In [ ]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings('ignore')

%matplotlib inline
```

In [2]:
```
df = pd.read_csv("AirQualityUCI.csv")
df.sample(10)
```

Out[2]:

	Date	Time	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)
3583	7/8/2004	1:00:00	1.2	925	-200	6.4	831	57	789	57	1662	975
4710	23/9/2004	0:00:00	1.6	1037	-200	8.0	898	152	793	87	1539	836
932	18/4/2004	14:00:00	-200.0	944	-200	3.4	677	-200	1140	-200	1436	501
6011	16/11/2004	5:00:00	0.5	737	-200	0.7	468	50	1481	40	869	395
656	7/4/2004	2:00:00	0.4	830	30	1.4	538	21	1475	30	1214	392
92	14/3/2004	14:00:00	1.8	1207	84	7.5	879	103	1104	102	1490	872
7086	31/12/2004	0:00:00	-200.0	987	-200	6.2	823	-200	827	-200	1027	893
9233	30/3/2005	11:00:00	1.8	1109	-200	7.9	896	287	629	137	1308	1027
7017	28/12/2004	3:00:00	-200.0	961	-200	3.0	654	-200	925	-200	1084	755
2129	7/6/2004	11:00:00	2.5	998	-200	14.1	1119	207	832	120	1780	1057

In [3]:
```
df.info()
```

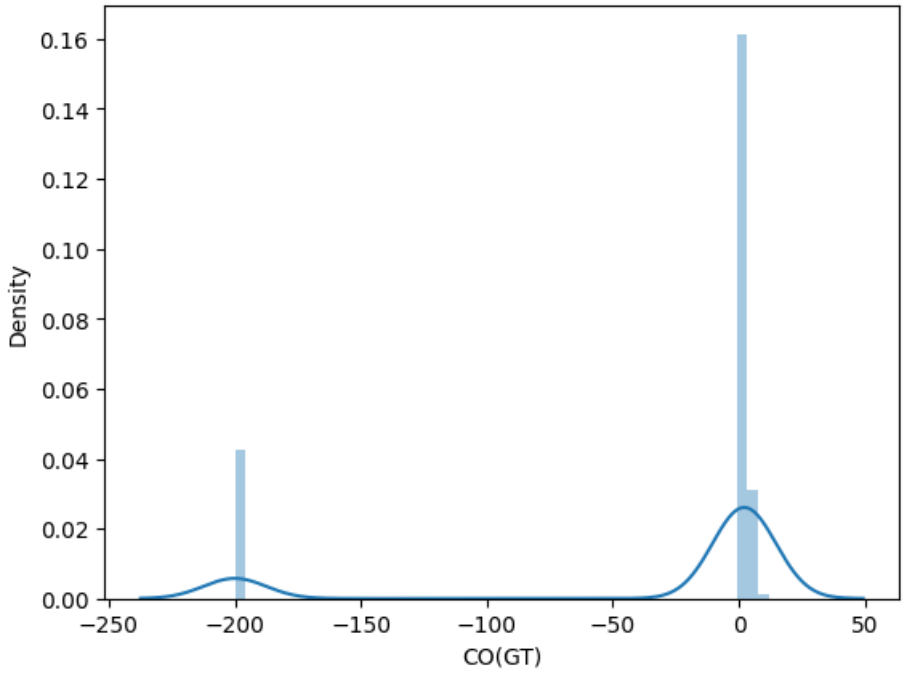
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9357 entries, 0 to 9356
Data columns (total 15 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Date           9357 non-null   object
 1   Time           9357 non-null   object
 2   CO(GT)         9357 non-null   float64
 3   PT08.S1(CO)    9357 non-null   int64
 4   NMHC(GT)       9357 non-null   int64
 5   C6H6(GT)       9357 non-null   float64
 6   PT08.S2(NMHC)  9357 non-null   int64
 7   NOx(GT)        9357 non-null   int64
 8   PT08.S3(NOx)   9357 non-null   int64
 9   NO2(GT)        9357 non-null   int64
 10  PT08.S4(NO2)   9357 non-null   int64
 11  PT08.S5(O3)    9357 non-null   int64
 12  T              9357 non-null   float64
 13  RH             9357 non-null   float64
 14  AH             9357 non-null   float64
dtypes: float64(5), int64(8), object(2)
memory usage: 1.1+ MB
```

In [4]:
```
df.isnull().sum()
```

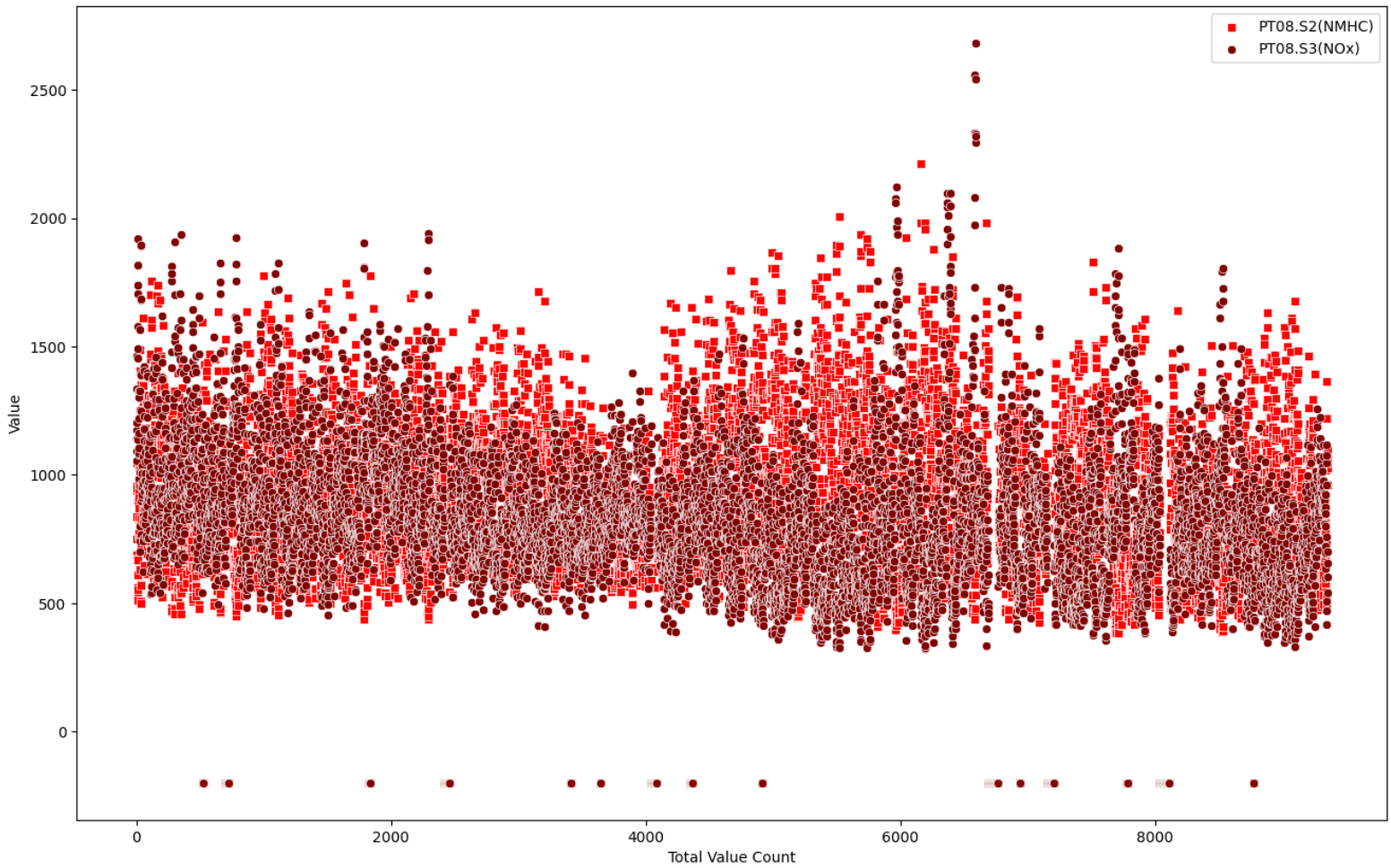
Out[4]:
```
Date             0
Time             0
CO(GT)           0
PT08.S1(CO)      0
NMHC(GT)         0
C6H6(GT)         0
PT08.S2(NMHC)    0
NOx(GT)          0
PT08.S3(NOx)     0
NO2(GT)          0
PT08.S4(NO2)     0
PT08.S5(O3)      0
T                0
RH               0
AH               0
dtype: int64
```

Plotting

In [5]:
```
sns.distplot(df["CO(GT)"])
plt.show()
```

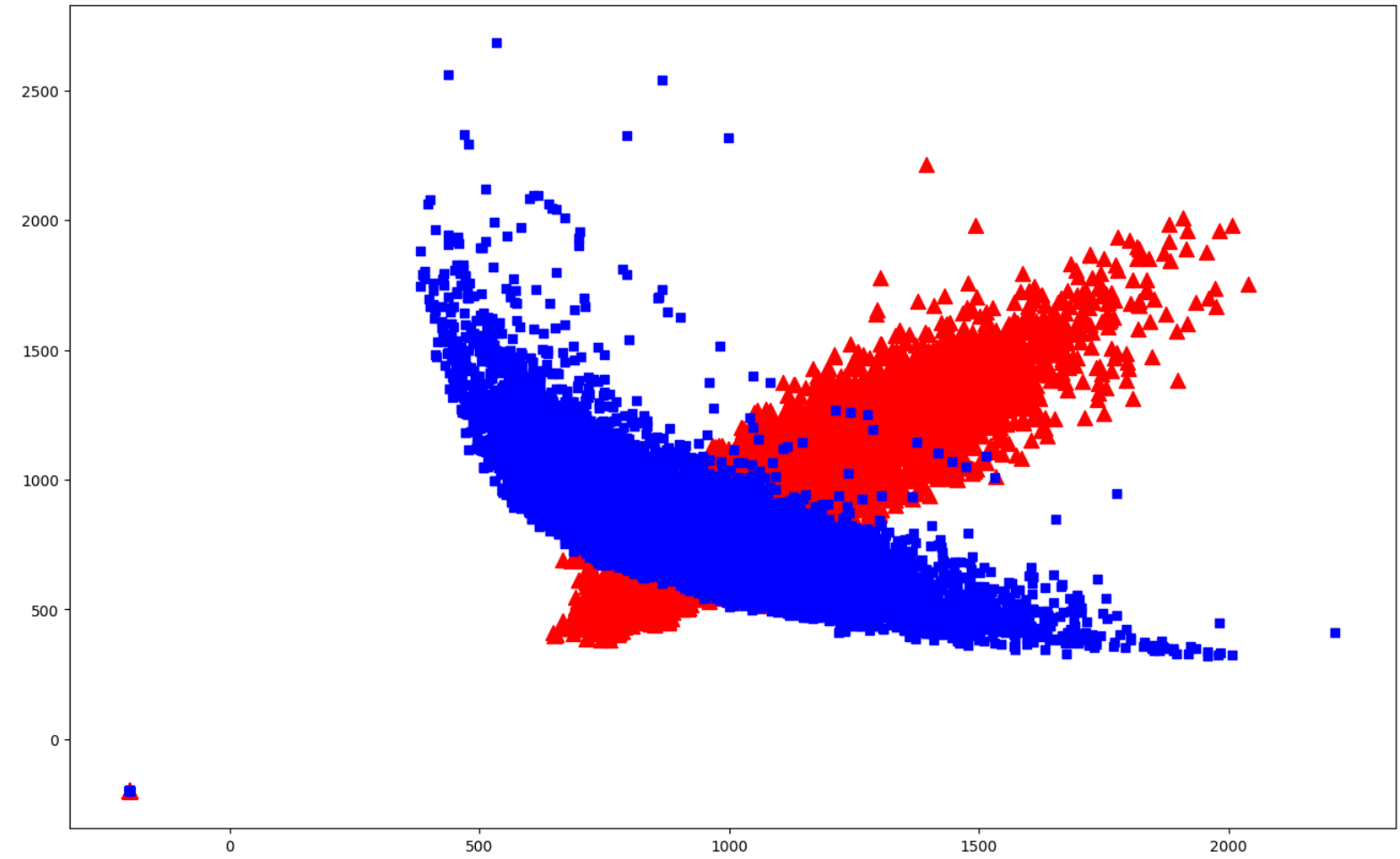


In [6]:
```
plt.figure(figsize=(16, 10))
# sns.set(style='whitegrid')
sns.scatterplot(df["PT08.S2(NMHC)"], color="red", marker="s")
sns.scatterplot(df["PT08.S3(NOx)"], color="maroon")
plt.legend(["PT08.S2(NMHC)","PT08.S3(NOx)"])
plt.xlabel("Total Value Count")
plt.ylabel("Value")
plt.show()
```



In [7]:
```
plt.figure(figsize=(16, 10))
plt.scatter(df['PT08.S1(CO)'], df['PT08.S2(NMHC)'], c=["r"], marker ="^", s=100)
plt.scatter(df['PT08.S2(NMHC)'],df['PT08.S3(NOx)'] , c=["b"], marker ="s", s=30)
```

Out[7]:<matplotlib.collections.PathCollection at 0x1d09d2e5fd0>



In [8]:
```
df.columns
```

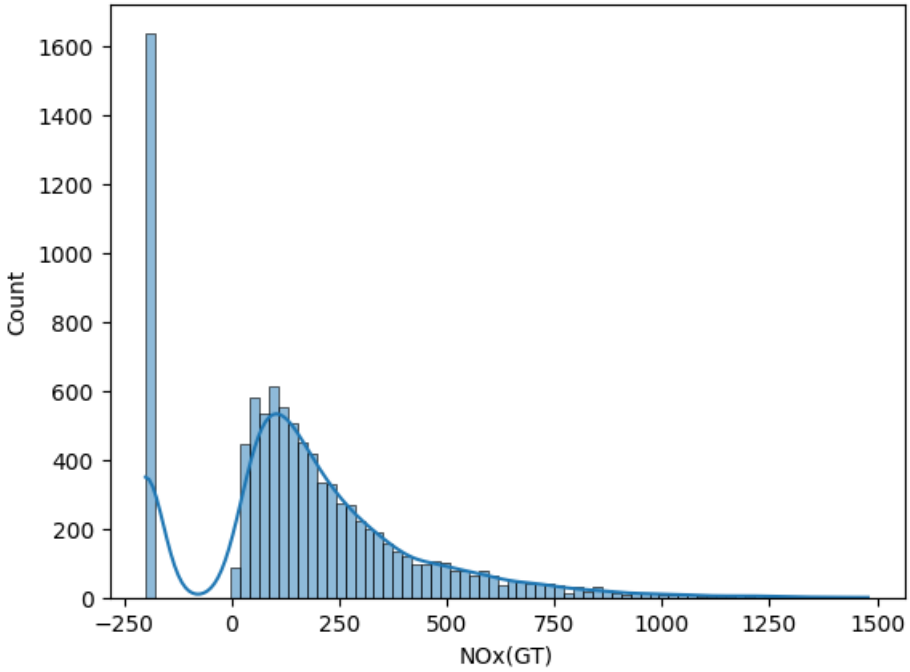
Out[8]:
```
Index(['Date', 'Time', 'CO(GT)', 'PT08.S1(CO)', 'NMHC(GT)', 'C6H6(GT)',
       'PT08.S2(NMHC)', 'NOx(GT)', 'PT08.S3(NOx)', 'NO2(GT)', 'PT08.S4(NO2)',
       'PT08.S5(O3)', 'T', 'RH', 'AH'],
      dtype='object')
```

In [9]:
```
df.dtypes
```

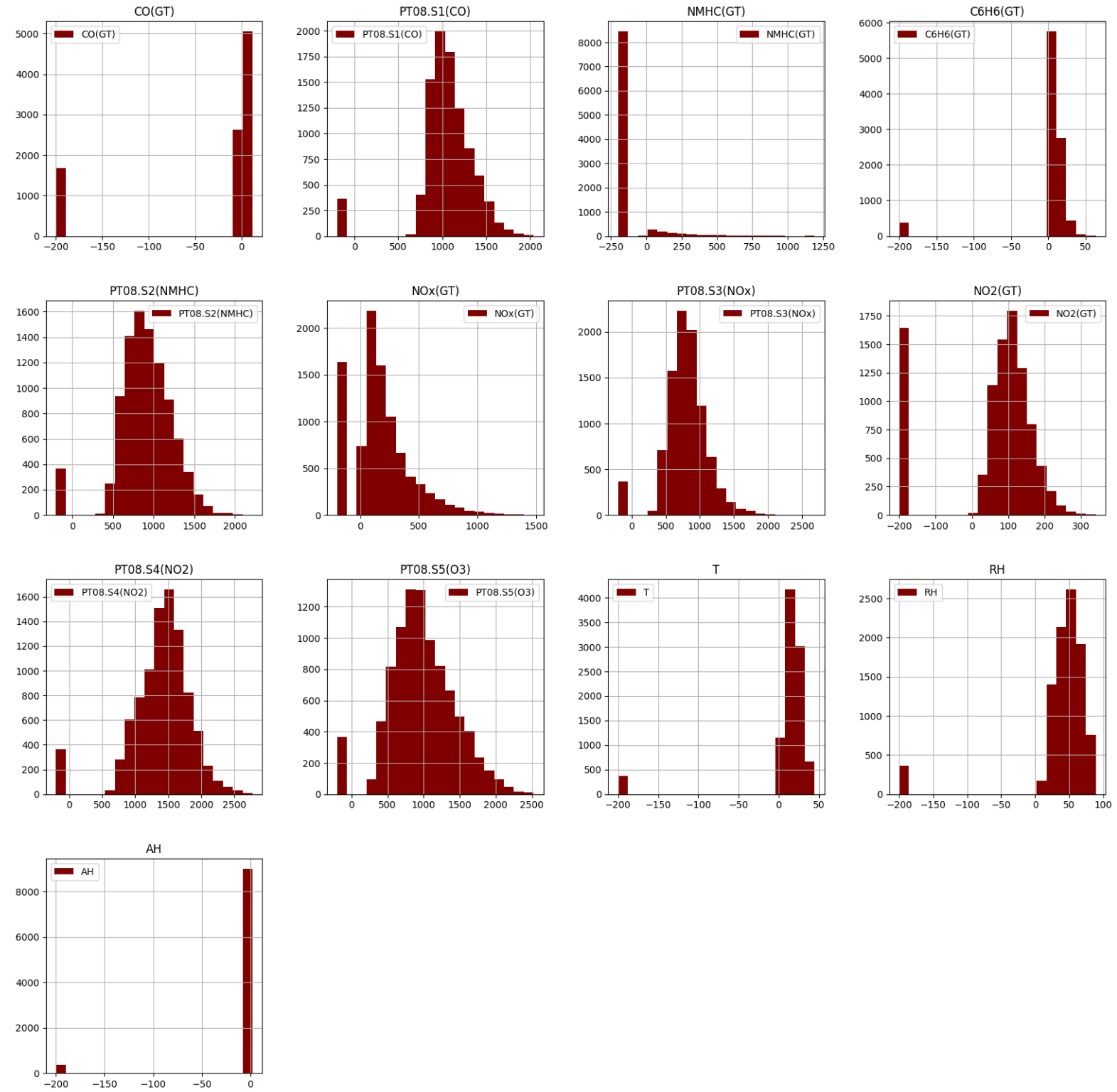
Out[9]:
```
Date            object
Time            object
CO(GT)         float64
PT08.S1(CO)      int64
NMHC(GT)         int64
C6H6(GT)       float64
PT08.S2(NMHC)    int64
NOx(GT)          int64
PT08.S3(NOx)     int64
NO2(GT)          int64
PT08.S4(NO2)     int64
PT08.S5(O3)      int64
T              float64
RH             float64
AH             float64
dtype: object
```

In [10]:
```
sns.histplot(df['NOx(GT)'], kde=True)
```

Out[10]:<Axes: xlabel='NOx(GT)', ylabel='Count'>

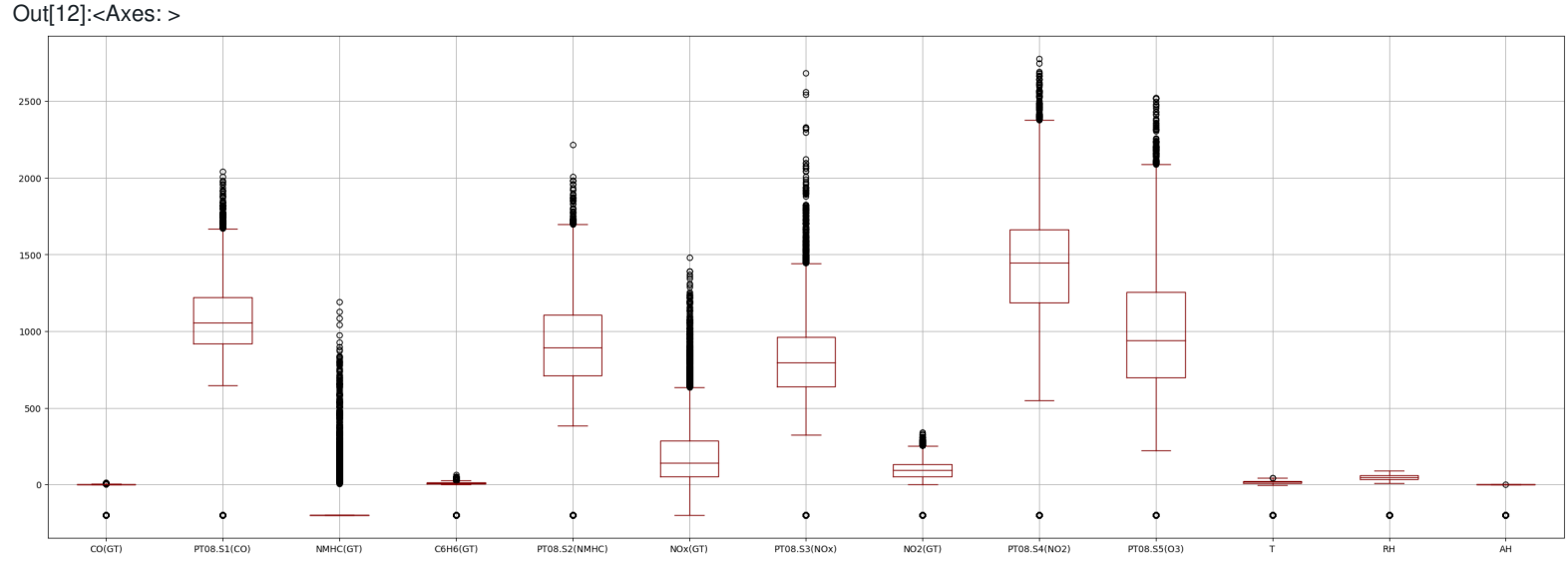


In [11]:
```
df.hist(figsize=(20,20),
        grid=True,
        bins=20,
        color="maroon",
        legend=True)
plt.show()
```



In [12]:
```
df.boxplot(figsize=(30,10),
           color="maroon",
           grid=True)
```

Out[12]:<Axes: >



In [13]:
```
x=df['PT08.S2(NMHC)']
y=df['PT08.S3(NOx)']
```

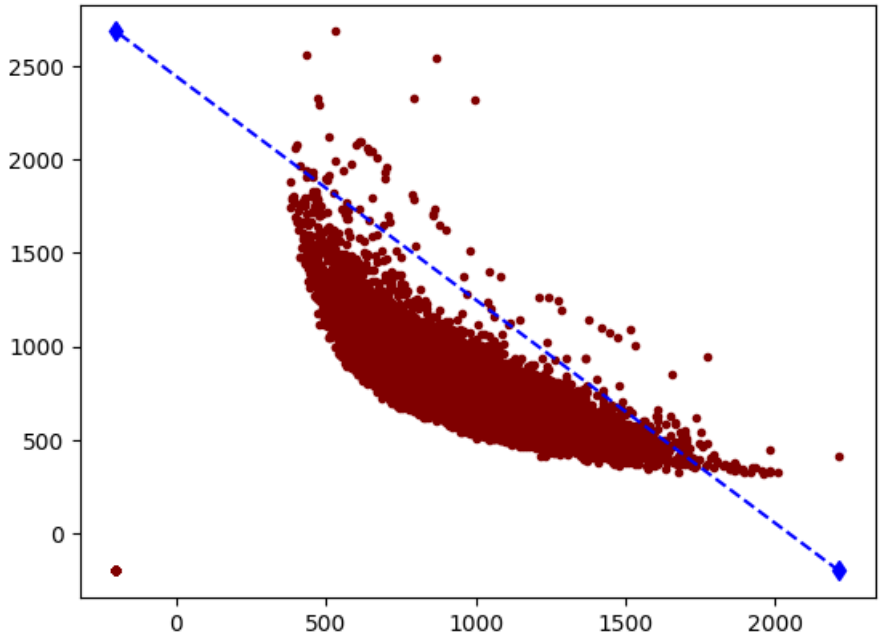
In [14]:
```
df.describe().T
```

Out[14]:

	count	mean	std	min	25%	50%	75%	max
CO(GT)	9357.0	-34.207524	77.657170	-200.0	0.6000	1.5000	2.6000	11.900
PT08.S1(CO)	9357.0	1048.990061	329.832710	-200.0	921.0000	1053.0000	1221.0000	2040.000
NMHC(GT)	9357.0	-159.090093	139.789093	-200.0	-200.0000	-200.0000	-200.0000	1189.000
C6H6(GT)	9357.0	1.865683	41.380206	-200.0	4.0000	7.9000	13.6000	63.700
PT08.S2(NMHC)	9357.0	894.595276	342.333252	-200.0	711.0000	895.0000	1105.0000	2214.000
NOx(GT)	9357.0	168.616971	257.433866	-200.0	50.0000	141.0000	284.0000	1479.000
PT08.S3(NOx)	9357.0	794.990168	321.993552	-200.0	637.0000	794.0000	960.0000	2683.000
NO2(GT)	9357.0	58.148873	126.940455	-200.0	53.0000	96.0000	133.0000	340.000
PT08.S4(NO2)	9357.0	1391.479641	467.210125	-200.0	1185.0000	1446.0000	1662.0000	2775.000
PT08.S5(O3)	9357.0	975.072032	456.938184	-200.0	700.0000	942.0000	1255.0000	2523.000
T	9357.0	9.778305	43.203623	-200.0	10.9000	17.2000	24.1000	44.600
RH	9357.0	39.485380	51.216145	-200.0	34.1000	48.6000	61.9000	88.700
AH	9357.0	-6.837604	38.976670	-200.0	0.6923	0.9768	1.2962	2.231

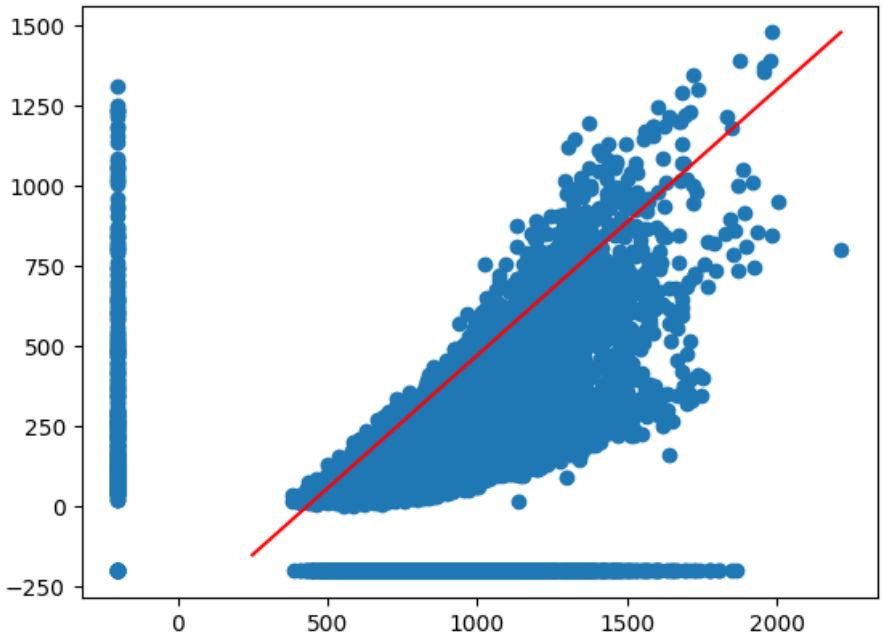
In [15]:
```
plt.scatter(df['PT08.S2(NMHC)'],df['PT08.S3(NOx)'], color='maroon', marker='.')
plt.plot([x.max(),x.min()], [y.min(),y.max()],'--', c='b', marker="d")
```

Out[15]:[<matplotlib.lines.Line2D at 0x1d0c321ac10>]



In [16]:
```
plt.scatter(df['PT08.S2(NMHC)'], df['NOx(GT)'])
plt.plot([250,2214],[-150, 1479], "r-")
```

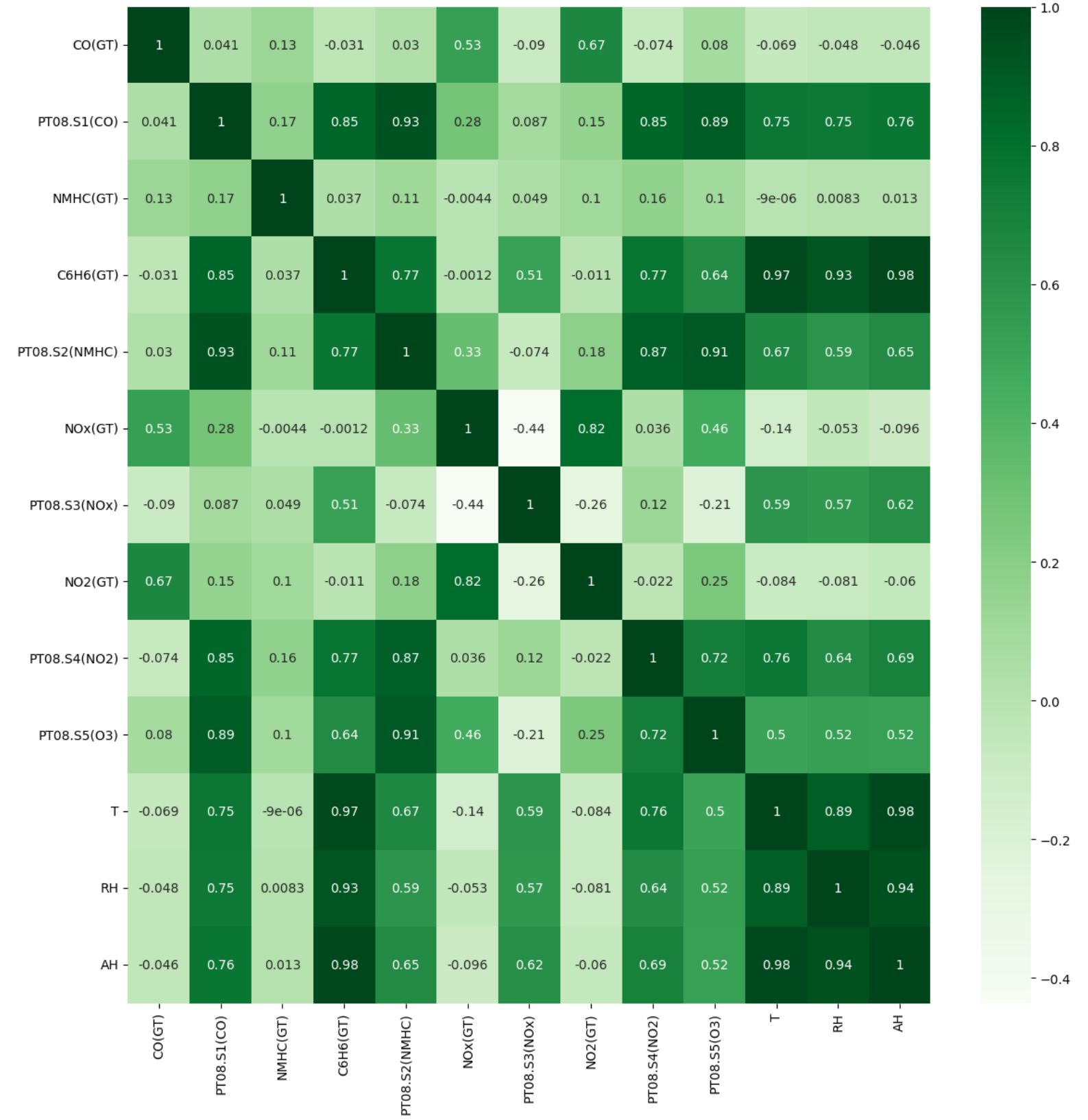
Out[16]:[<matplotlib.lines.Line2D at 0x1d0c381dd90>]



In [ ]:

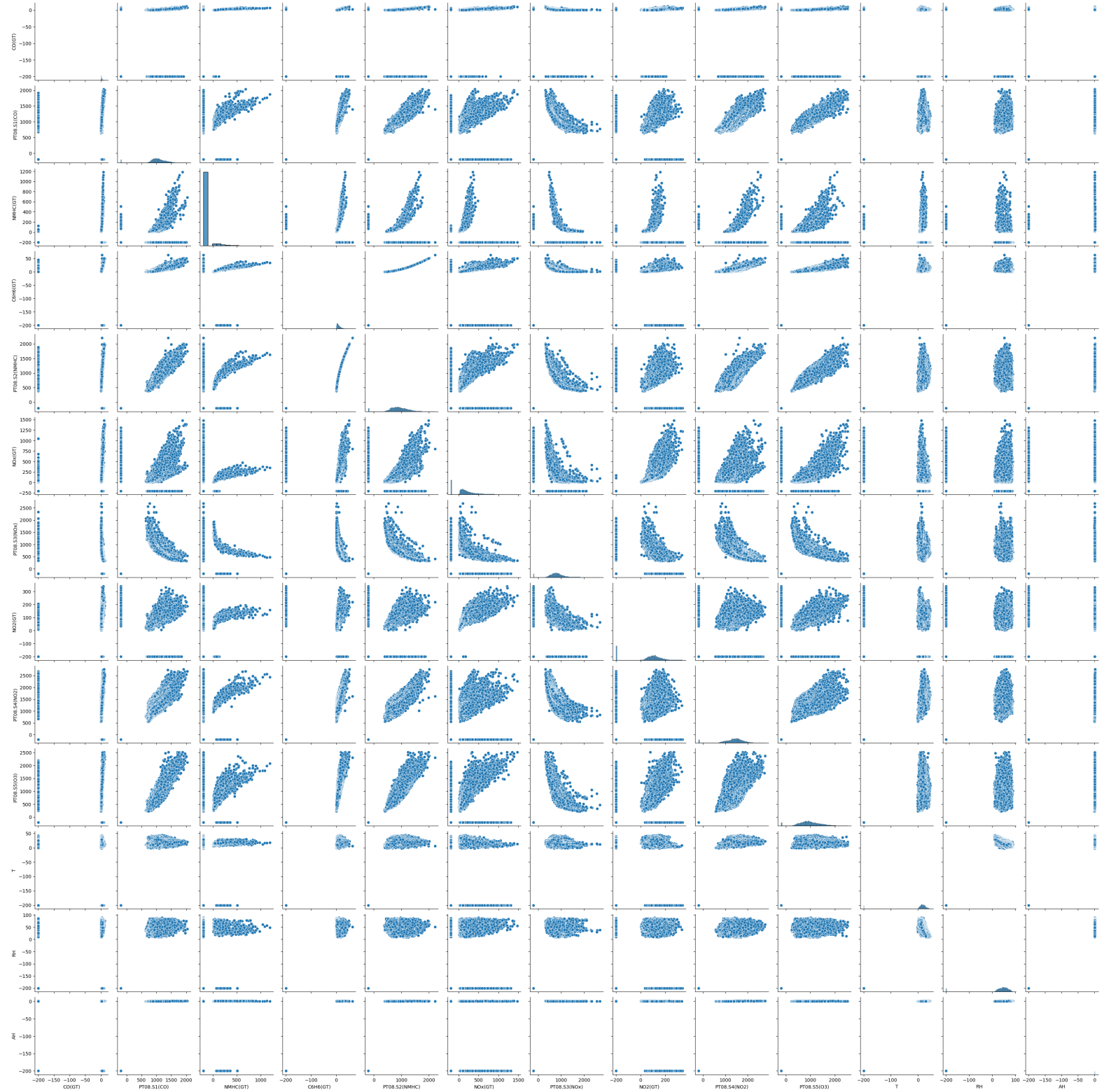
In [17]:
```
plt.figure(figsize=(14, 14))
sns.heatmap(df.corr(), annot=True, cmap="Greens")
```

Out[17]:<Axes: >



In [19]:
```
sns.pairplot(df)
```

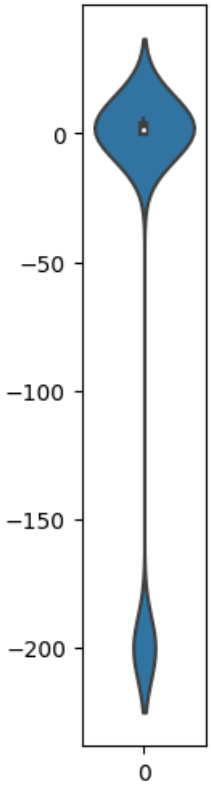
Out[19]:<seaborn.axisgrid.PairGrid at 0x1d0c375c750>



In [45]:
```
plt.figure(figsize=(1, 6))

sns.violinplot(df["CO(GT)"])
```

Out[45]:<Axes: >



In [38]:
```
fig = plt.figure(figsize=(2,4))

ax = fig.add_axes([0,0,1,1])
pl = ax.violinplot(df['PT08.S4(NO2)'])
plt.show()
```