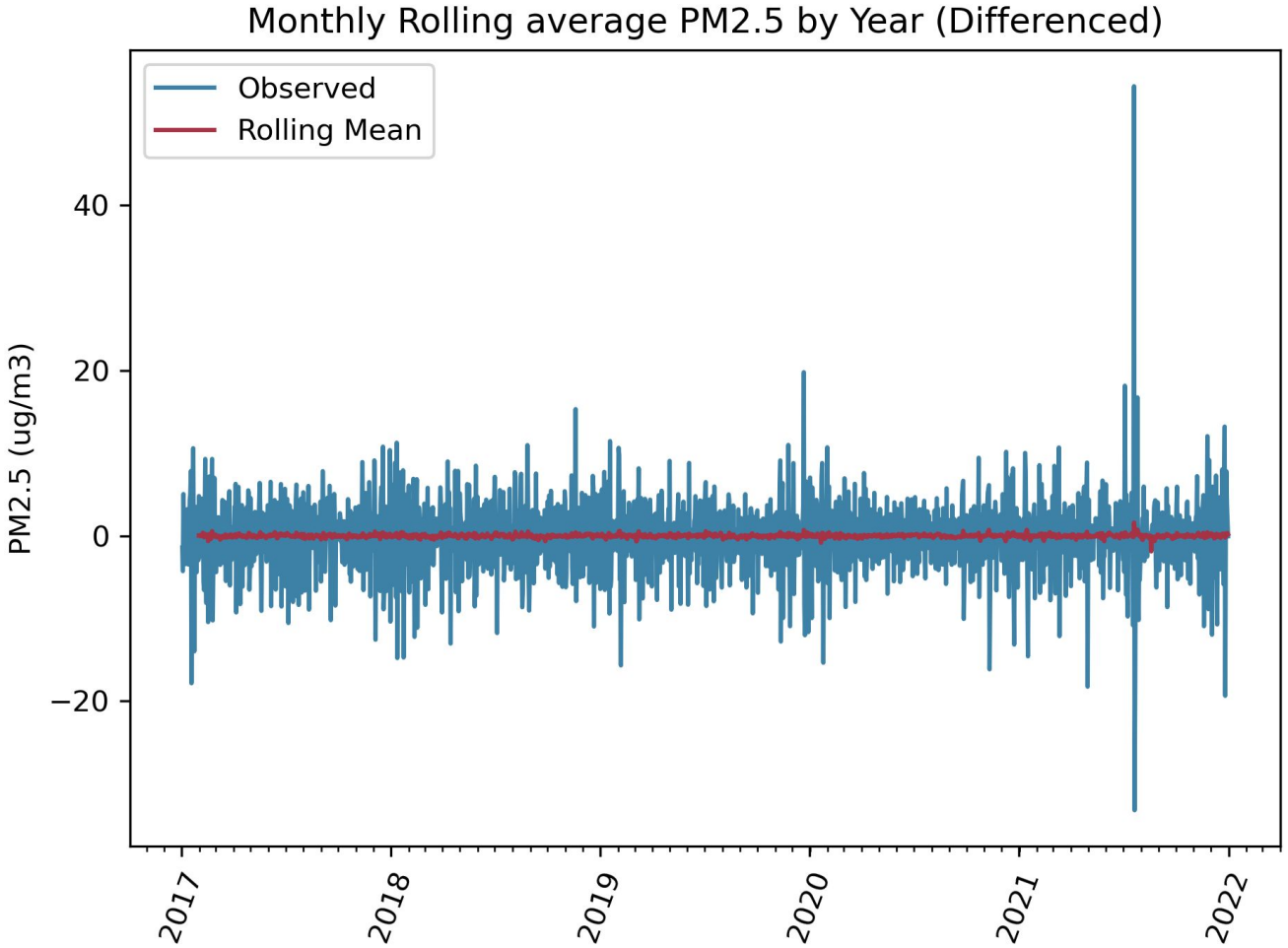

Rolling Average PM2.5 by Year
PM2.5 (ug/m3)
60
40
20
0
2018
2020
2022



Results of Dickey-Fuller test:	
Test Statistic	-5.411293
p-value	0.000003
#Lags Used	24.000000
Number of Observations Used	1800.000000
Critical Value (1%)	-3.433988
Critical Value (5%)	-2.863147
Critical Value (10%)	-2.567626

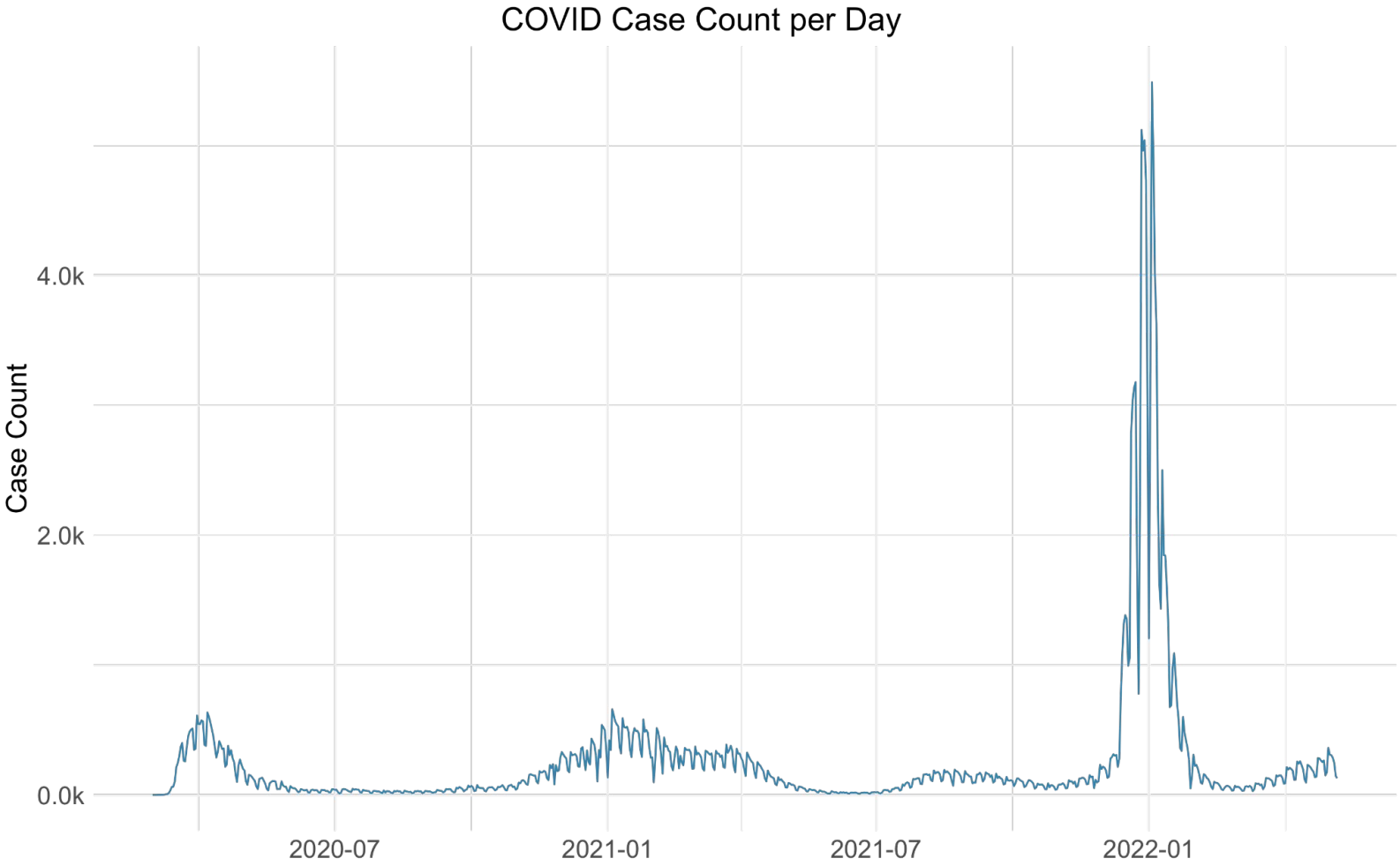


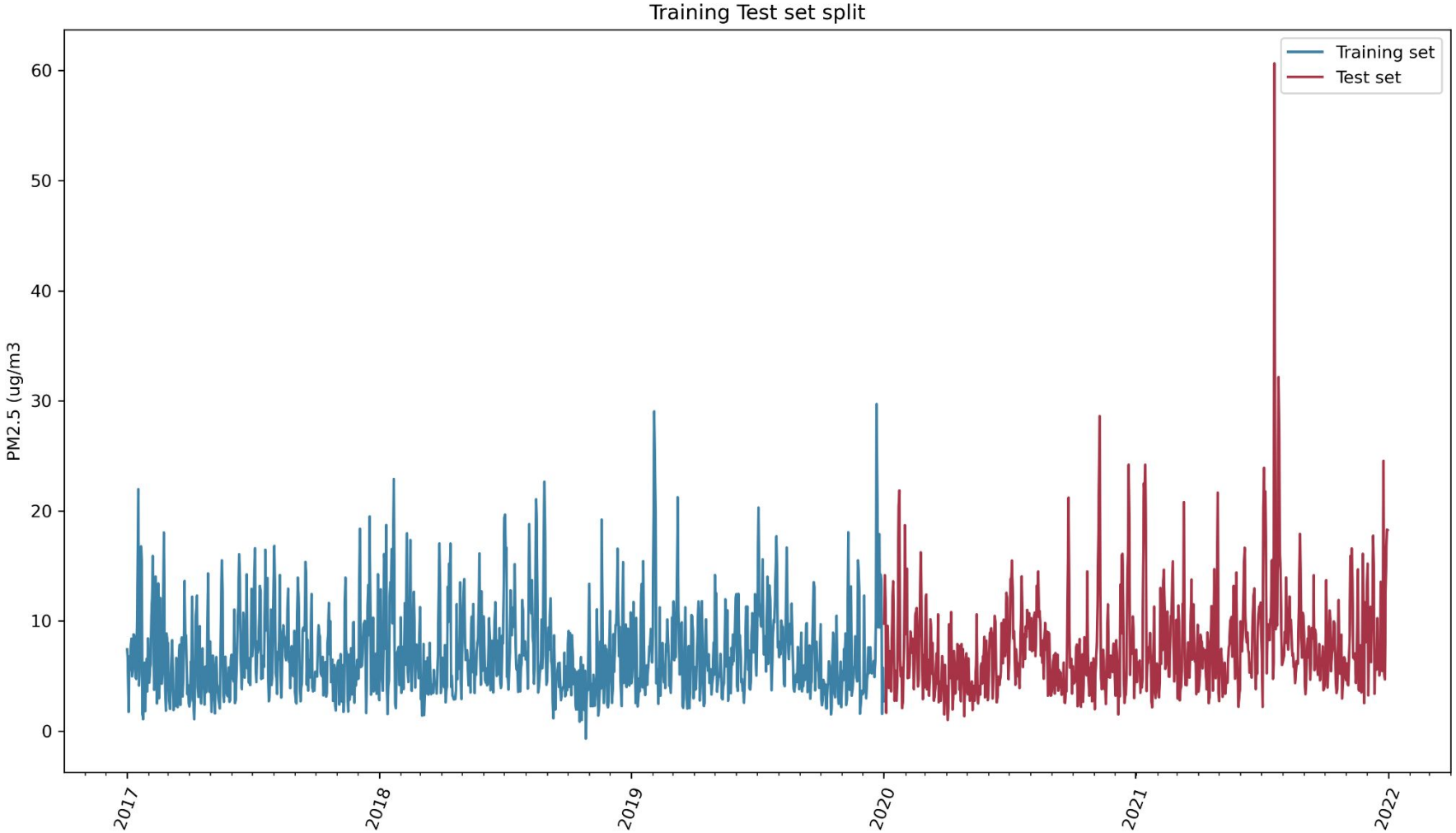
Results of Dickey-Fuller test:	
Test Statistic	-1.012625e+01
p-value	9.165461e-18
#Lags Used	2.400000e+01
Number of Observations Used	1.772000e+03
Critical Value (1%)	-3.434046e+00
Critical Value (5%)	-2.863172e+00
Critical Value (10%)	-2.567639e+00

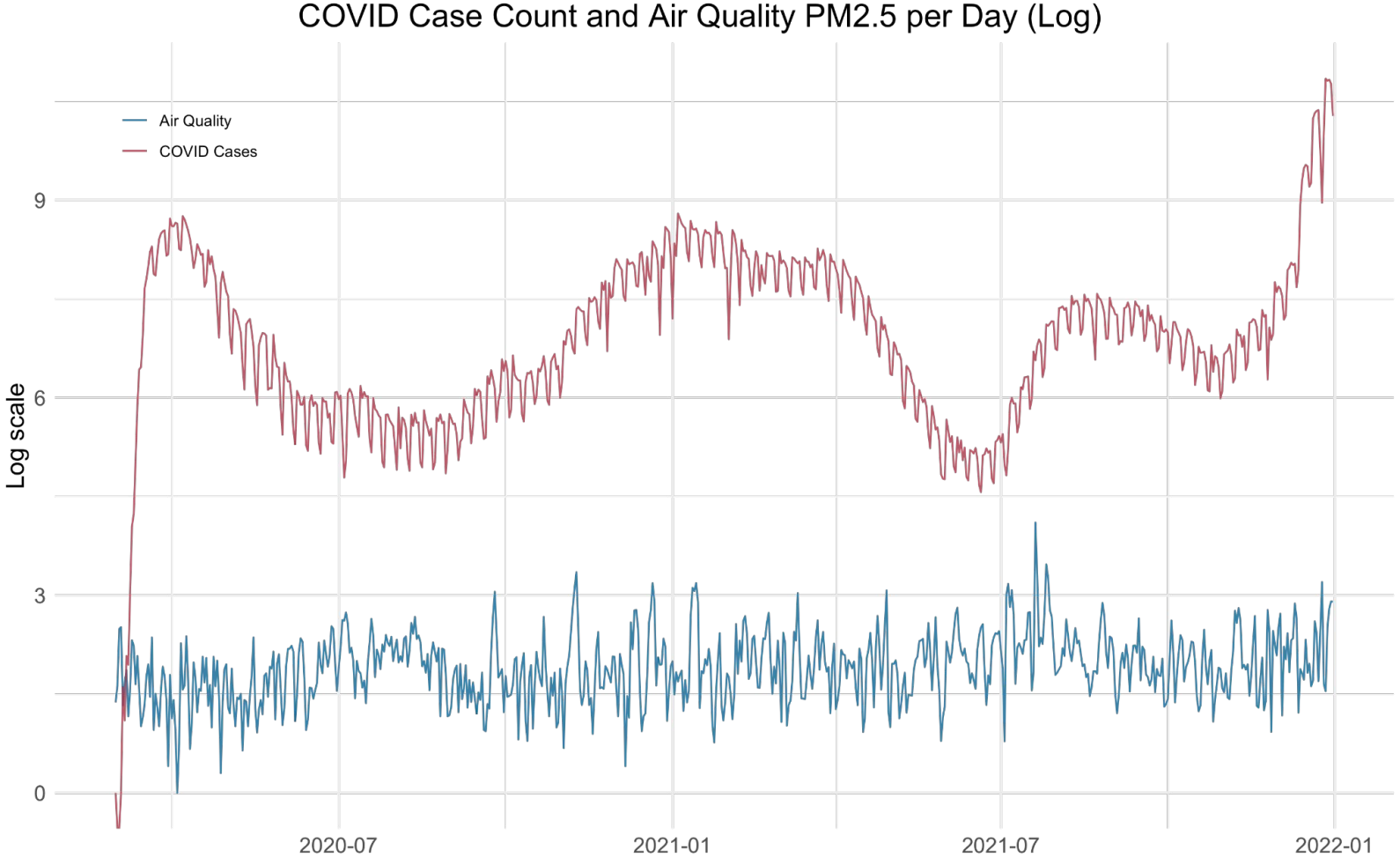


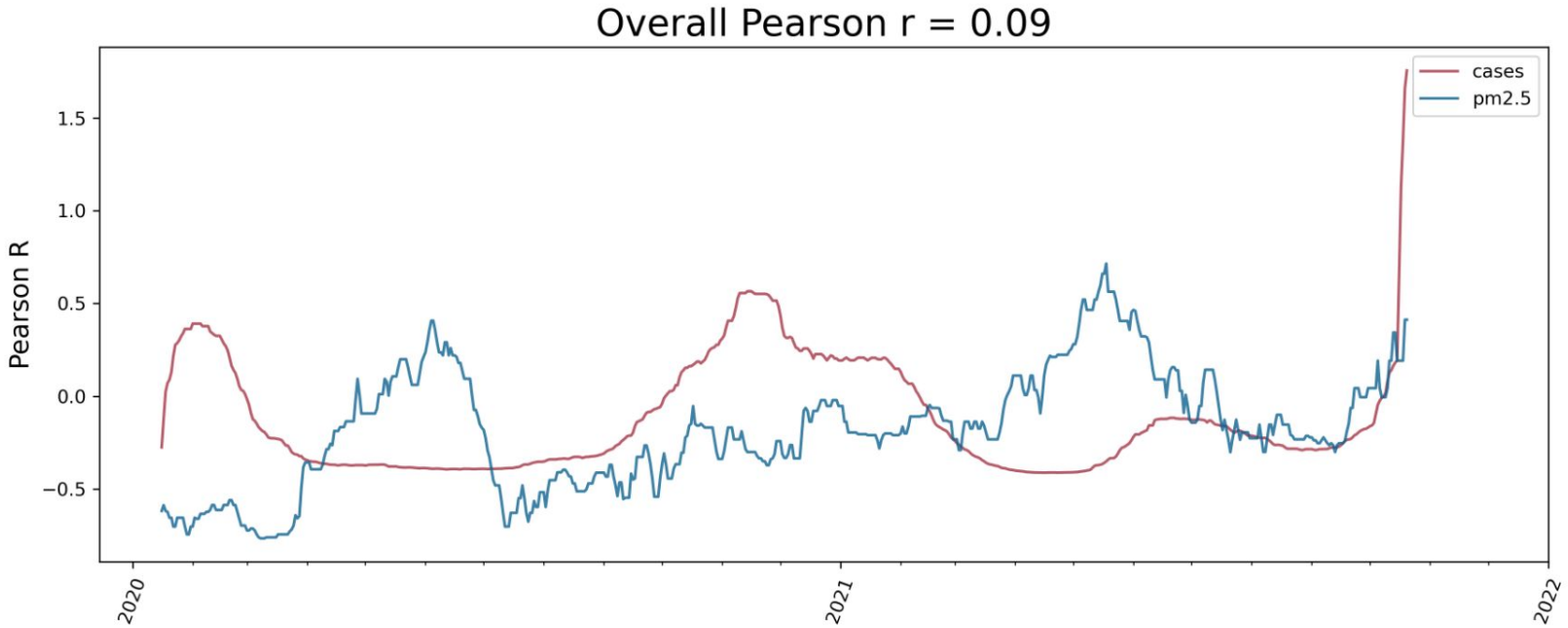
Results of Dickey-Fuller test:	
Test Statistic	-1.527813e+01
p-value	4.617527e-28
#Lags Used	2.100000e+01
Number of Observations Used	1.803000e+03
Critical Value (1%)	-3.433982e+00
Critical Value (5%)	-2.863144e+00
Critical Value (10%)	-2.567624e+00

Time Series Comparisons


COVID Case Count per Day
Case Count
4.0k
2.0k
0.0k
2020-07
2021-01
2021-07
2022-01

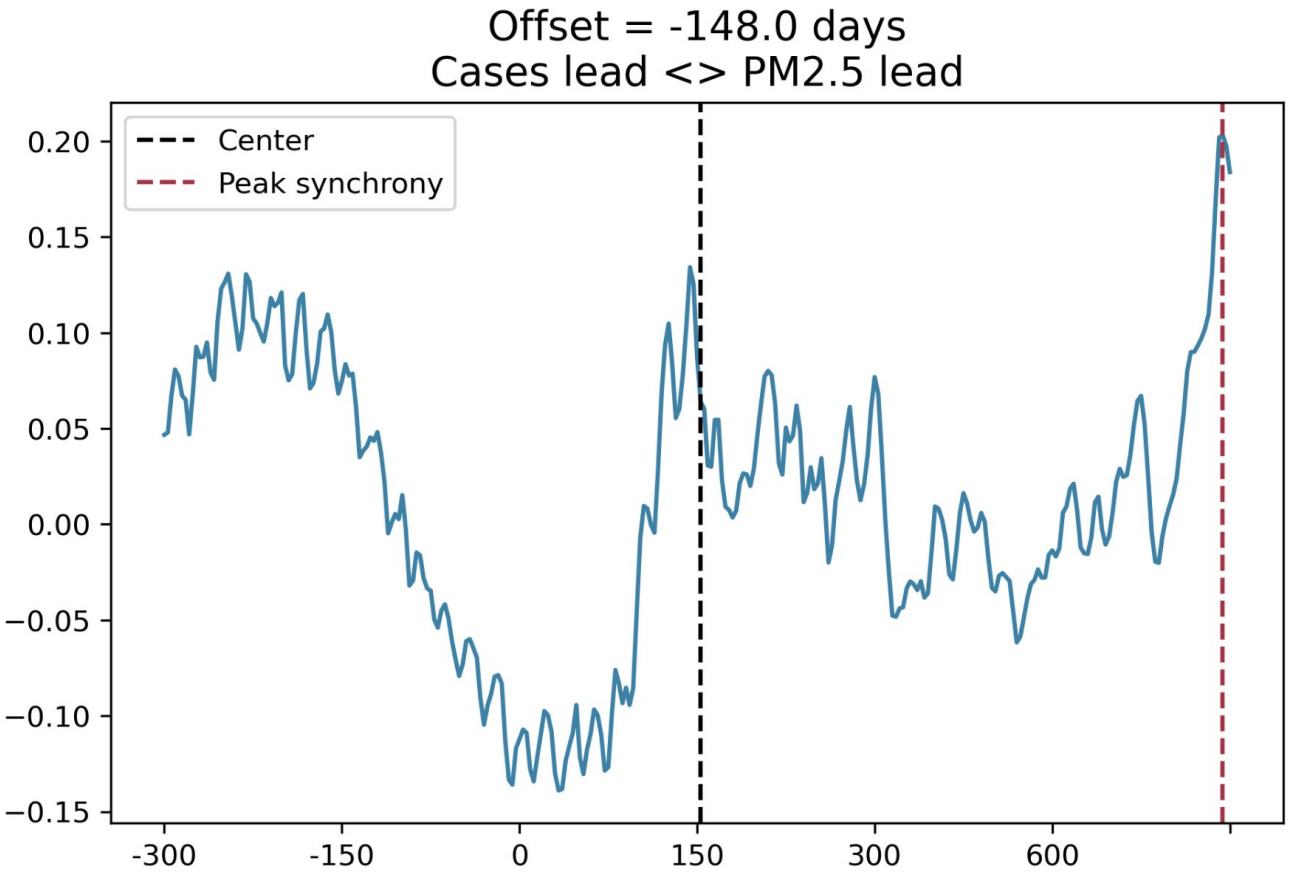

Training Test set split
Training set
Test set
PM2.5 (ug/m3
0
10
20
30
40
50
60
2017
2018
2019
2020
2021
2022


COVID Case Count and Air Quality PM2.5 per Day (Log)
Air Quality
COVID Cases
Log scale
9
6
3
0
2020-07
2021-01
2021-07
2022-01



Pearson's R

- Low correlation overall of 0.09



Time Lagged Cross Correlation (TLCC)

- Positive effect offset that PM2.5 is leading the interaction
- R is maximized when Cases are pulled forward by 148 days.

Limitations

- **NYC only:**
  - Originally wanted to do multiple cities (i.e. LA, Chicago)
  - Would different cities present different results?

- **Data integrity:**
  - COVID case reporting has been spotty at best, perhaps data is not accurate with reality

- **Modeling**
  - Did not explore any full fledged models, would like to expand in the future (i.e. Facebook Prophet, SARIMA, SARIMAX etc…)

Conclusion and Next Steps

There is a definite seasonality to concentrations of PM2.5 in NYC, shifting up in the summer and winter months, at about a 6 month cycle (180 day).

Underlying trend of PM2.5 in NYC went up from 2017-2019, went down in the beginning of 2020 and has since sky-rocketed through 2021.

While 2020 - 2021 time series of COVID and PM2.5 are not highly correlated, we can almost see a call and response type relationship. However, this may not reflect true causality. That said TLCC could mean folks are more susceptible during times of higher PM2.5 in air.

Dynamic Time Warping would better help in matching signals.

Spatial autocorrelation would account for test site location proximity.

Resources

EPA Air Data

RAQSAPI

NYC Open Data

SOCRATA API

EPA Particulate Matter (PM) Pollution

Air Pollution and the Health of New Yorkers

Comparing Time Series Data by Jin Hyun Cheong

Thank you