

NYair analysis

Ricardo Barros

2/24/2022

Introduction

The report aims to analyse the four pollutants observations made by 25 air quality monitors that are scattered across New York city. The measurements were made between 01-01-1990 and 31-12-2019, and the pollutants analysed here are:

- Carbon monoxide (CO) [ppm]
- Sulfur dioxide (SO2) [ppb]
- Ozone (O3) [ppm]
- Nitrogen dioxide (NO2) [ppb]

ppm: parts per million ppb: parts per billion

Data source: NOAA Historical Air Quality

(<https://console.cloud.google.com/marketplace/product/epa/historical-air-quality?project=fiery-webbing-340301>)

See also `sql_code` (<https://github.com/ricardovobarros/NYair>) file to see how the database was queried.

Data Processing

Looking into the first row:

```
##  site_num  parameter_name  concentration  date_local  longitude  latitude
## 1    0071  Carbon monoxide    1.173684  1990-01-01  -73.98368  40.69578
## 2    0011  Sulfur dioxide    14.712500  1990-01-01  -73.94722  40.73277
## 3    0073  Sulfur dioxide    17.958333  1990-01-01  -73.90958  40.81149
## 4    0056  Carbon monoxide    1.568421  1990-01-01  -73.96661  40.75912
## 5    0063      Ozone        0.024706  1990-01-01  -74.01264  40.71149
## 6    0056  Carbon monoxide    1.608696  1990-01-01  -73.96661  40.75912
```

What is the sampling time period?

```
##           min           max
## 1 1990-01-01 2019-12-31
```

How many measurements were made and what are their time period?

```
##           parameter           period  observations
## 1  Carbon monoxide 1990-01-01 to 2019-12-31    104192
## 2  Sulfur dioxide  1990-01-01 to 2019-12-31    102677
## 3      Ozone       1990-01-01 to 2019-12-31     50026
## 4 Nitrogen dioxide (NO2) 1990-01-01 to 2019-12-31    38892
## 5 Outdoor Temperature 1992-06-09 to 2019-12-31     21793
```

What is the average number of observations between 1990 and 2019?

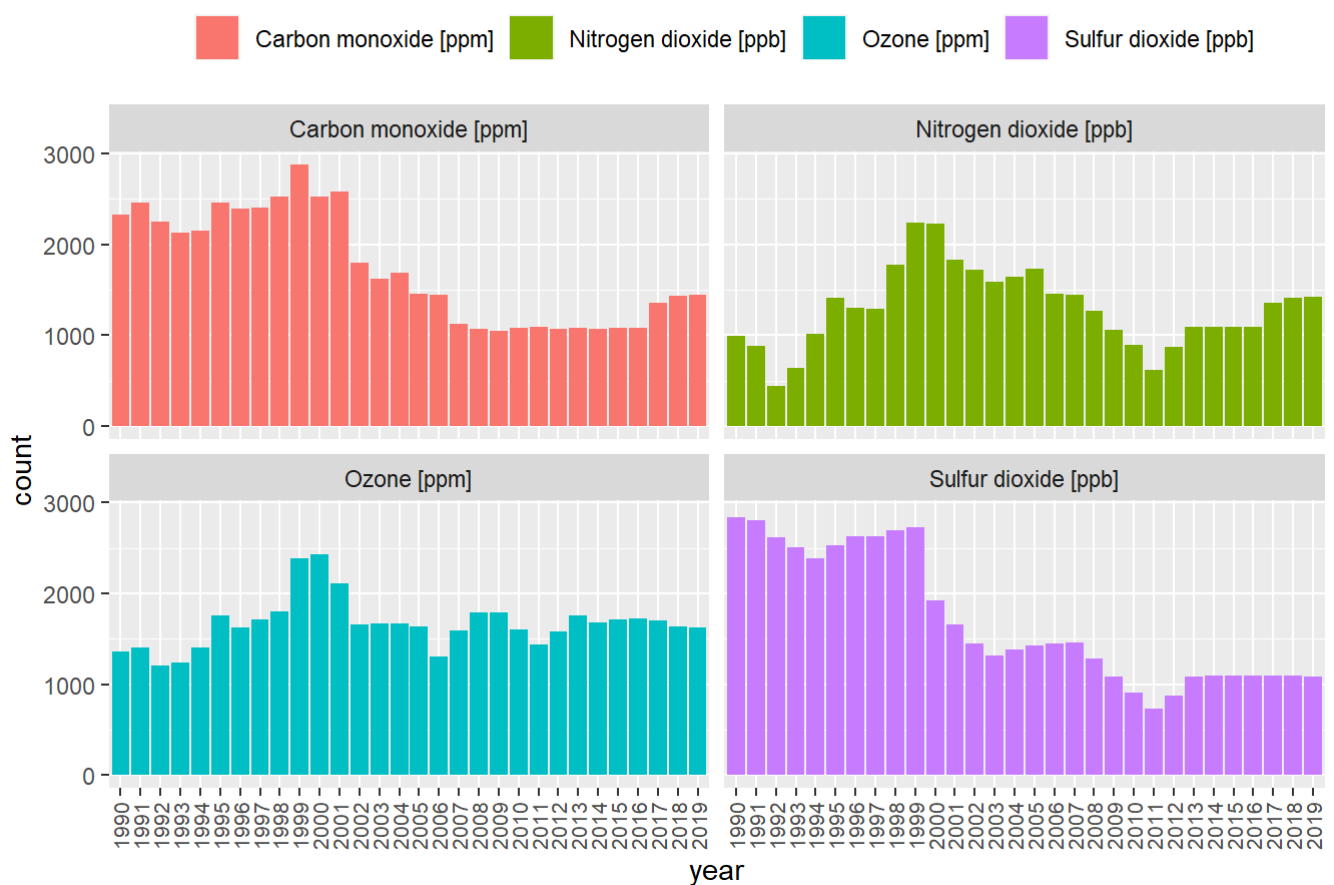
- On average, Carbon Dioxide had the large number of observations.

```
## # A tibble: 5 x 2
##   parameter_name      mean
##   <chr>              <dbl>
## 1 Carbon monoxide [ppm] 4.76
## 2 Nitrogen dioxide [ppb] 3.59
## 3 Outdoor Temperature[°F] 2.24
## 4 Ozone [ppm]          4.57
## 5 Sulfur dioxide [ppb]  4.65
```

How frequent were the measurements throughout the years?

- CO and SO₂ had approximately 2500 observation per year until around the year 2000, when the total number of observations drop to around 1000.
- O₃ observations kept almost constant over the year (around 1500 observations per year) with a peak between 1999 and 2001.

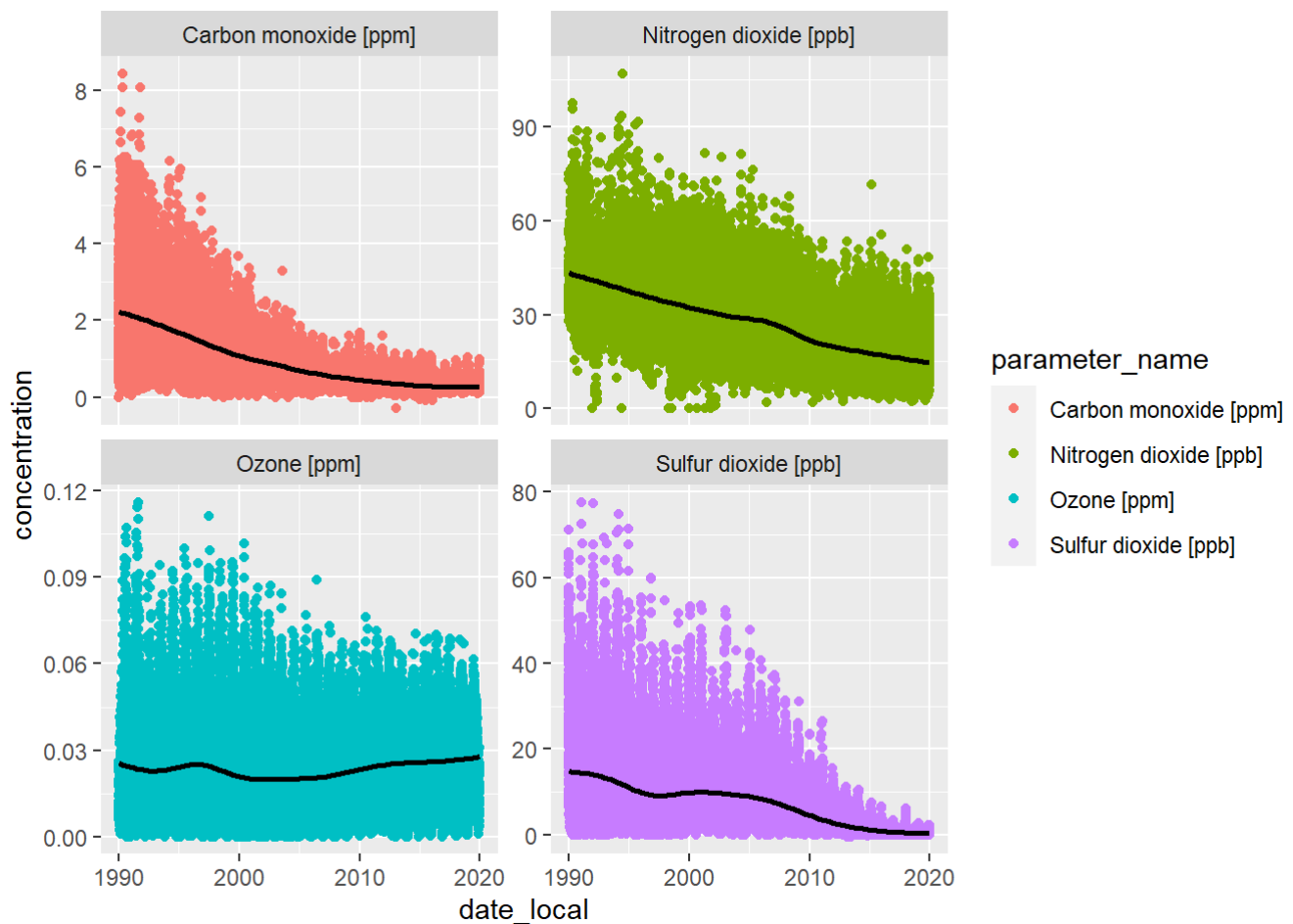
Observations totals over the years



Data Analysis

How the concentration of pollutants evolve over time?

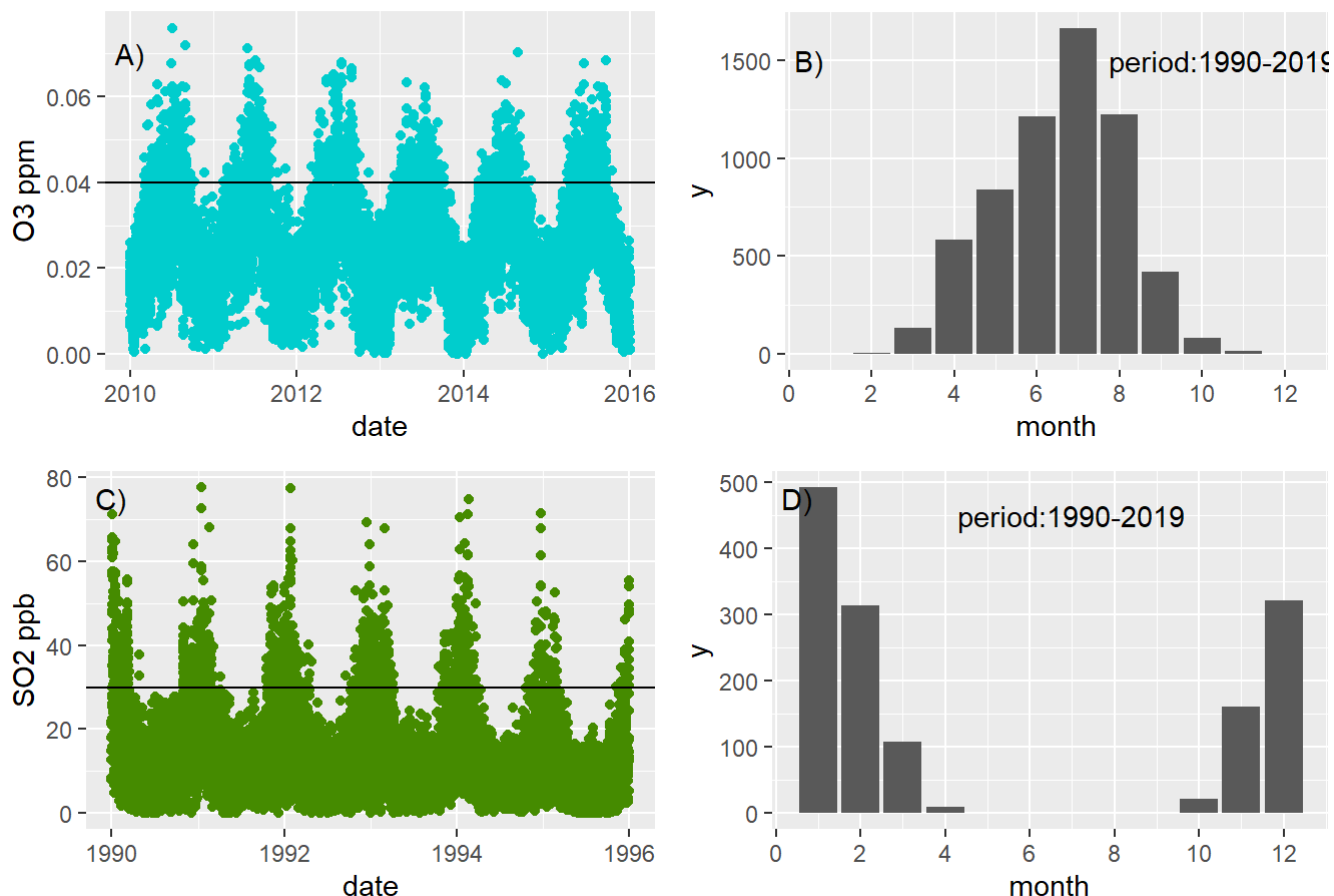
- The trend of all pollutants, but ozone, seems to decrease over time.
- Specially Sulfur dioxide (SO₂) and Ozone (O₃) present regular oscillation peaks and valleys.



Have the peaks and valleys of O3 and SO2 a pattern?

- Looking closely between 2010 and 2016, O3 presents a pattern where peaks occur round the middle of the year and valleys around the turning of the years (**Figure A**).
- Figure B** shows the frequency distribution of the months that presented mean concentration over 0.04 ppm of O3 (black line in **Figure A**) within 1990 and 2019. Clearly, monitors measure on average higher concentrations of O3 between April and September.
- A mirror phenomena occurs when measuring SO3. Within the first and last months, monitors detect a higher concentration of SO3. For instance, **Figure C** show the weave-pattern of SO3 between 1990 and 1996, and **Figure D** shows the frequency distribution of the months that have daily average concentration above 30 ppb between 1990 and 2019.

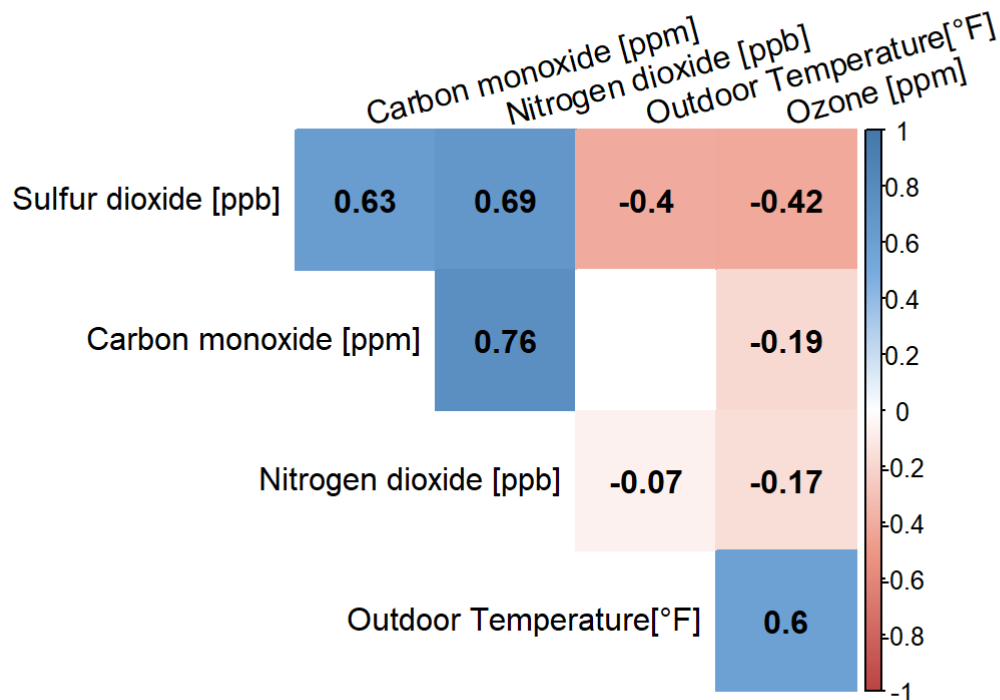
O3 and SO2 concentration patterns



What is the linear correlation among the parameters between 1990 and 2019?

- CO, SO2 and NO2 have a strong correlation among themselves. That may indicate that these tree pollutants share common sources.
- Ozone (O3) and temperature also present strong correlation. It is common knowledge that ozone is unstable substance on atmosphere pressure and high temperatures. Thus, one could expect that O3 and temperature have negative correlation, if the source of ozone were constant over the years. However, the positive correlation of 0.6 indicates the temperature is positive correlated with the amount of O3 released on New York proximity.

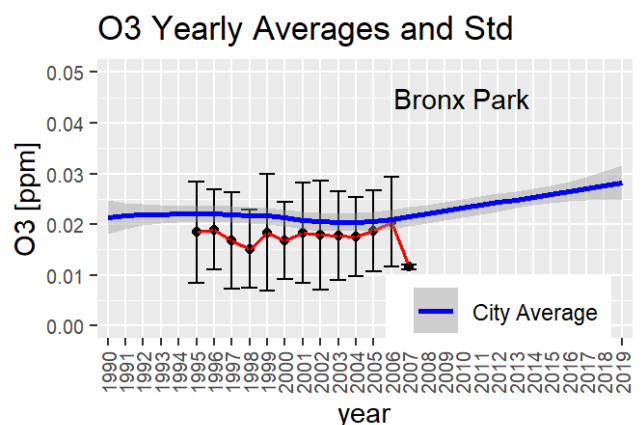
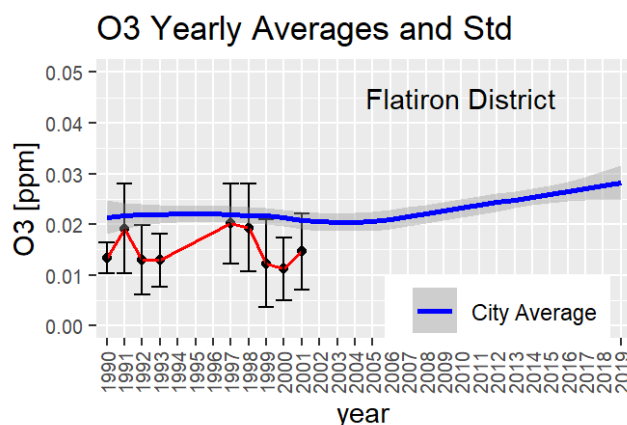
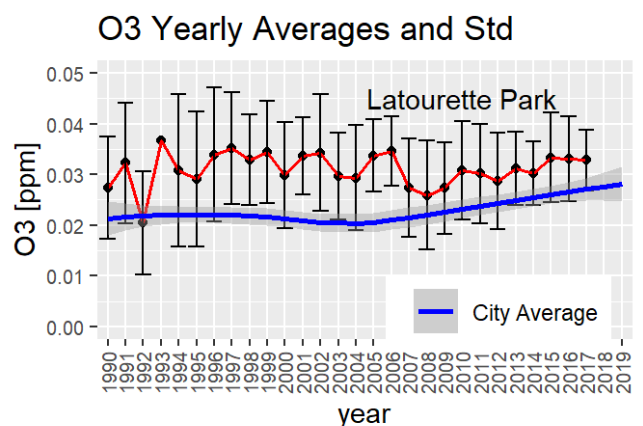
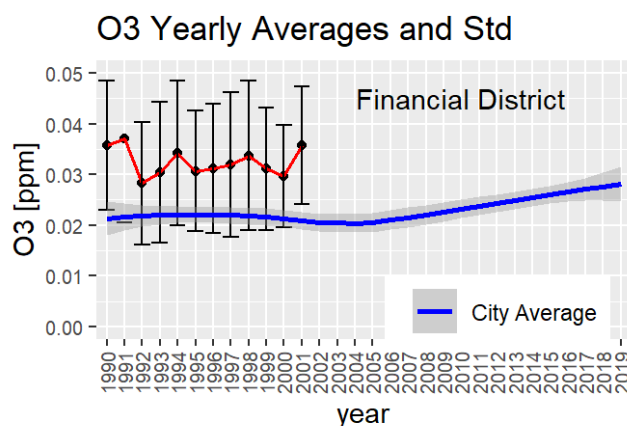
Parameter Correlation with significance p-value <= .01



What are the four most relevant Ozone observations and their neighborhood?

- Neighborhoods Financial District and Latourette Park measure consistently higher concentration of ozone than the city average.
- The opposite is observed for the neighborhoods Bronx Park and Flatiron.

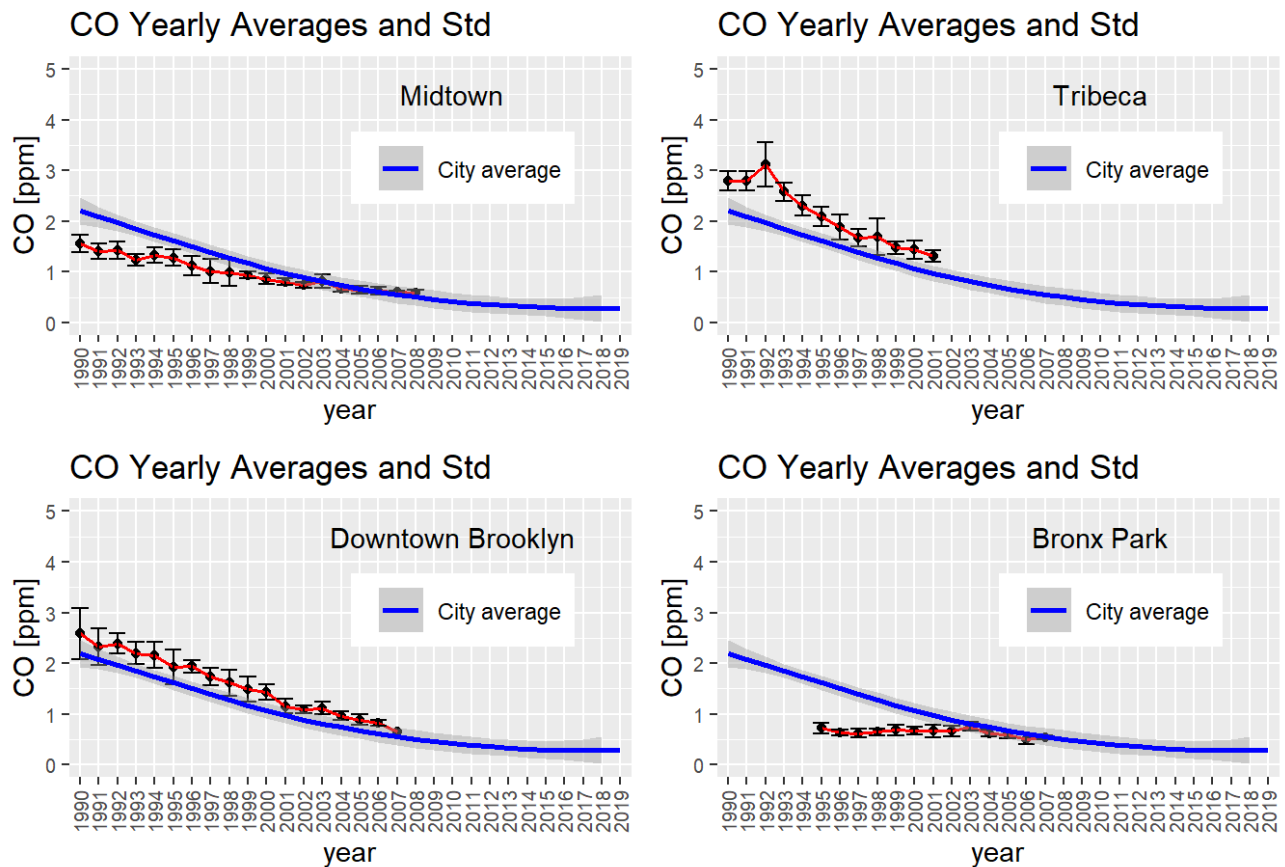
O3 Anual Average and Standard Deviation in Four Neighborhoods



What are the four most relevant Carbon Monoxide observations and their neighborhood?

- Neighborhoods Tribeca and Downtown Brooklyn measure slightly higher concentration of CO than the city average.
- The opposite is observed for the neighborhoods Midtown and Bronx Park.

CO Annual Average and Standard Deviation in Four Neighborhoods

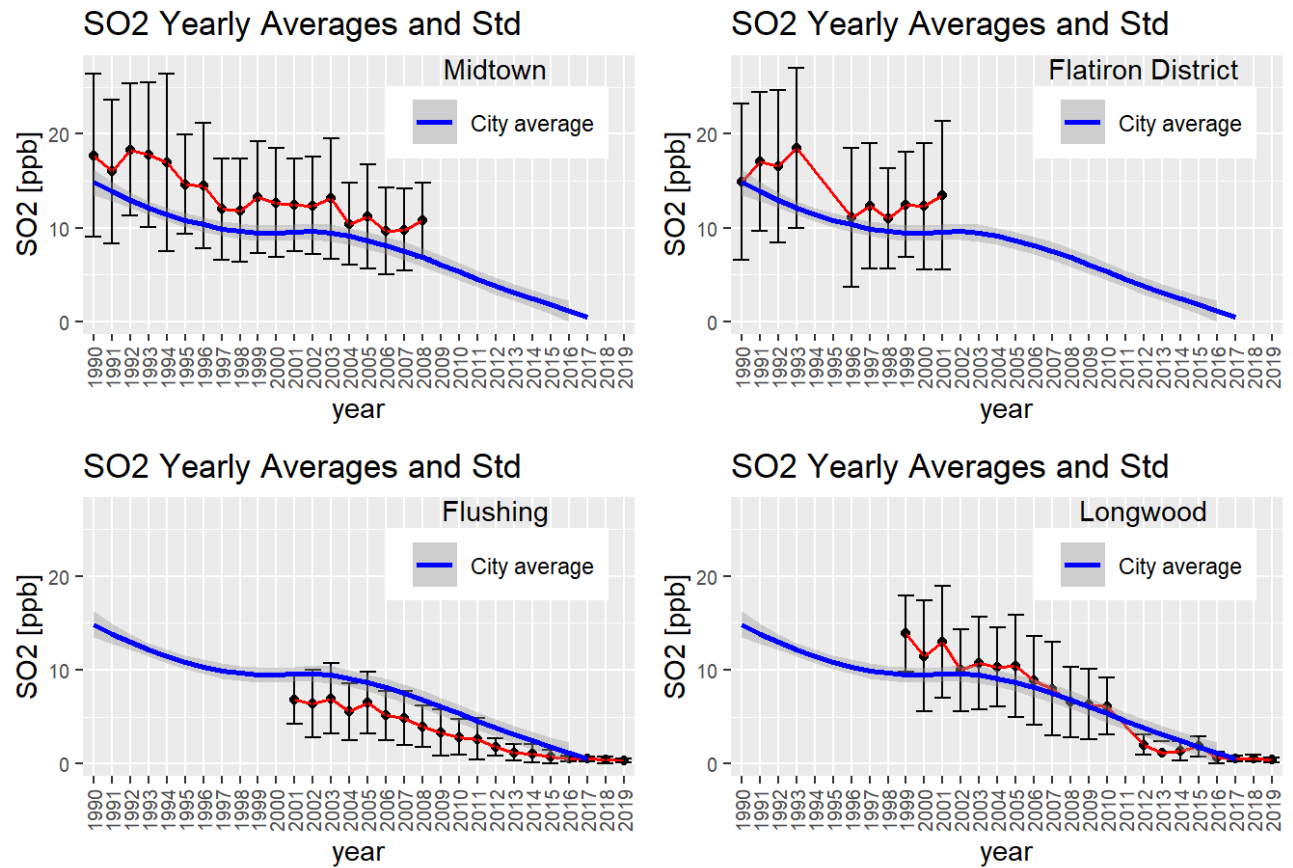


What are the four most relevant Sulfur Dioxide observations and their neighborhood?

- Neighborhoods Midtown and Flatiron District measure higher concentration of SO₂ than the city average in the 90s and part of the 2000s.
- The monitor located in Flushing detected consistently lower values than the city average.
- The monitor located in Longwood detected SO₂ concentrations around average. Around the year 2000 SO₂ yearly average crosses downwards the city average.
- The City Average decreasing trend may be biased because the monitors with historical data from the 90s to middle 2000s are different than the one that have historical data from middle 2000s onwards.

However, the trends of monitors separately have a downward trend.

SO2 Anual Average and Standard Deviation in Four Neighborhoods

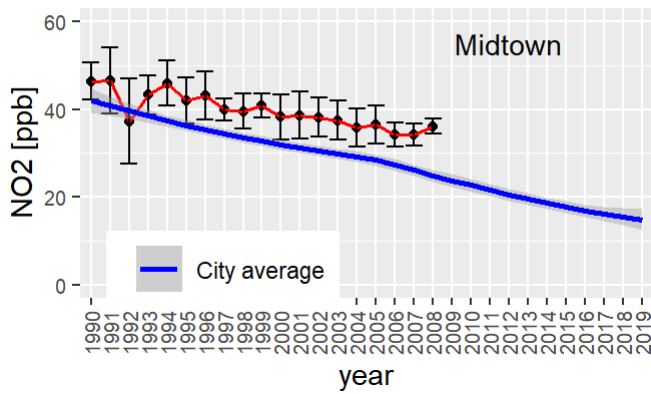


What are the four most relevant Nitrogen Dioxide observations and their neighborhood?

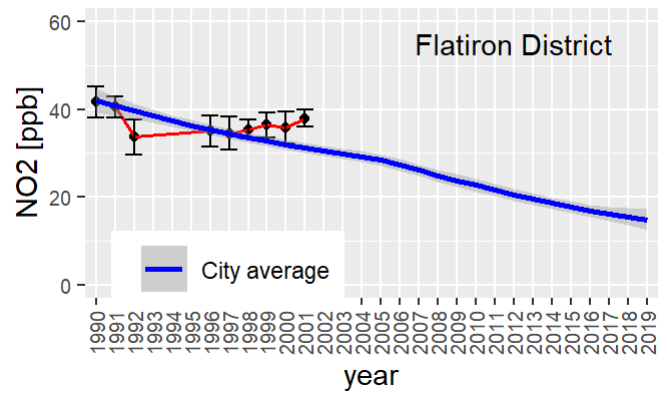
- Midtown monitor measured consistently concentrations of NO₂ above the city average.
- Contrarily, the monitor by FLushing measured consistently yearly mean concentration below city average.
- The monitor located in Bay Side neighborhood has data of only four years.

NO2 Annual Average and Standard Deviation in Four Neighborhoods

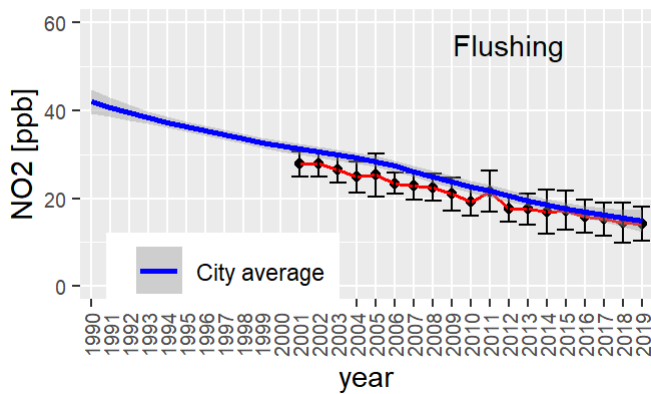
NO2 Yearly Averages and Std



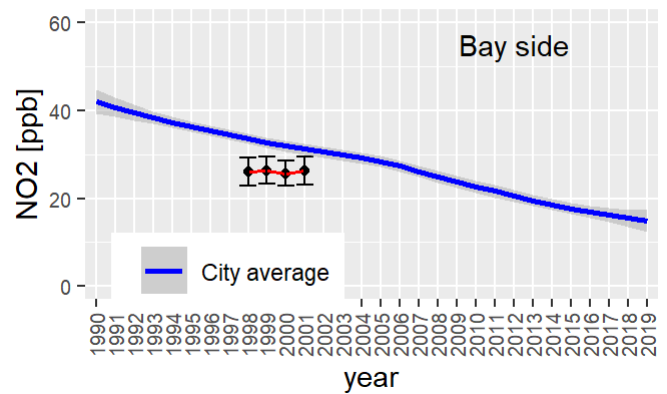
NO2 Yearly Averages and Std



NO2 Yearly Averages and Std

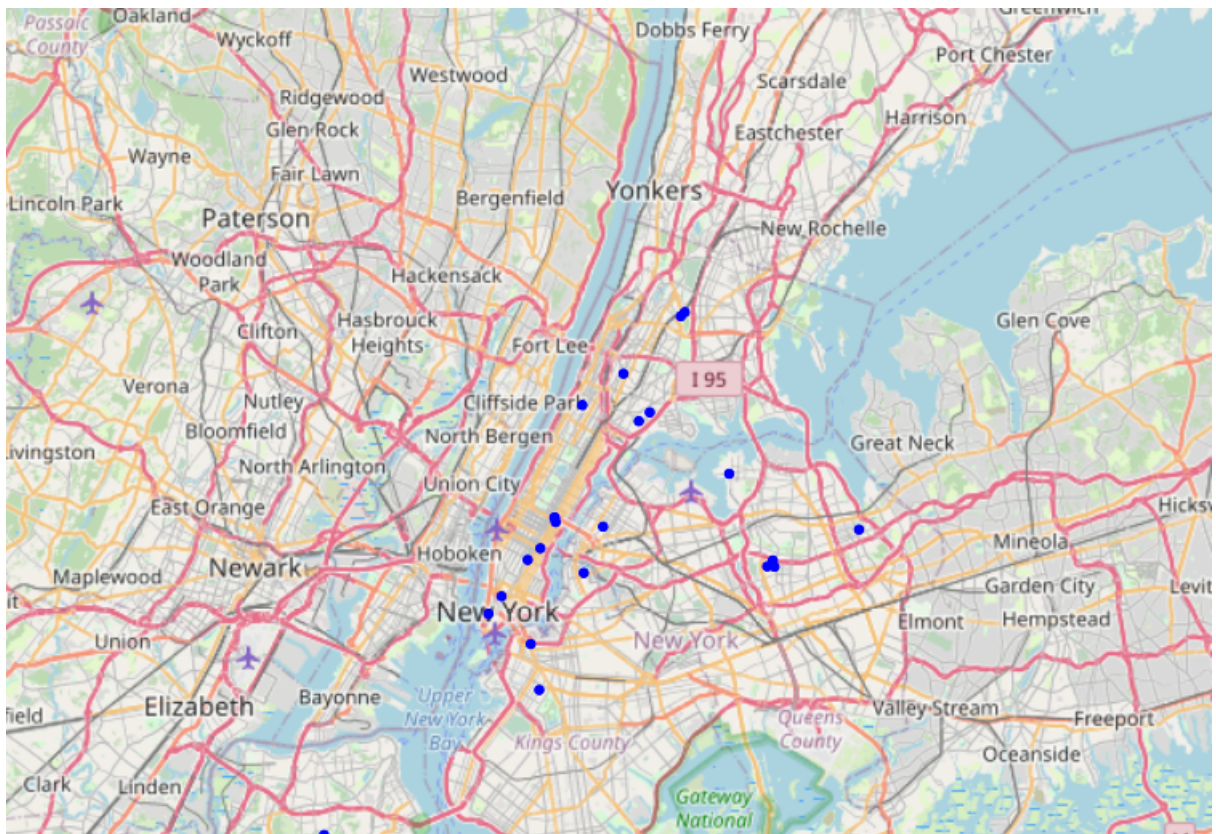


NO2 Yearly Averages and Std

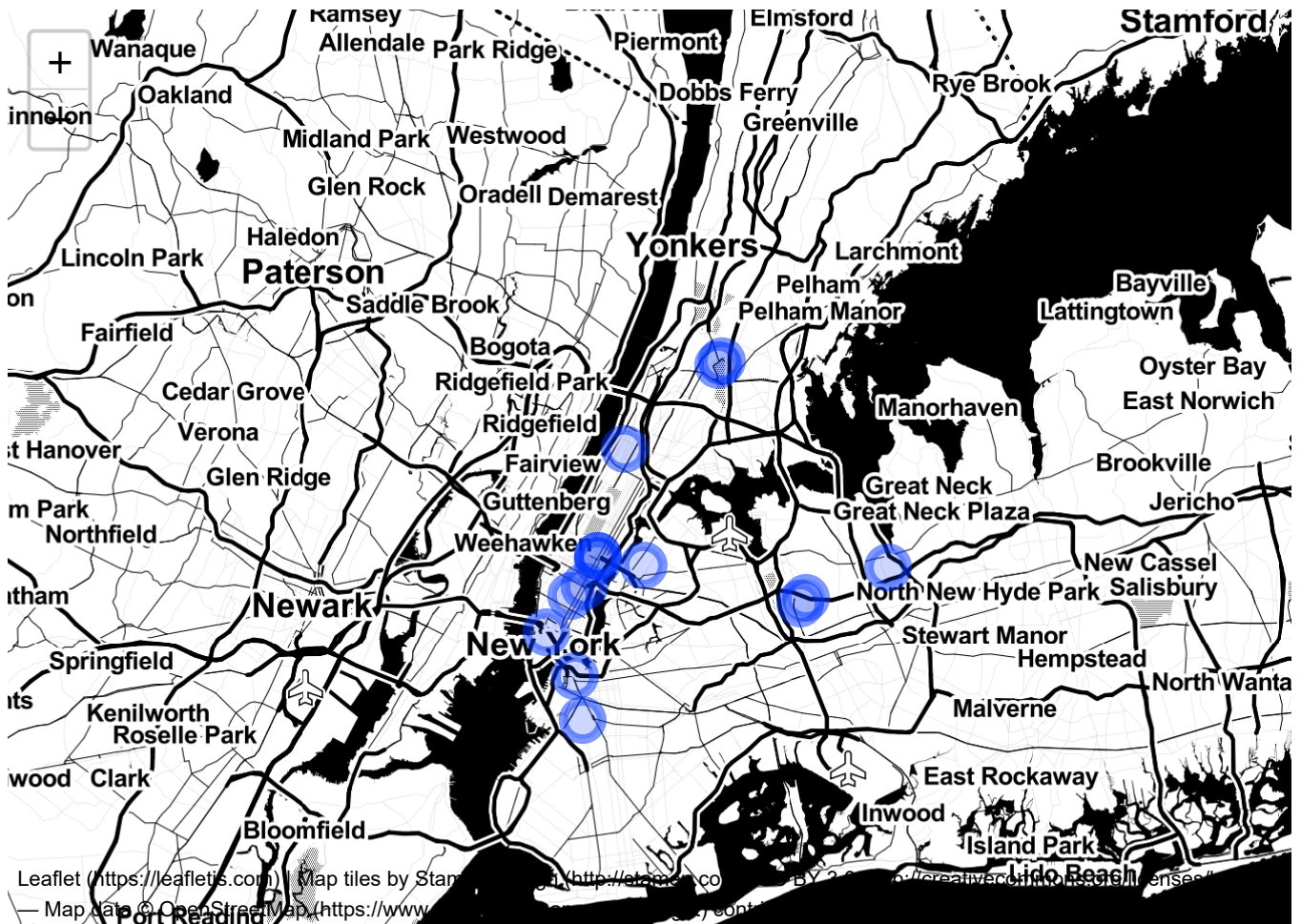


Geospatial Visualization

Air quality monitors across NY



CO monitors across New York



Relative CO pollution among the neighborhoods of NY between 1990 and 2000

In the map below, the darker the red, the higher the CO average concentration in ppm between 1990 and 2000. Since the majority of the neighborhoods do not have a monitor, their CO concentration are estimated through weighted average of all the other monitors. The weights are function of the distances to the other monitors (Read more here).

The *nearest monitor* distance, which is displayed together of the estimation and neighborhood name, can be interpreted as a measurement of approximation occur. It is expected that, the closer a neighborhood is from a monitor, the more accurate is the estimation of the reality. Looking to the Choropleth (https://en.wikipedia.org/wiki/Choropleth_map#:~:text=A%20choropleth%20map%20) below, it is possible to conclude:

- Manhattan and proximity have the highest values of CO on average between 1990 and 2000.
- Neighborhoods near to the coast (most southern) have on average the lowest concentrations of CO.