

MEXICO HOURLY AIR POLLUTION (2010-2021)

รายชื่อสมาชิก

- นาย กรพัล ธนากรณัต์ 650510597
- นาย พีรวัชร โนนตาปิง 650510633
- นาย ศุภวิชญ์ ขันธิกุล 650510688

CONTENT

01

ALL DATA

02

DATA DESCRIPTION

03

CLEAR DATA

04

DATA ANALYSIS

ALL DATA



- Daily Data
- Hourly Data
- Information about weather measurement stations



DATA DESCRIPTION

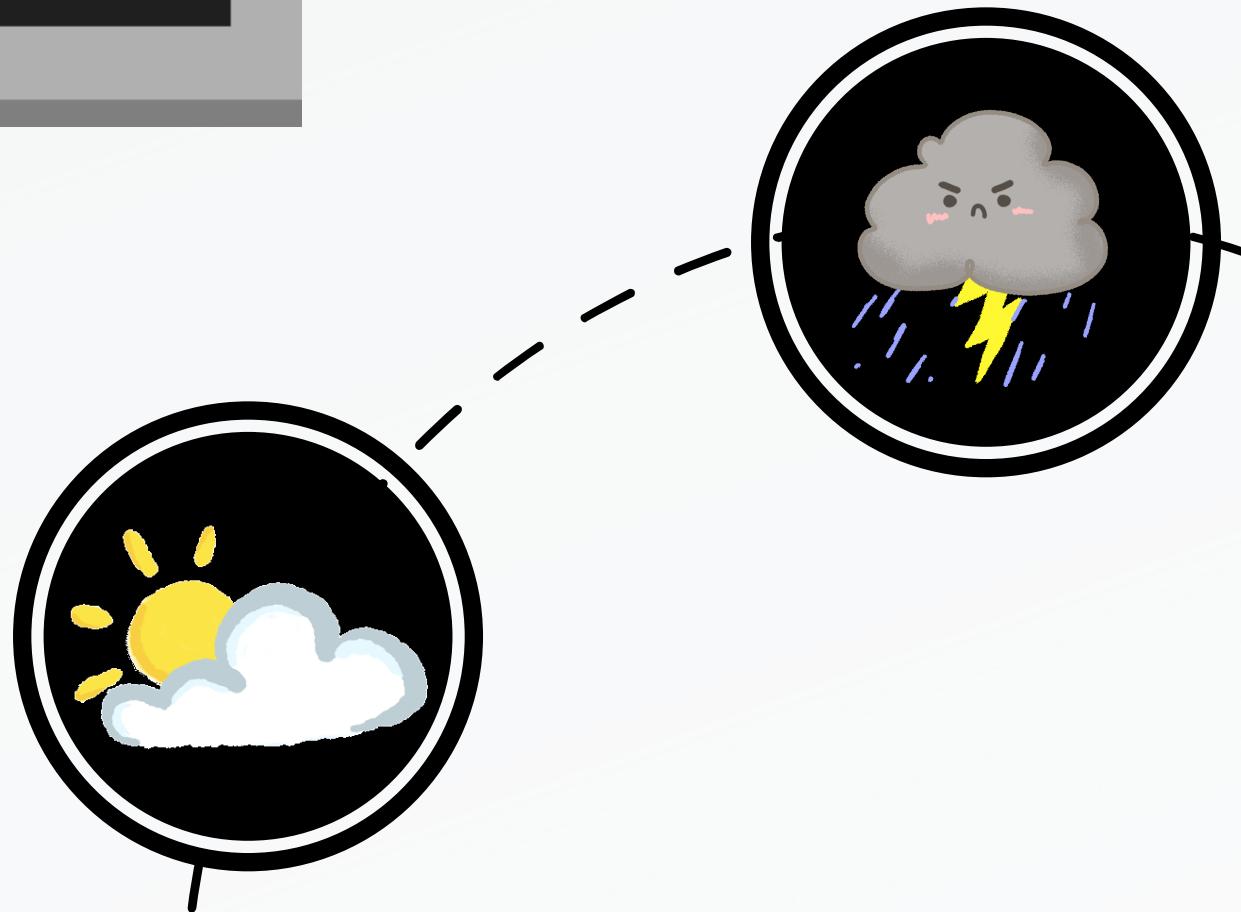
ข้อมูลที่เก็บรายวัน

ข้อมูลที่เก็บรายชั่วโมง

Output :

มี column ดังนี้

```
Index(['datetime', 'station_id', 'PM2.5', 'PM10', 'NOx', 'O3', 'CO', 'HR',
       'NO', 'NO2', 'TMP', 'BEN', 'CH4', 'CN', 'CO2', 'H2S', 'HCNM', 'HCT',
       'HRI', 'IUV', 'PB', 'PP', 'PST', 'RS', 'TMPI', 'UVA', 'XIL'],
      dtype='object')
```

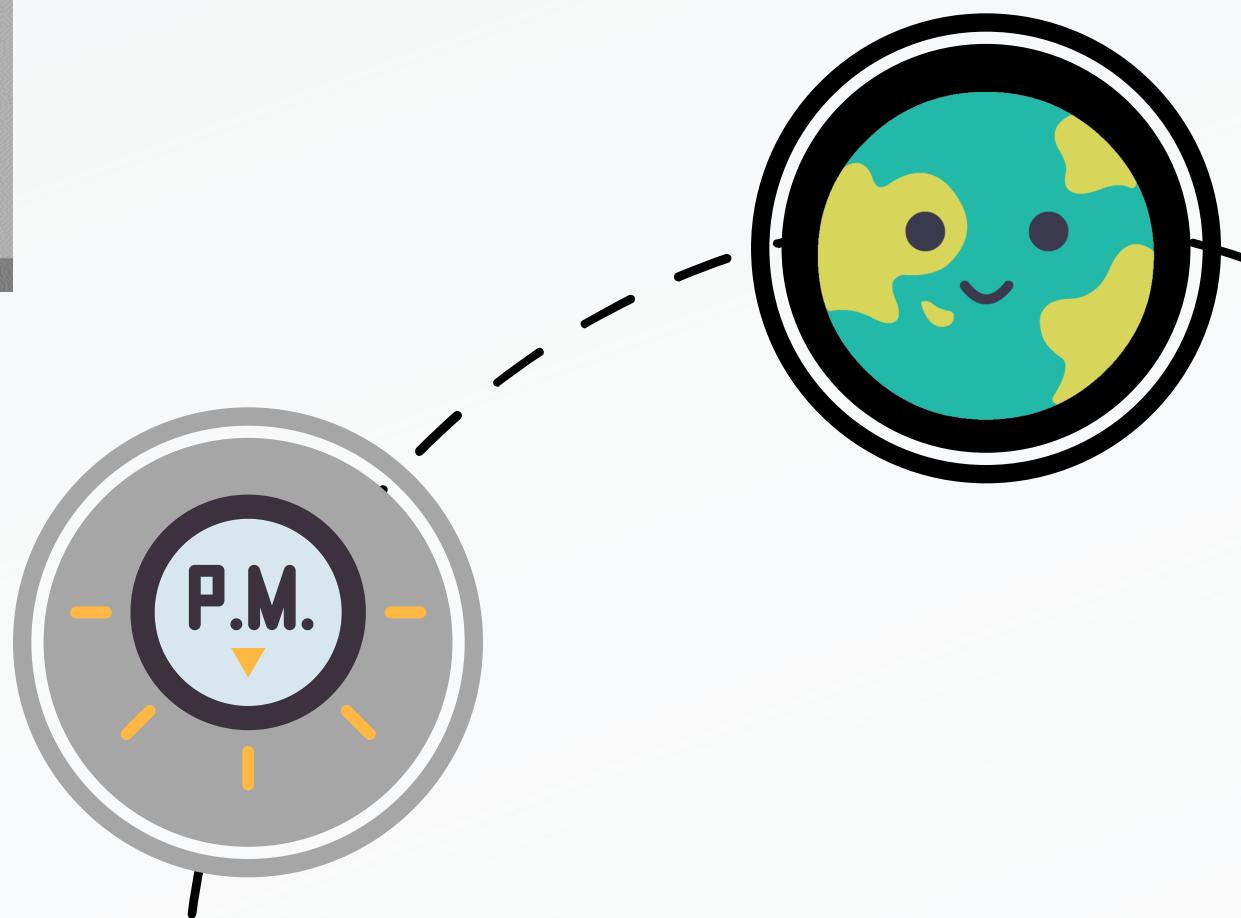


DATA DESCRIPTON

ข้อมูลเกี่ยวกับสถานีเก็บวัดสภาพอากาศ

Output : มี column ดังนี้

```
Index(['station_id', 'station_name', 'station_code', 'network_id',
       'network_name', 'network_code', 'street', 'ext', 'interior', 'colonia',
       'zip', 'state_code', 'municipio_code', 'year_started', 'altitude',
       'address', 'date_validated', 'date_validated2', 'passed_validation',
       'video', 'lat', 'lon', 'date_started', 'timezone', 'street_view',
       'video_interior', 'color'],
      dtype='object')
```



CLEAN DATA

จากข้อมูลที่มีการเก็บบันทึก เราพบว่ามีการเก็บข้อมูลตั้งแต่ปี 2000
แต่เราต้องการข้อมูลตั้งแต่ปี 2010 จึงต้องทำการ clean data

```
dt.date.min()  
>> datetime.date(2000, 1, 1)
```

จากการ process code จะได้ข้อมูลตั้งแต่ปี 2000
ซึ่งเป็นข้อมูลดิบที่น้อยที่สุดจากวันที่

ค่าสัดส่วนของจำนวนช่องข้อมูลที่ว่างต่อช่องทั้งหมด

จากขนาดข้อมูลเราเลือกใช้ไฟล์รายช่วงไม่เนื่องจากมีข้อมูลที่มากกว่า

Hourly data



```
datetime      0.000000
station_id    0.000000
PM2.5        0.319223
PM10         0.220759
NOx          0.488096
O3           0.174247
CO            0.236877
HR            0.430760
NO            0.473294
NO2           0.270147
TMP           0.417410
BEN           0.996921
CH4           0.999292
CN            0.997875
CO2           0.999778
H2S           0.999215
HCNM          0.999264
HCT           1.000000
HRI           0.869417
IUV           0.981712
PB            0.647174
PP            0.619081
PST           1.000000
RS            0.722184
TMPI          0.768464
UVA           0.936919
XIL           0.996921
dtype: float64
```

Daily data



```
datetime      0.000000
station_id    0.000000
PM2.5        0.301798
PM10         0.220694
NOx          0.512587
O3           0.166824
CO            0.242232
HR            0.467663
NO            0.489818
NO2           0.269033
TMP           0.451557
BEN           0.996654
CH4           0.998834
CN            0.997431
CO2           0.999758
H2S           0.998791
HCNM          0.998795
HCT           1.000000
HRI           0.955935
IUV           0.978687
PB            0.603881
PP            0.674661
PST           1.000000
RS            0.673965
TMPI          0.851528
UVA           0.930524
XIL           0.996654
dtype: float64
```

CLEAN DATA

ค่าลัดล่วงของจำนวนช่องข้อมูลที่ว่างต่อช่องทั้งหมด

คัดข้อมูลที่มีค่าเกิน ครึ่งหนึ่ง (<0.5)

Hourly data

```
datetime      0.000000
station_id    0.000000
PM2.5        0.319223
PM10         0.220759
NOx          0.488096
O3            0.174247
CO            0.236877
HR            0.430760
NO            0.473294
NO2           0.270147
TMP           0.417410
BEN           0.996921
CH4           0.999292
CN            0.997875
CO2           0.999778
H2S           0.999215
HCNM          0.999264
HCT           1.000000
HRI           0.869417
IUV           0.981712
PB             0.647174
PP             0.619081
PST           1.000000
RS             0.722184
TMPI          0.768464
UVA           0.936919
XIL           0.996921
dtype: float64
```



```
datetime      0.000000
station_id    0.000000
PM2.5        0.319223
PM10         0.220759
NOx          0.488096
O3            0.174247
CO            0.236877
HR            0.430760
NO            0.473294
NO2           0.270147
TMP           0.417410
dtype: float64
```

CLEAN DATA

พิจารณาข้อมูลโดยใช้ไฟล์รายช่วงไม่งedly เป็นรายวัน
จากนั้นตัดค่า outlier โดยใช้ percentile ที่ 98.5 และ 5
แล้วเติมค่าในช่องที่ไม่มีข้อมูลด้วยค่าเฉลี่ยนั้นๆ

	station_id	PM2.5	PM10	NOx	O3	CO	HR	NO	NO2
count	5.848182e+06								
mean	1.643294e+02	2.441384e+01	5.270074e+01	3.431426e-02	2.547988e-02	9.878910e-01	6.036240e+01	1.781668e-02	1.777698e-02
std	8.867297e+01	2.157374e+01	3.676655e+01	2.557446e-02	1.711418e-02	6.587089e-01	1.685173e+01	1.962701e-02	1.073153e-02
min	3.200000e+01	3.012900e+00	9.400056e+00	3.200100e-03	2.000081e-03	1.560028e-01	1.663104e+01	5.000100e-04	2.100028e-03
25%	1.020000e+02	1.450119e+01	3.105655e+01	1.864065e-02	1.200000e-02	5.350000e-01	5.154000e+01	6.000000e-03	1.000000e-02
50%	1.370000e+02	2.110023e+01	4.790000e+01	3.036875e-02	2.431242e-02	8.200000e-01	6.036240e+01	1.339321e-02	1.620000e-02
75%	2.450000e+02	2.723187e+01	5.900000e+01	4.083036e-02	3.200000e-02	1.202126e+00	7.000000e+01	2.163440e-02	2.300000e-02
max	4.260000e+02	3.018800e+02	3.270700e+02	2.069858e-01	9.399300e-02	4.166400e+00	9.999999e+01	1.729900e-01	6.339660e-02

datetime	0
station_id	0
PM2.5	251524
PM10	577548
NOx	595181
O3	294199
CO	394821
HR	602107
NO	570415
NO2	399948
TMP	558123
dtype:	int64

DATA ANALYSIS



DATA CORRELATION



DATA COMPARISON

compare between example station

DATA CORRELATION

What do the guidelines recommend?

The WHO Air quality guidelines recommend levels and interim targets for common air pollutants: PM, O₃, NO₂, and SO₂.

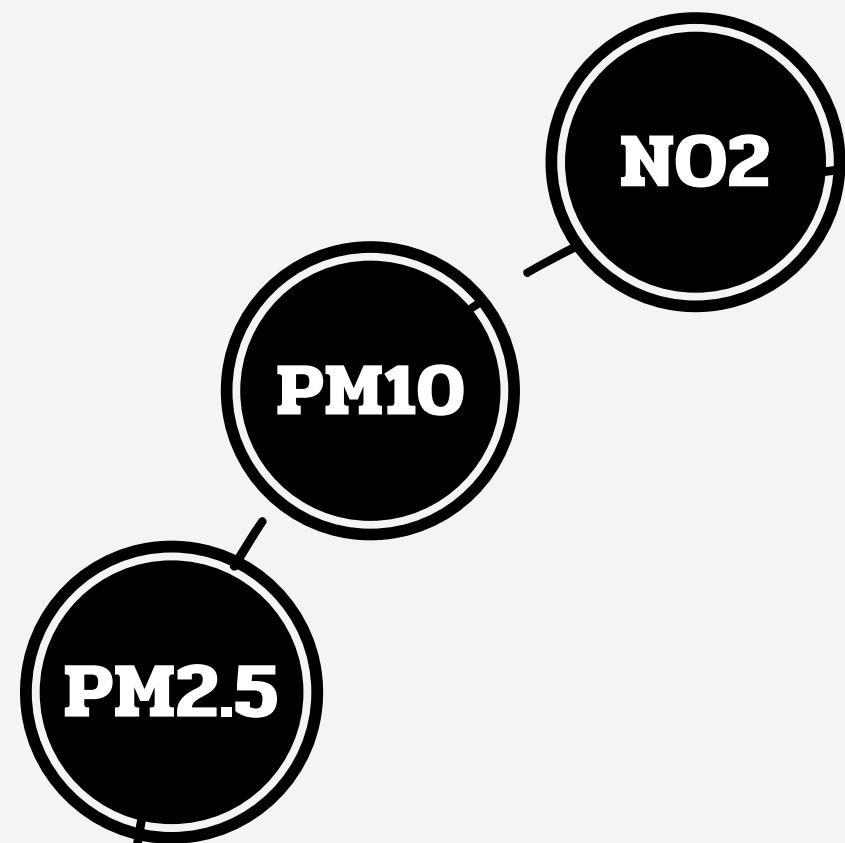
Recommended 2021 AQG levels compared to 2005 air quality guidelines

Pollutant	Averaging Time	2005 AQGs	2021 AQGs
PM _{2.5} , µg/m ³	Annual	10	5
	24-hour ^a	25	15
PM ₁₀ , µg/m ³	Annual	20	15
	24-hour ^a	50	45
O ₃ , µg/m ³	Peak season ^b	-	60
	8-hour ^a	100	100
NO ₂ , µg/m ³	Annual	40	10
	24-hour ^a	-	25
SO ₂ , µg/m ³	24-hour ^a	20	40
CO, mg/m ³	24-hour ^a	-	4

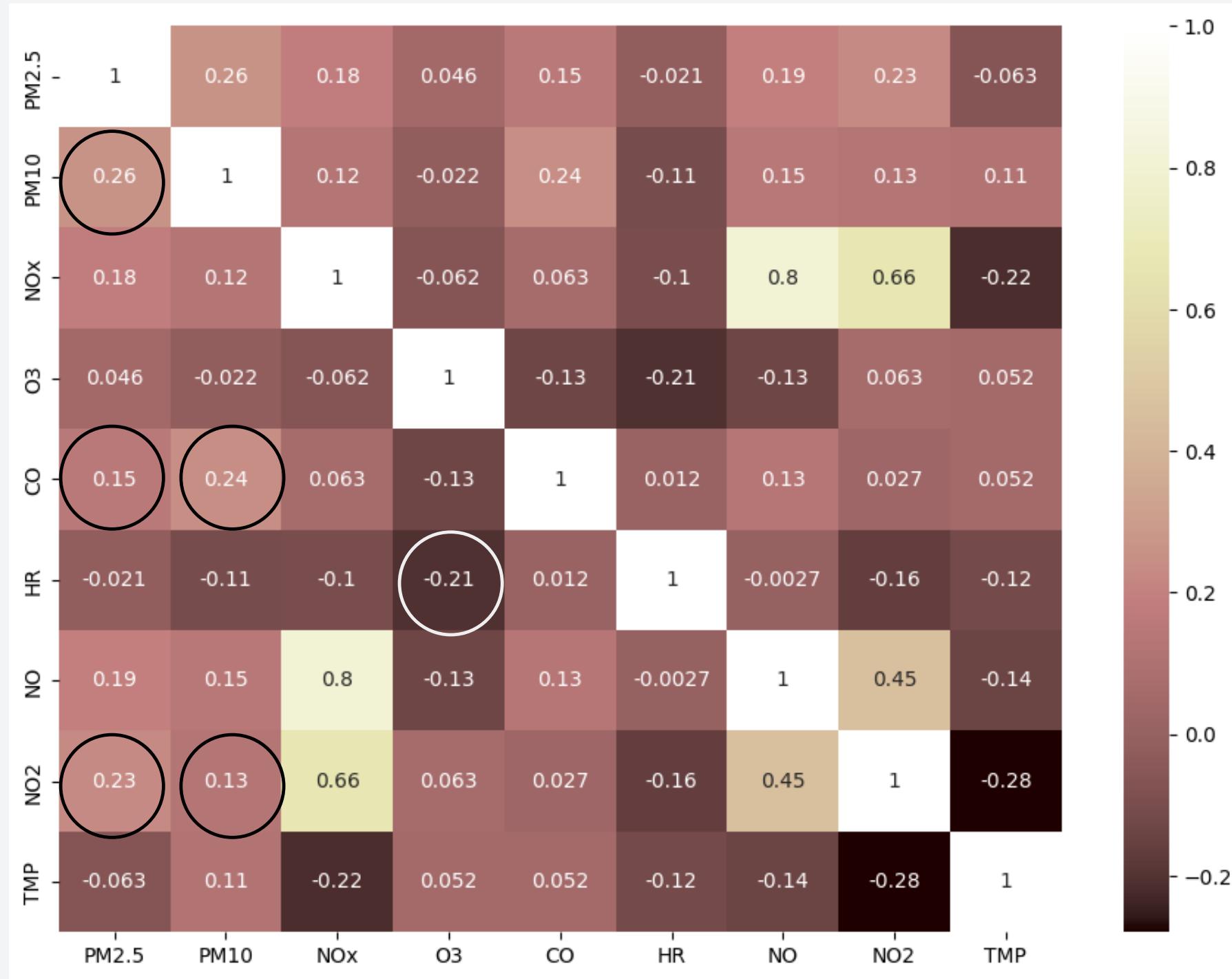
NO₂ ຈະຖືກແປລັງເປັນຫມົຍ ppm

$$\frac{\frac{10}{1000}}{0.0409 \times (14 + (16 \times 2))} \approx 0.005315$$

	PM2.5	PM10	NO2
2010	1.0	1.0	1.000000
2011	1.0	1.0	1.000000
2012	1.0	1.0	1.000000
2013	1.0	1.0	1.000000
2014	1.0	1.0	1.000000
2015	1.0	1.0	0.984127
2016	1.0	1.0	0.985507
2017	1.0	1.0	0.973333
2018	1.0	1.0	0.961538
2019	1.0	1.0	0.987179
2020	1.0	1.0	0.986842
2021	1.0	1.0	0.985915



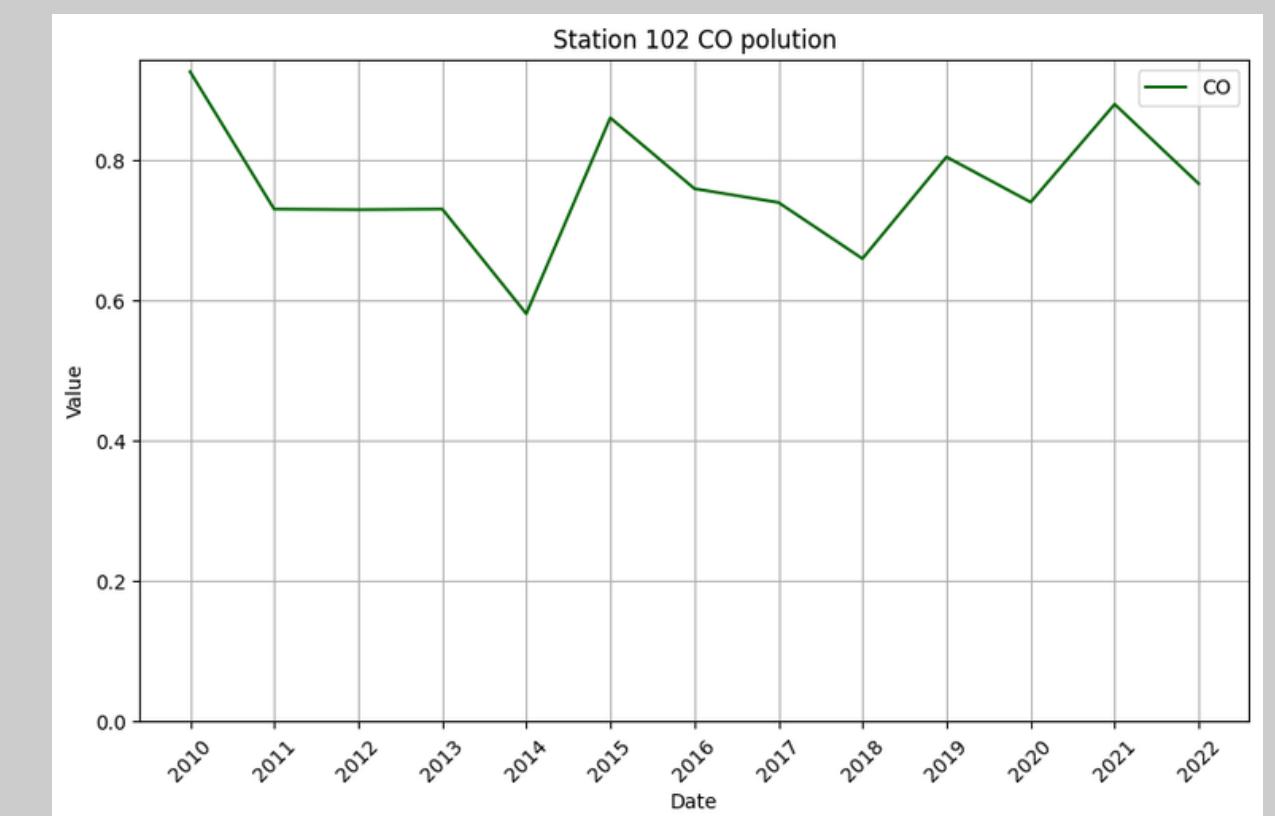
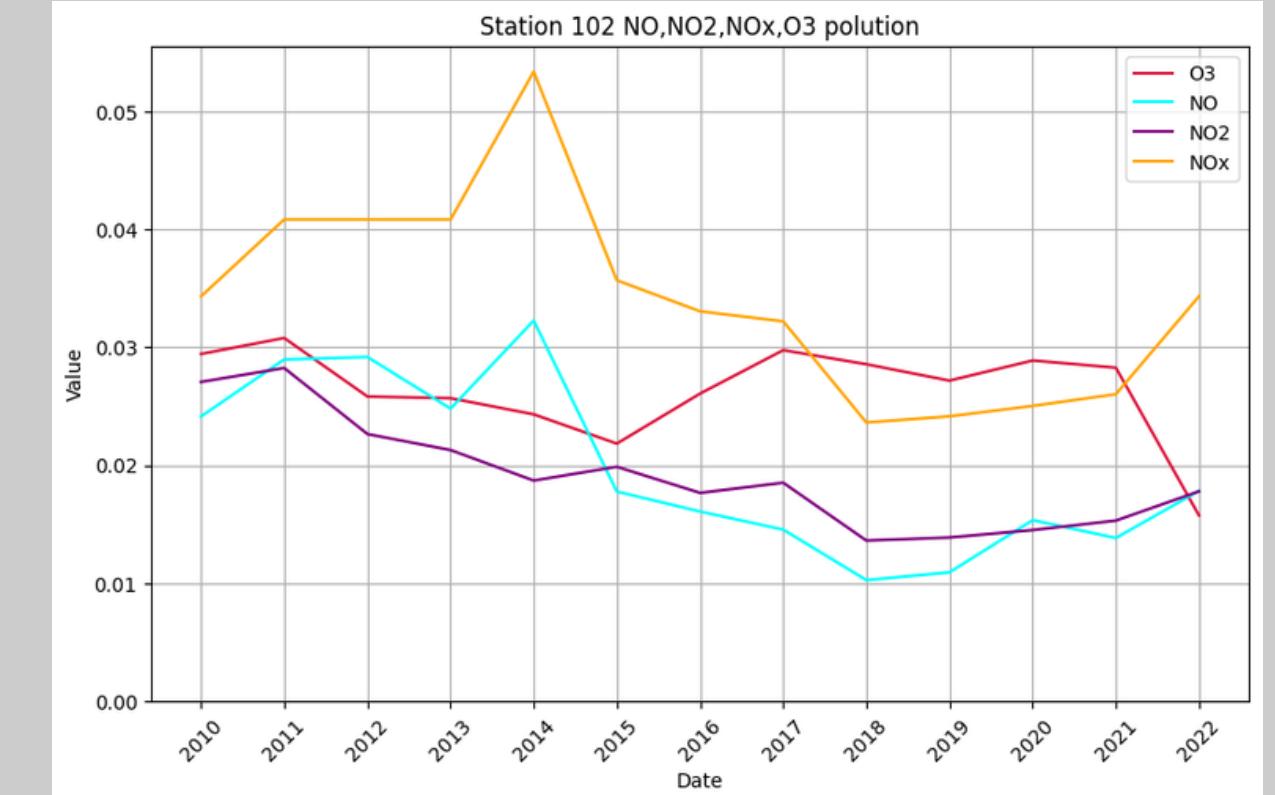
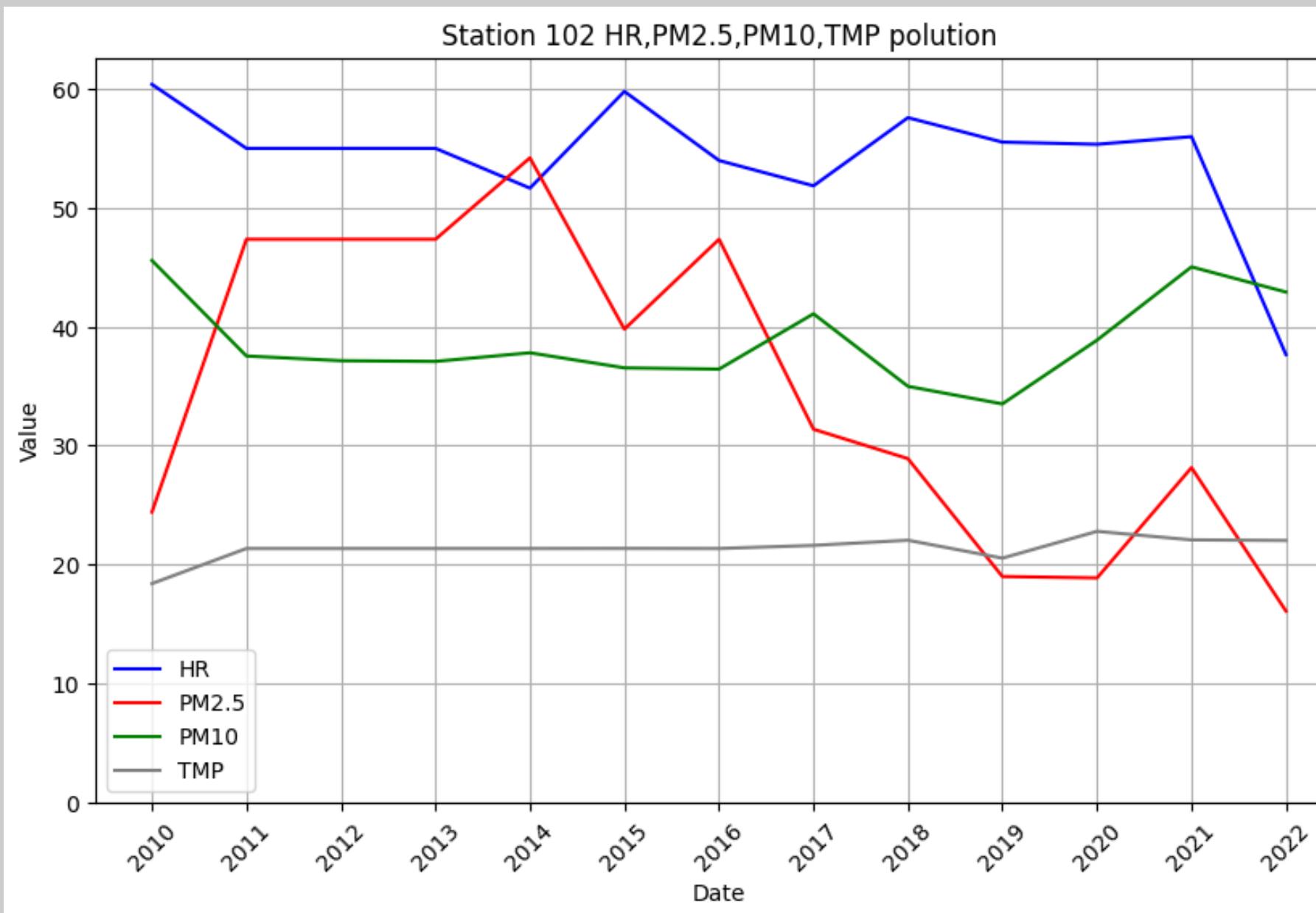
DATA CORRELATION



สังเกตว่า column ของ PM ต่างๆ และ NO2

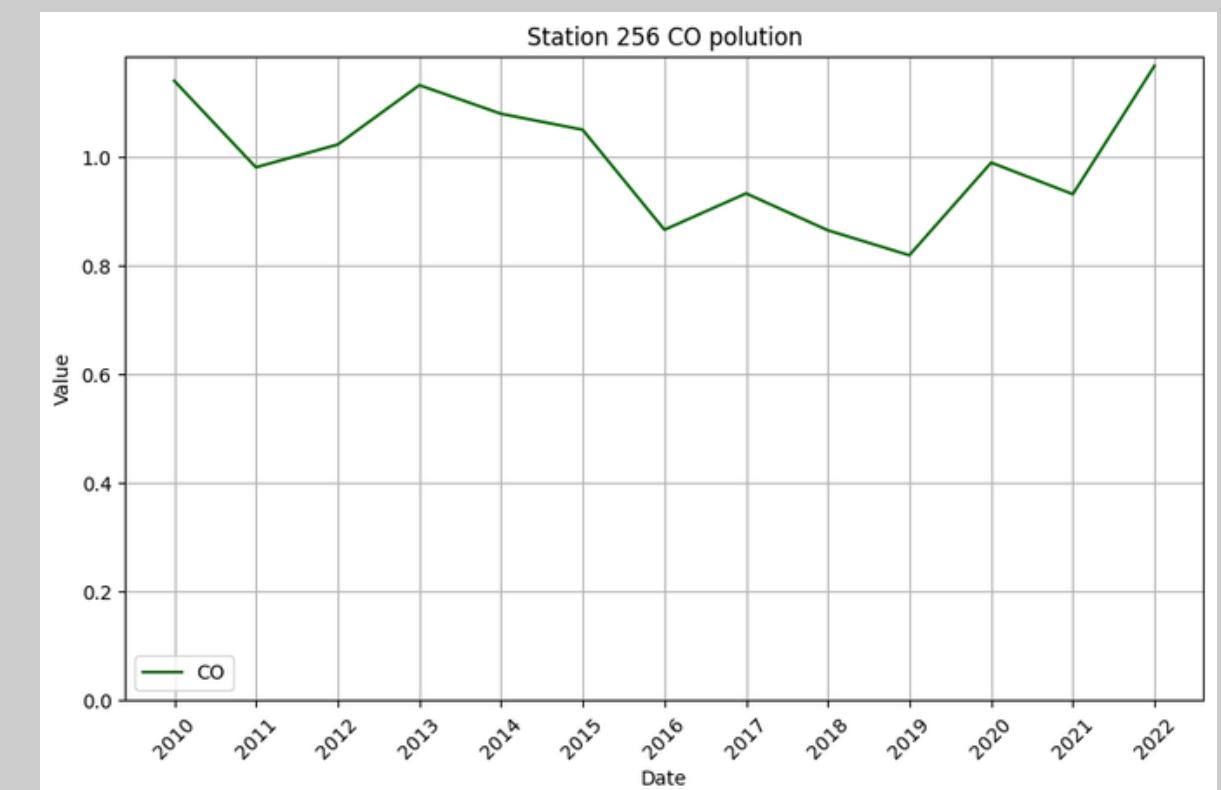
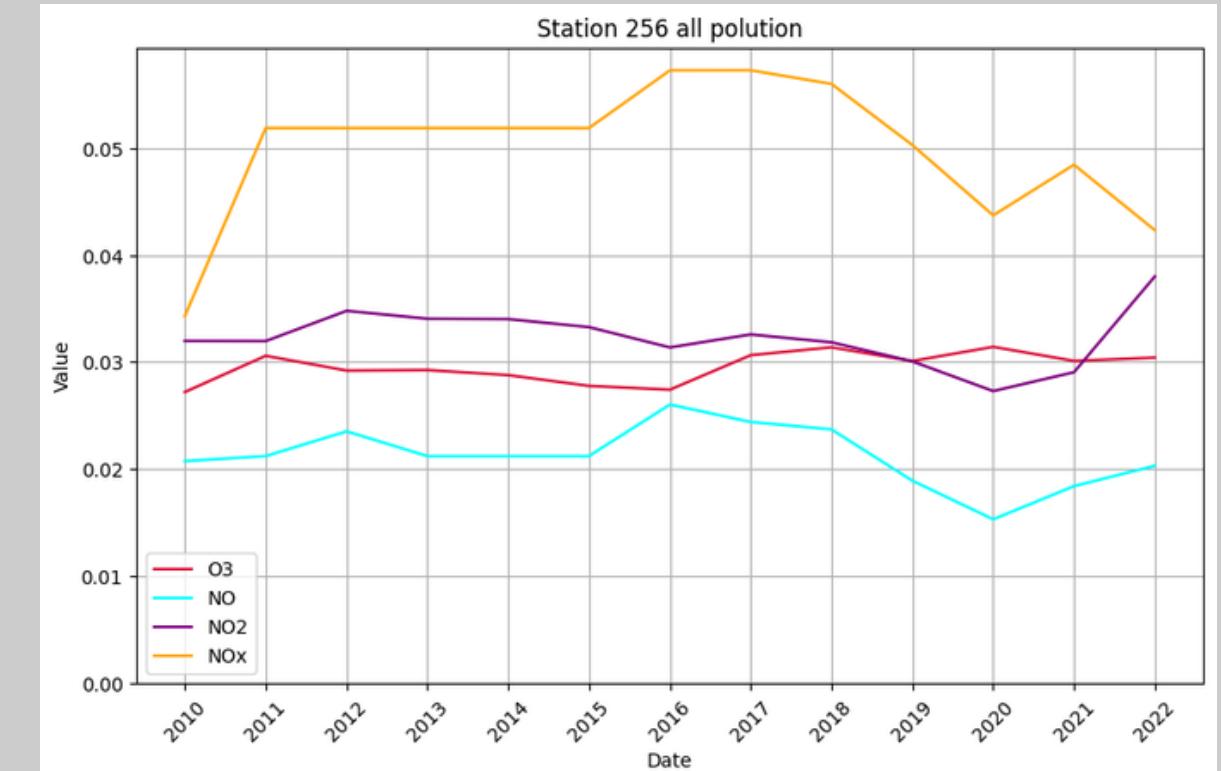
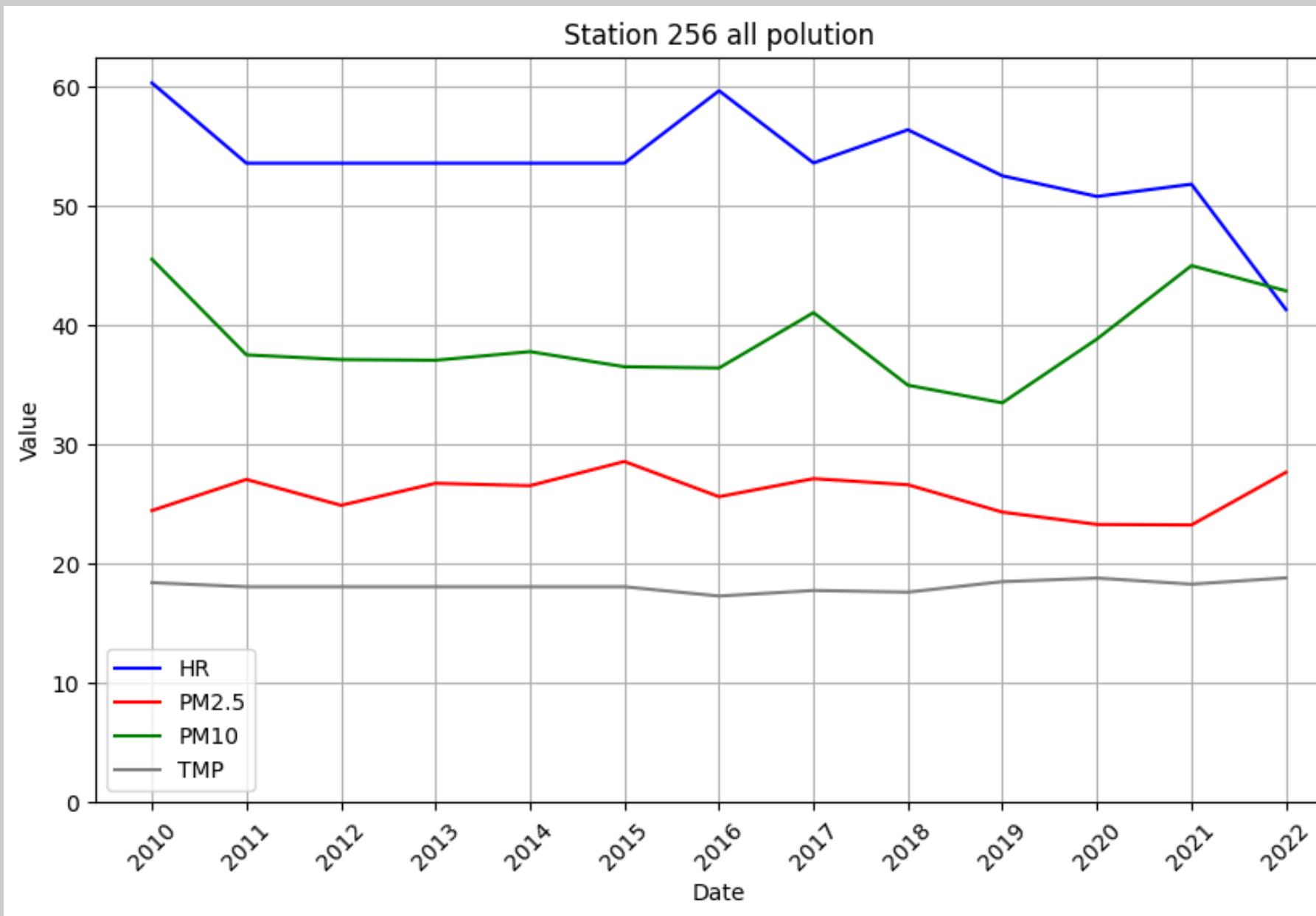
DATA COMPARISON

Station 102



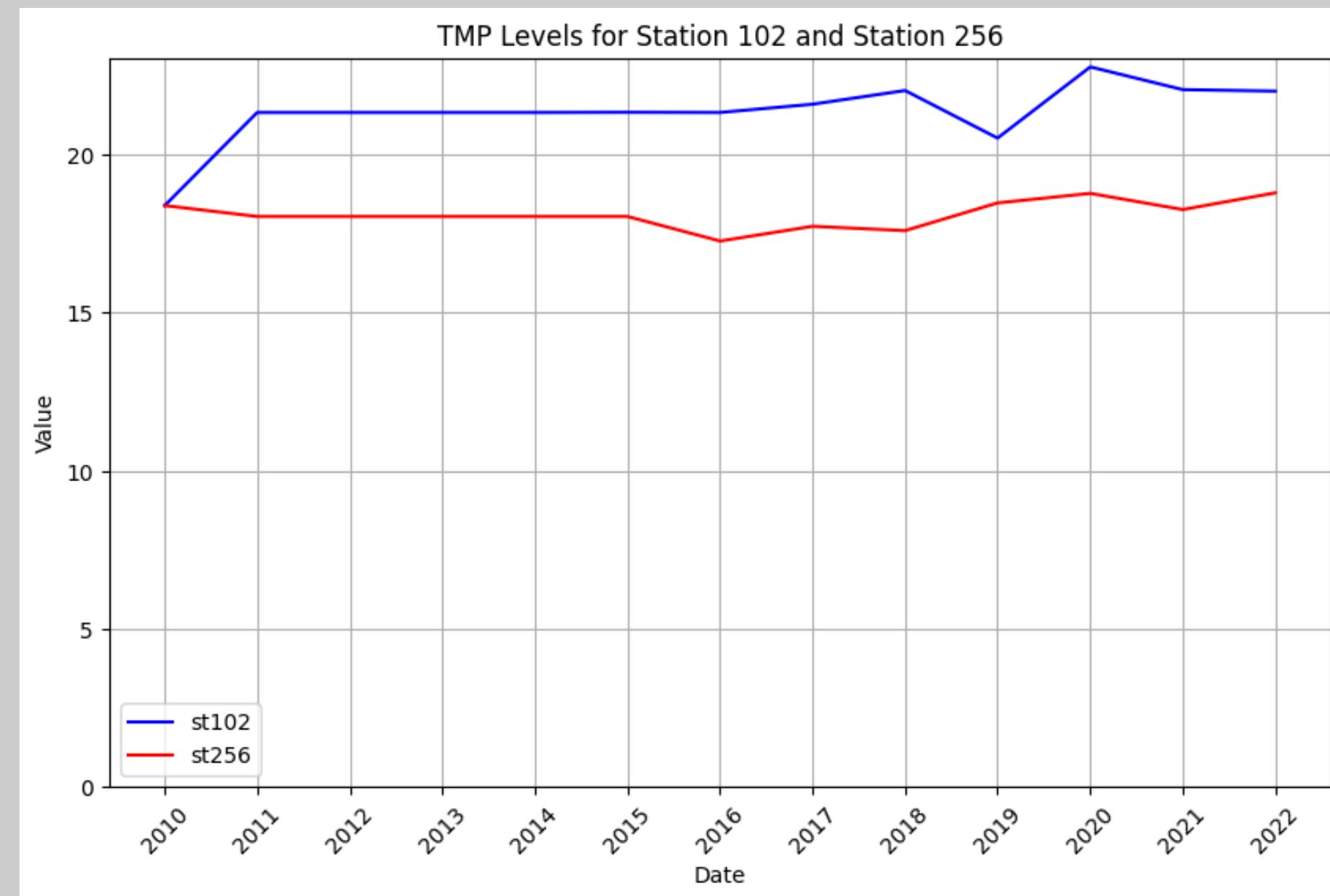
DATA COMPARISON

Station 256



DATA COMPARISON

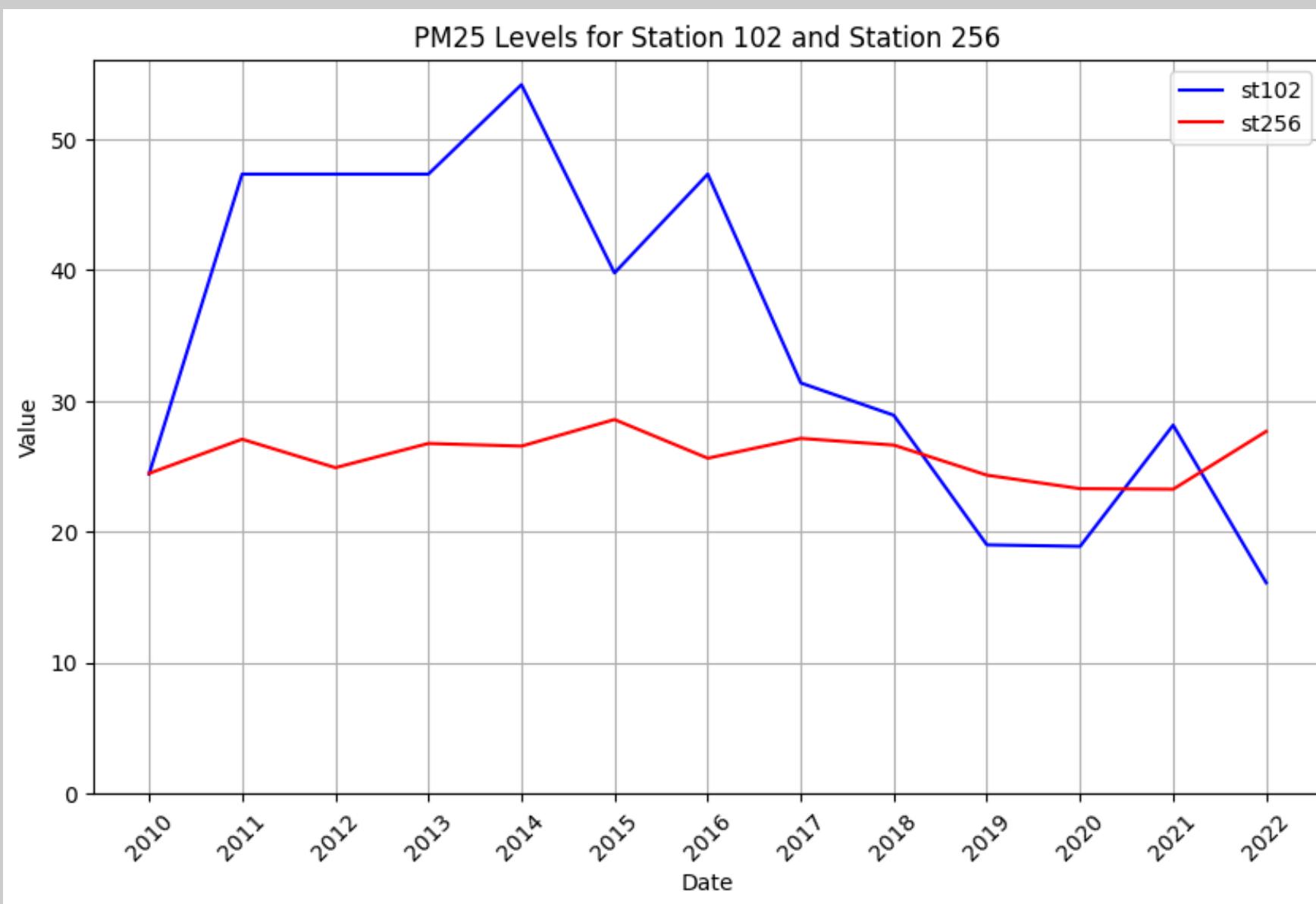
*TMP (Temperature)
Station 102 compare to Station 256*



DATA COMPARISON

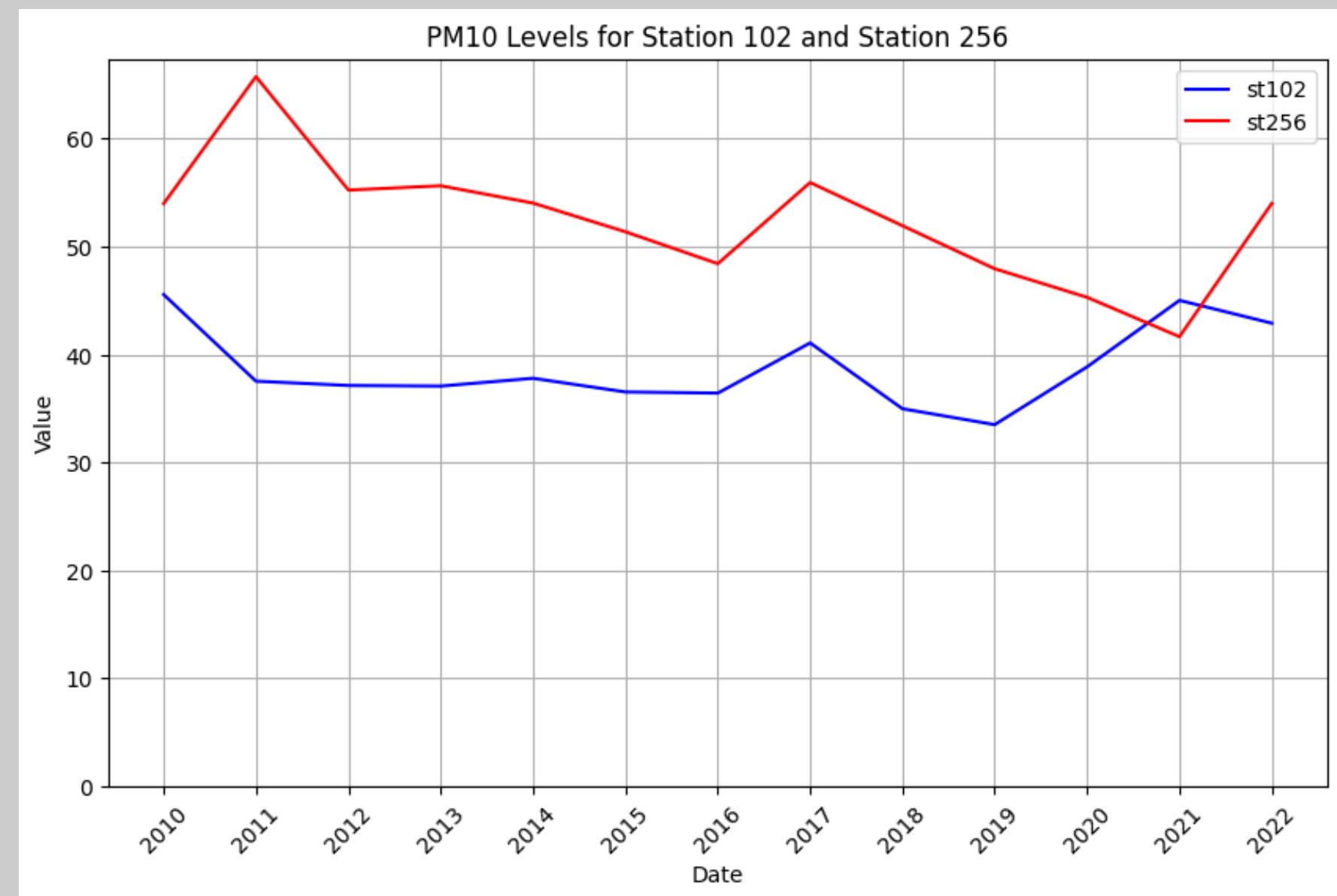
PM 2.5

Station 102 compare to Station 256



PM 10

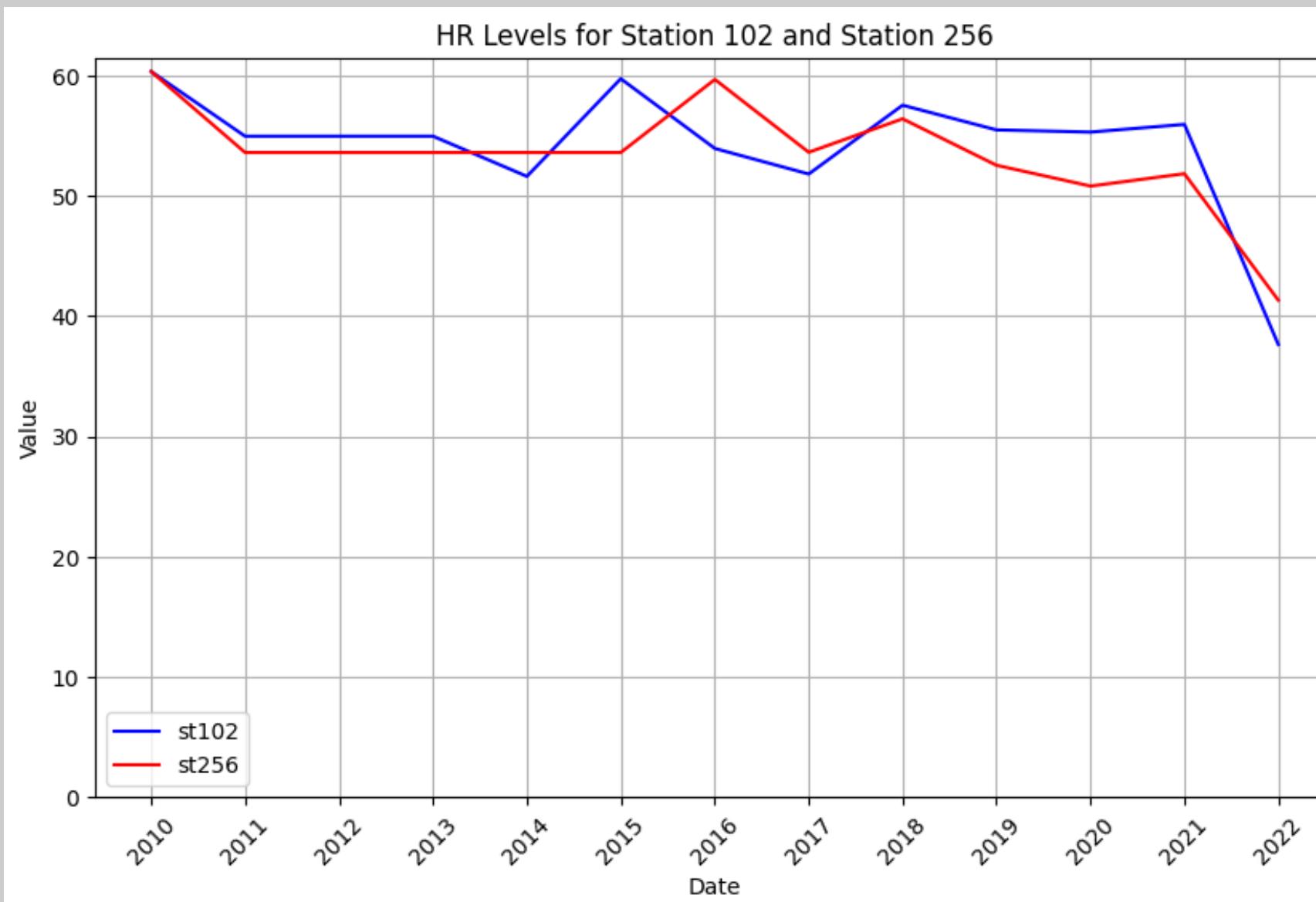
Station 102 compare to Station 256



DATA COMPARISON

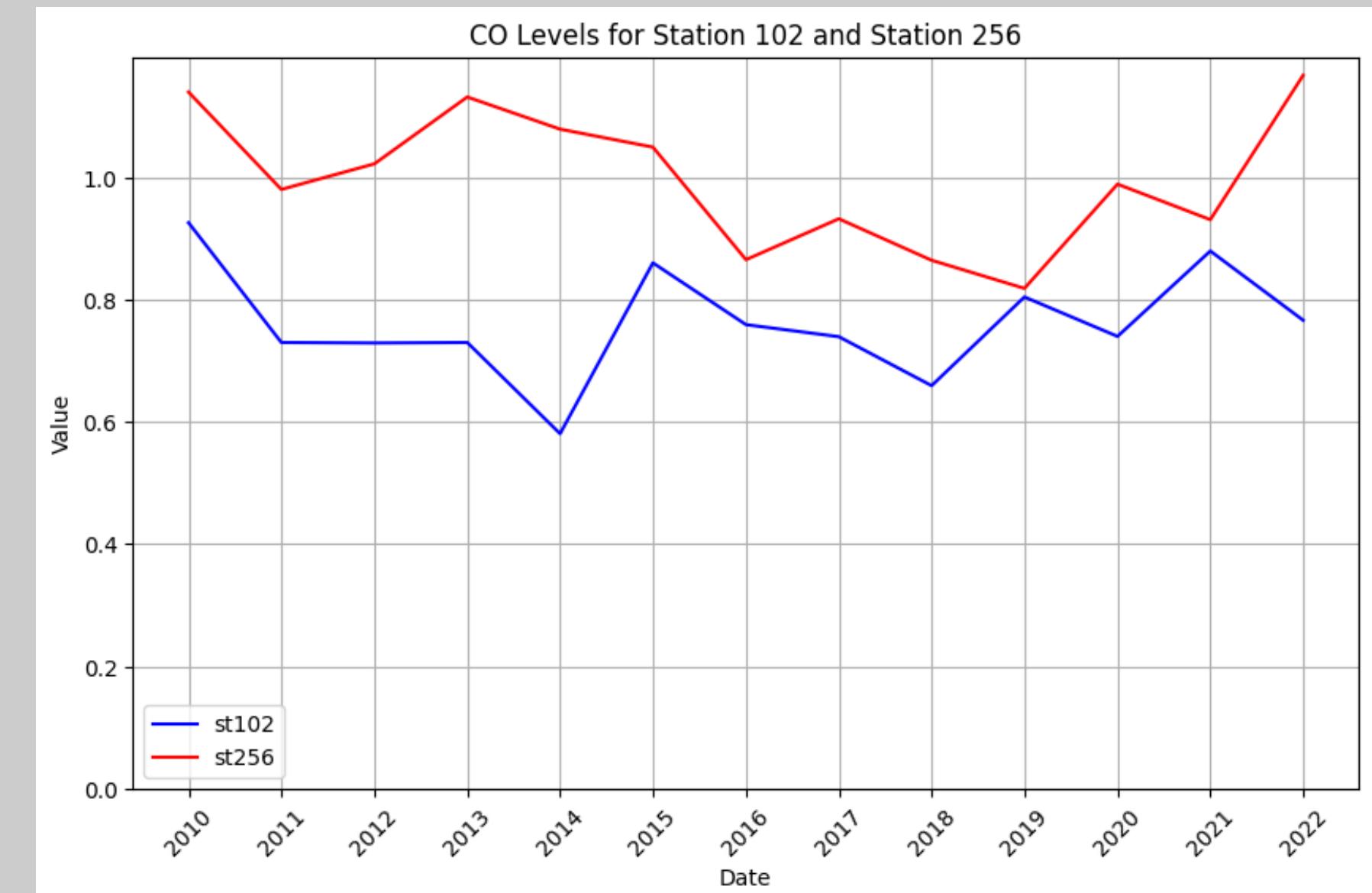
HR (relative humidity)

Station 102 compare to Station 256



CO (Carbon monoxide)

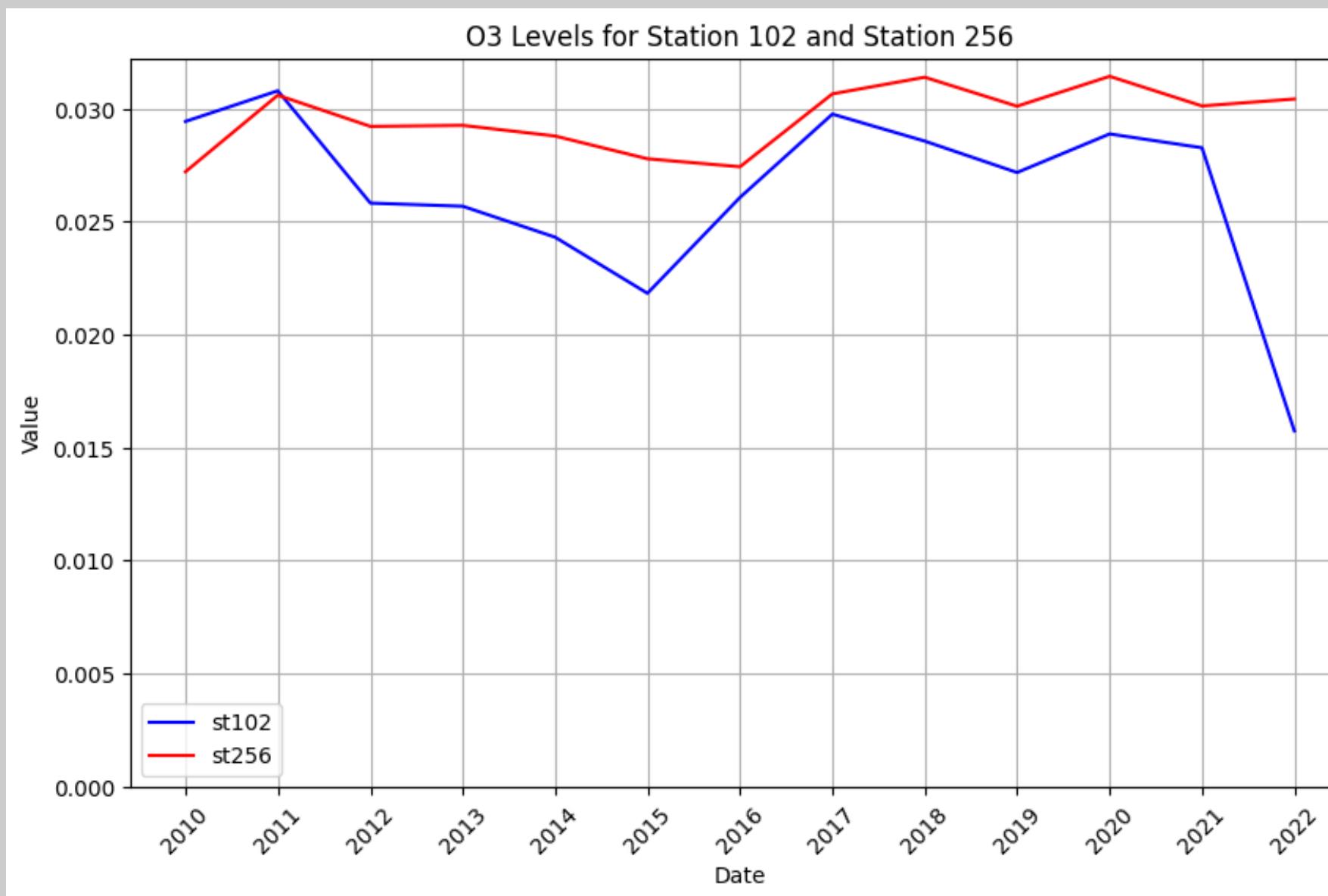
Station 102 compare to Station 256



DATA COMPARISON

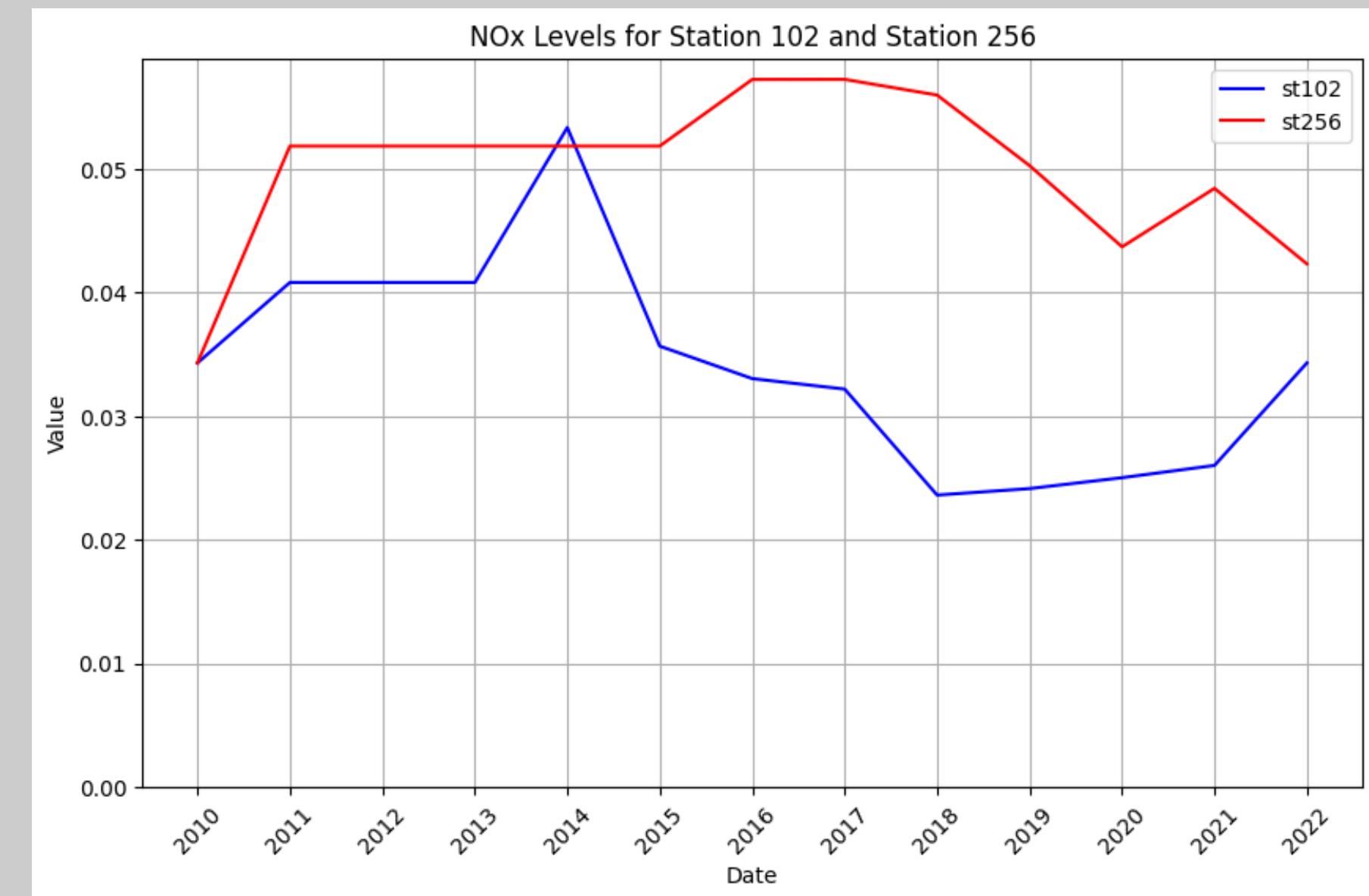
O3 (Ozone)

Station 102 compare to Station 256



NOx (Nitrogen oxide)

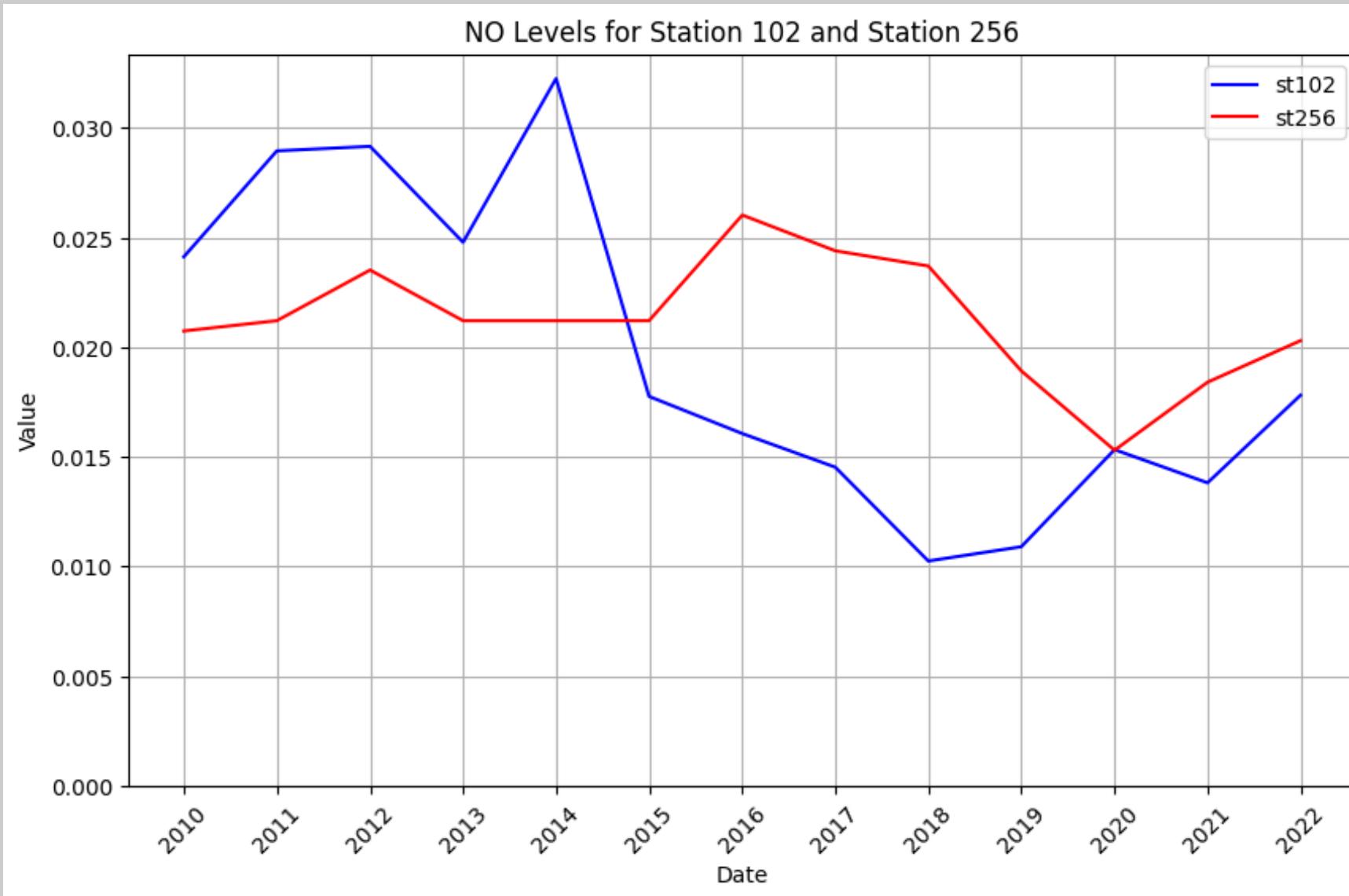
Station 102 compare to Station 256



DATA COMPARISON

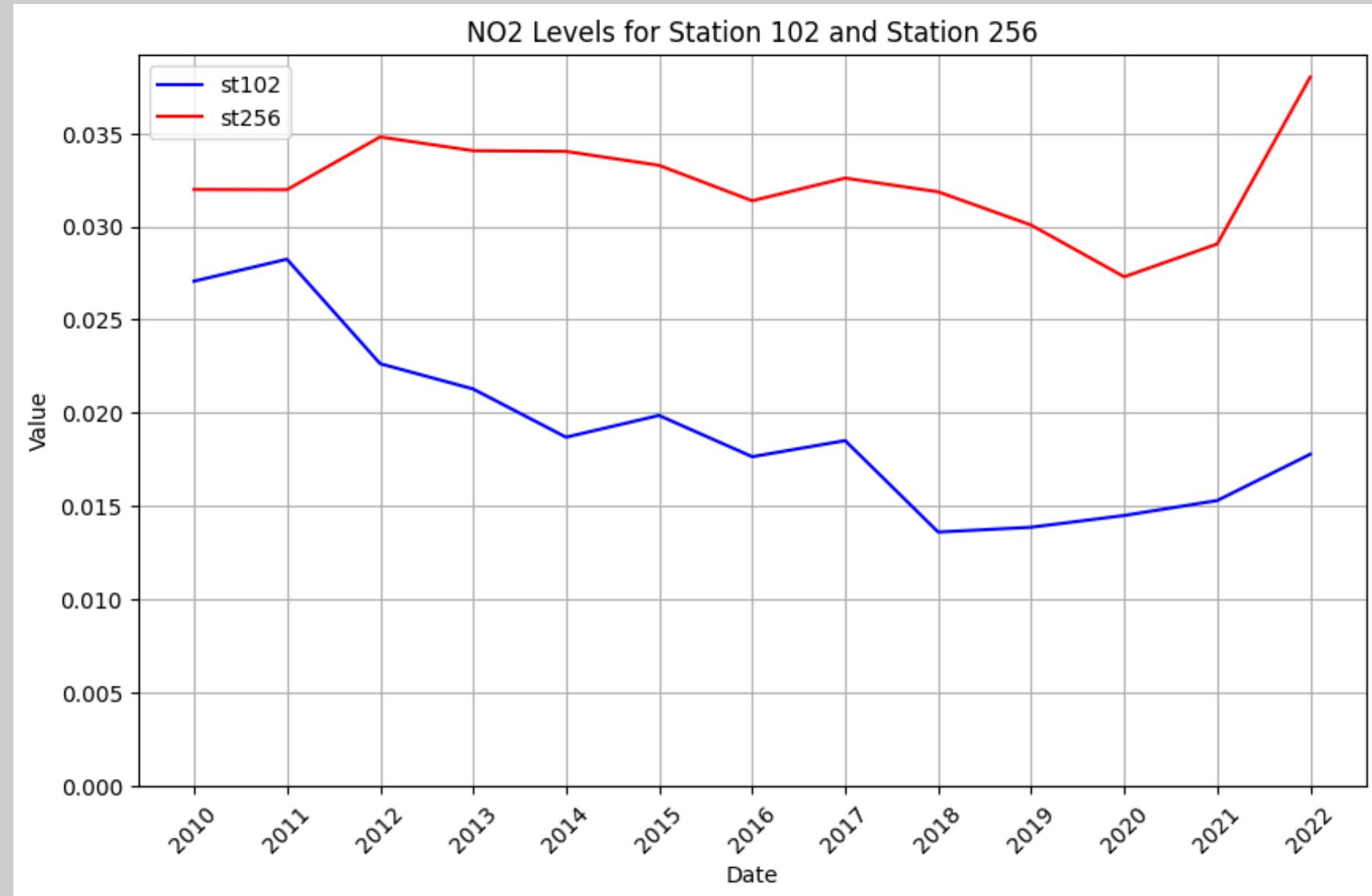
NO (*Nitric oxide*)

Station 102 compare to Station 256



NO₂ (*Nitrogen dioxide*)

Station 102 compare to Station 256



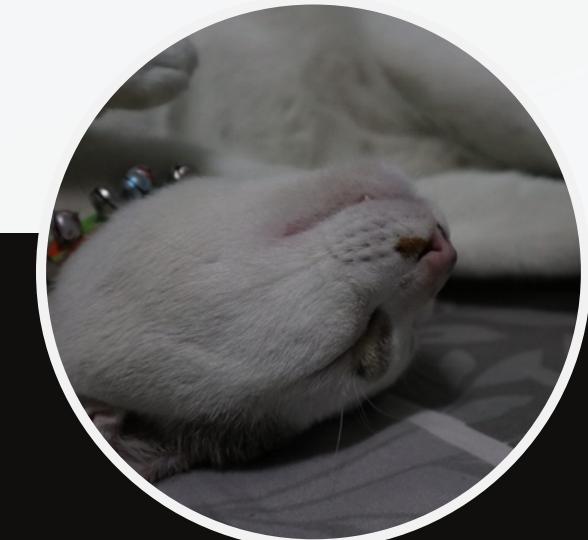
TEAM MEMBER



Koraphat
Thanakritnaphan
Tar



Peeraphat
Notaping
New



Supawit
Kantikul
Kev

**THANK'S FOR
WATCHING**

