



Projet ASI322:

Étude de l'impact de la pollution de l'air sur taux de décès maladies respiratoires

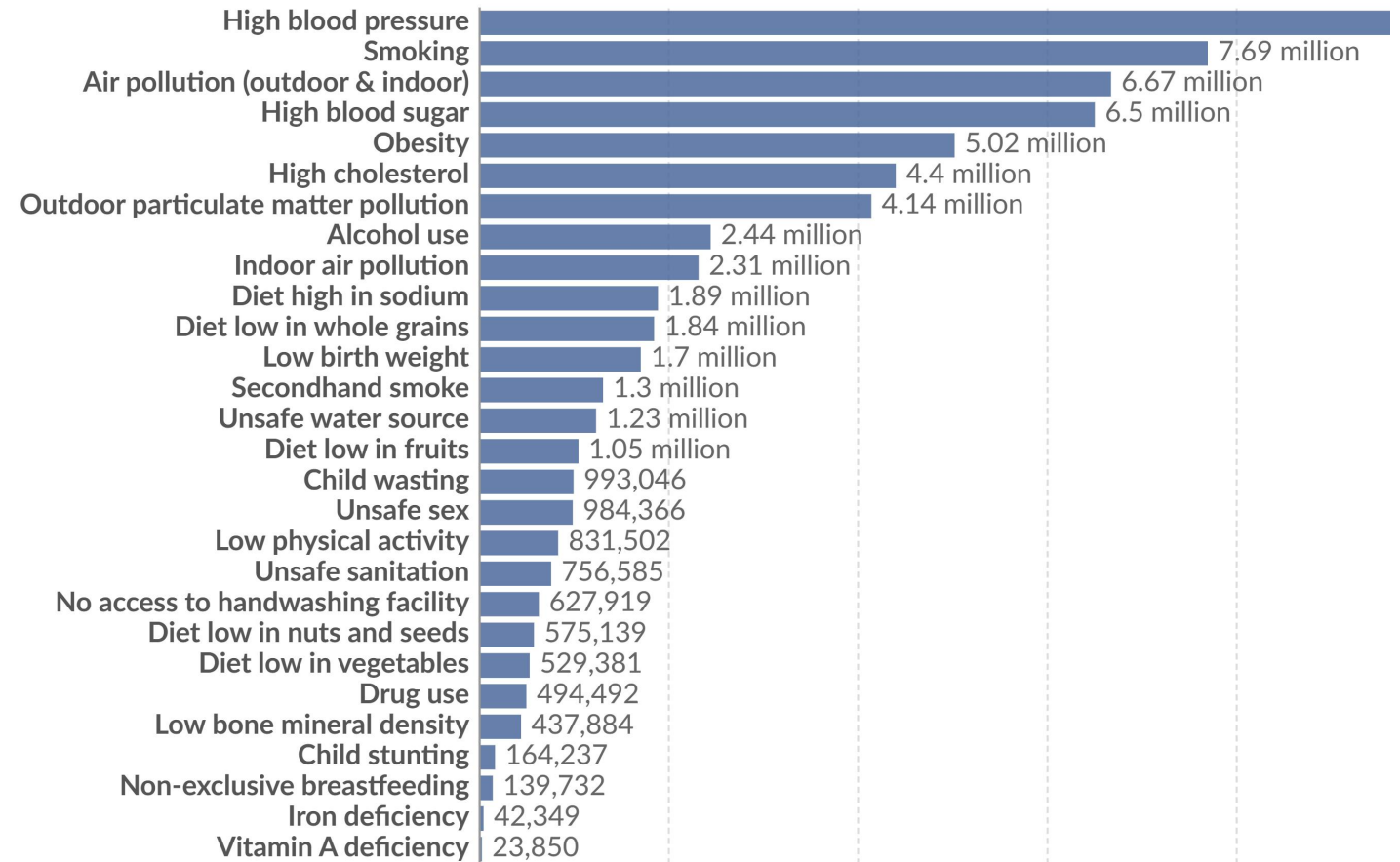
Jia CAO, 2024/01/09

Contenu:

- Partie 0: Problématique
- Partie 1: Acquisition des données et prétraitement
 - 1.1: problème rencontré
- Partie 2: Analyse et traitement des données
- Partie 3: Machine Learning
- Partie 4: Conclusion

Deaths by risk factor, World, 2019

The estimated annual number of deaths attributed to each risk factor¹. Estimates come with wide uncertainties especially for countries with poor vital registration².



Contenu:

- Partie 0: Problématique
- Partie 1: Acquisition des données et prétraitement
 - 1.1: problème rencontré
- Partie 2: Analyse et traitement des données
- Partie 3: Machine Learning
- Partie 4: Conclusion

Particulate Pollutants	Gaseous Pollutants	
<ol style="list-style-type: none">1. Lead2. Fly Ash3. Metallic Oxides4. Nanoparticles	<ol style="list-style-type: none">1. Carbon monoxide (CO)2. Carbon dioxide (CO₂)3. Chlorofluorocarbons (CFCs)4. Ozone (O₃)5. Nitrogen oxide (NO_x)6. Sulphur dioxide (SO₂)	<ul style="list-style-type: none">• Volatile organic compounds (VOCs)• Benzene• Ethylene• Biological pollutants• Asbestos• Radon

IQA Niveau de pollution de l'air

0 - 50	Bon
51 - 100	Modéré
101 - 150	Mauvais pour les groupes sensibles
151 - 200	Mauvais
201 - 300	Très mauvais
300 +	Dangereux

$$AQI = \max(AQI_{PM_{2.5}}, AQI_{PM_{10}}, AQI_{O_3}, \dots)$$

Ozone (ppm) – truncate to 3 decimal places

PM_{2.5} (µg/m³) – truncate to 1 decimal place

PM₁₀ (µg/m³) – truncate to integer

CO (ppm) – truncate to 1 decimal place

SO₂ (ppb) – truncate to integer

NO₂ (ppb) – truncate to integer

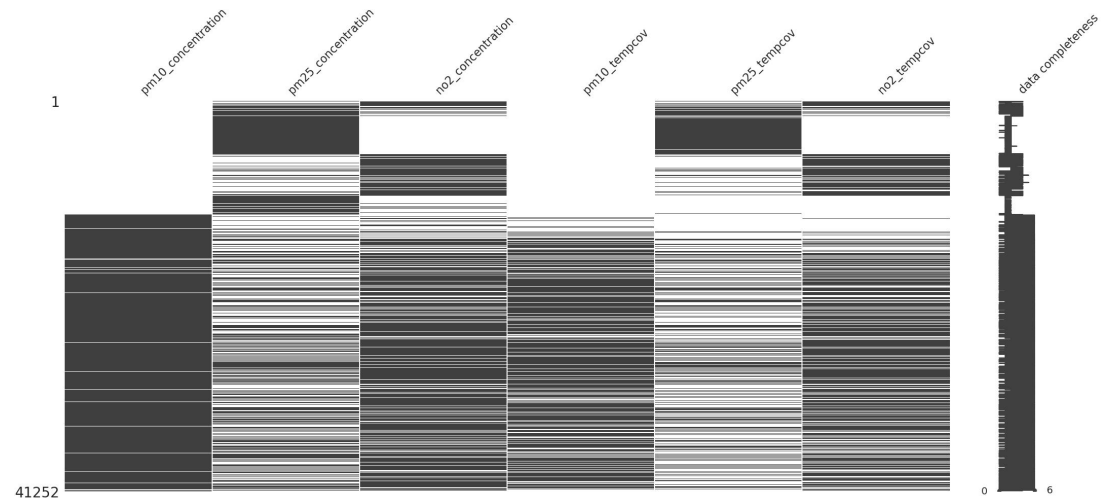
Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 aspects:

- polluants & qualité de l'air
- maladies respiratoire

Source de données:

1. (OMS) Base de données sur la qualité de l'air ambiant (données **annuelles**, selon les **villes**)
['pm10_concentration' 'pm25_concentration' 'no2_concentration' 'pm10_tempcov' 'pm25_tempcov' 'no2_tempcov' 'population']



Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 aspects:

- polluants & qualité de l'air
- maladies respiratoire

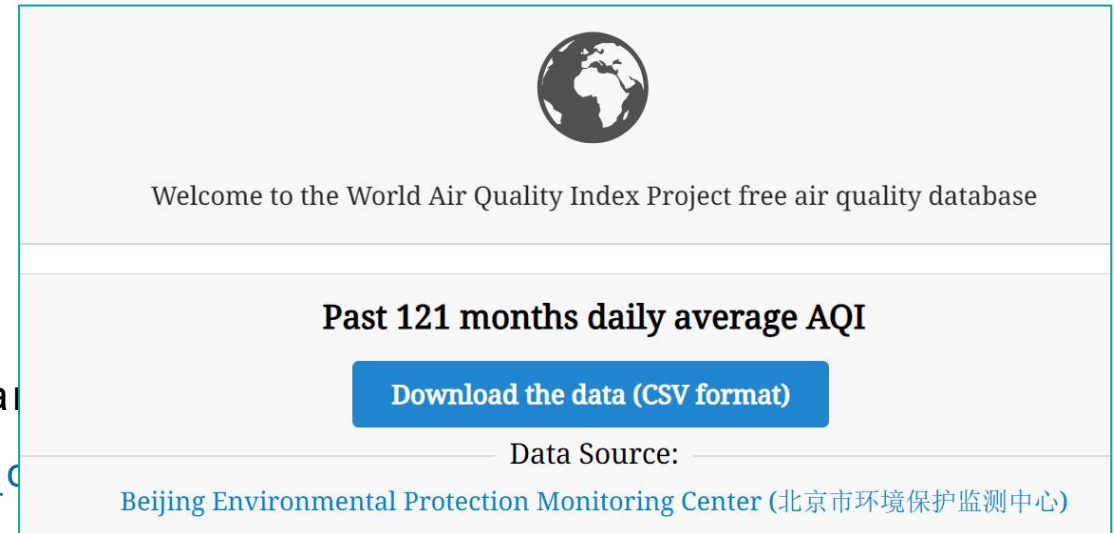
Source de données:

1. (OMS) Base de données sur la qualité de l'air a

`['pm10_concentration' 'pm25_concentration' 'no2_c'
'no2_tempcov' 'population']`

2. (WAQI) Base de données sur la qualité de l'air ambiant (données **quoditiennes**, selon les **villes**)

`[' pm25', ' pm10', ' o3', ' no2', ' so2', ' co']` NB: ici ce sont des IQA des polluants



Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 concepts:

- polluants & qualité de l'air

- maladies respiratoire

Source de données:

1. (OMS) Base de données sur l

`['pm10_concentration' 'pm25_co
'no2_tempcov' 'population']`

2. (WAQI) Base de données sur la qualité de l'air ambiant (données **quoditiennes**, selon les **villes**)

`[' pm2.5', ' pm10', ' o3', ' no2', ' so2', ' co']` NB: ici ce sont des IQA des polluants

3. [Berkeley Earth] données **horaires**, selon **ville** ou **pays**

`['UTC hour', 'pm2.5 (concentration)']`

Search for country, region or city:

RESET

Guadeloupe
Guatemala
Guernsey
Guinea
Honduras
Hong Kong
Hungary
Iceland

Palestine
Panama
Paraguay
Peru
Philippines
Poland
Portugal
Puerto Rico

BERKELEY EARTH™

Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 aspects:

- polluants & qualité de l'air
- maladies respiratoire

Source de données:

4. (aqistudy.cn) Base de données des villes chinoises (données **quoditiennes**, par **ville**)



日期	AQI	质量等级	PM2.5	PM10	SO2	CO	NO2	O3_8h
2014-01-01	195	中度污染	147	181	63	1.7	99	61
2014-01-02	147	轻度污染	113	131	37	1.6	95	60
2014-01-03	189	中度污染	142	163	56	1.4	96	45

Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 aspects:

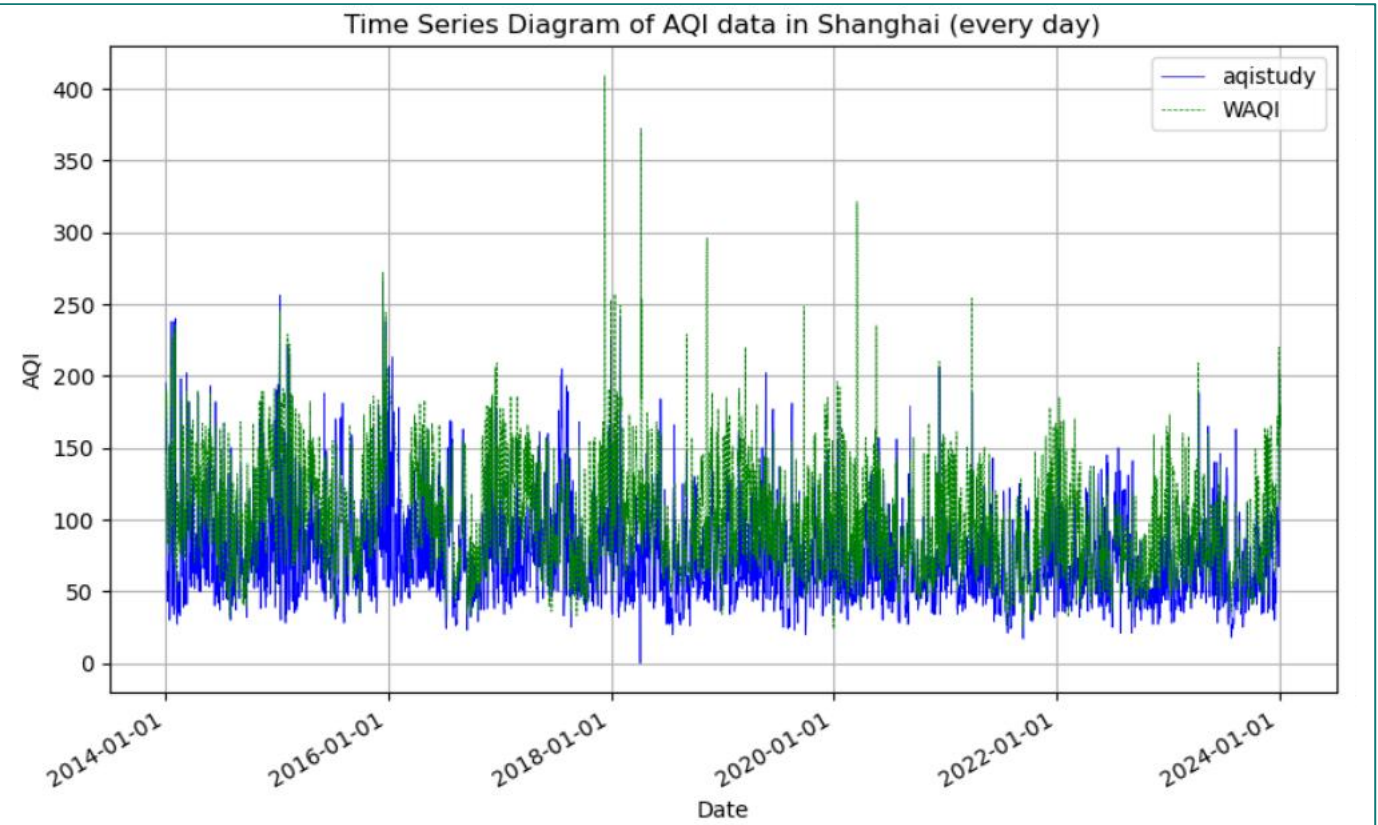
- polluants & qualité de l'air
- maladies respiratoire

Source de données:

4. (aqistudy.cn) Base de données des vill



日期	AQI	质量等级	PM2.5					
2014-01-01	195	中度污染	147	181	63	1.7	99	61
2014-01-02	147	轻度污染	113	131	37	1.6	95	60
2014-01-03	189	中度污染	142	163	56	1.4	96	45



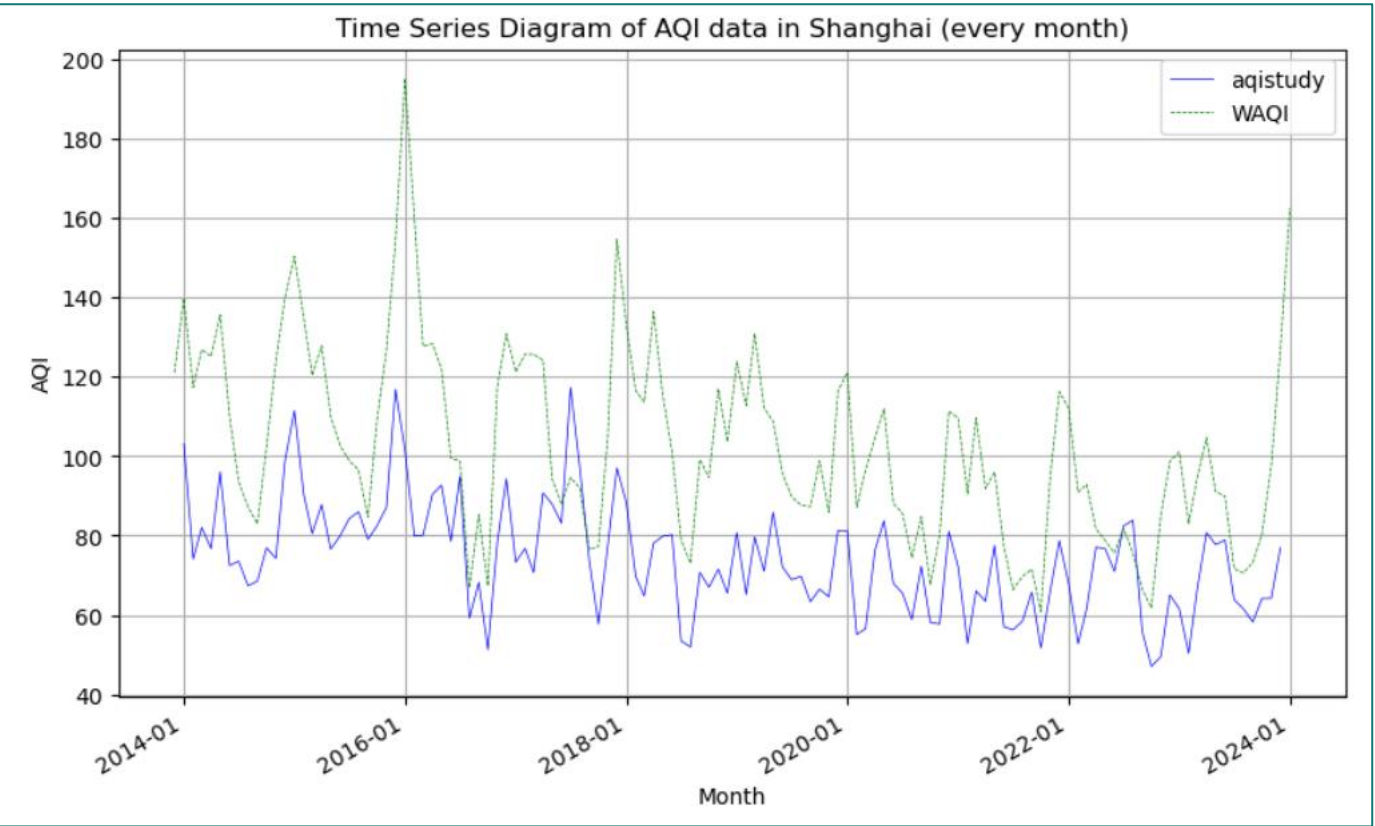
Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 aspects:

- polluants & qualité de l'air
- maladies respiratoire

Source de données:

4. (aqistudy.cn) Base de données des vill



日期	AQI	质量等级	PM2.5					
2014-01-01	195	中度污染	147	181	63	1.7	99	61
2014-01-02	147	轻度污染	113	131	37	1.6	95	60
2014-01-03	189	中度污染	142	163	56	1.4	96	45

Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 aspects

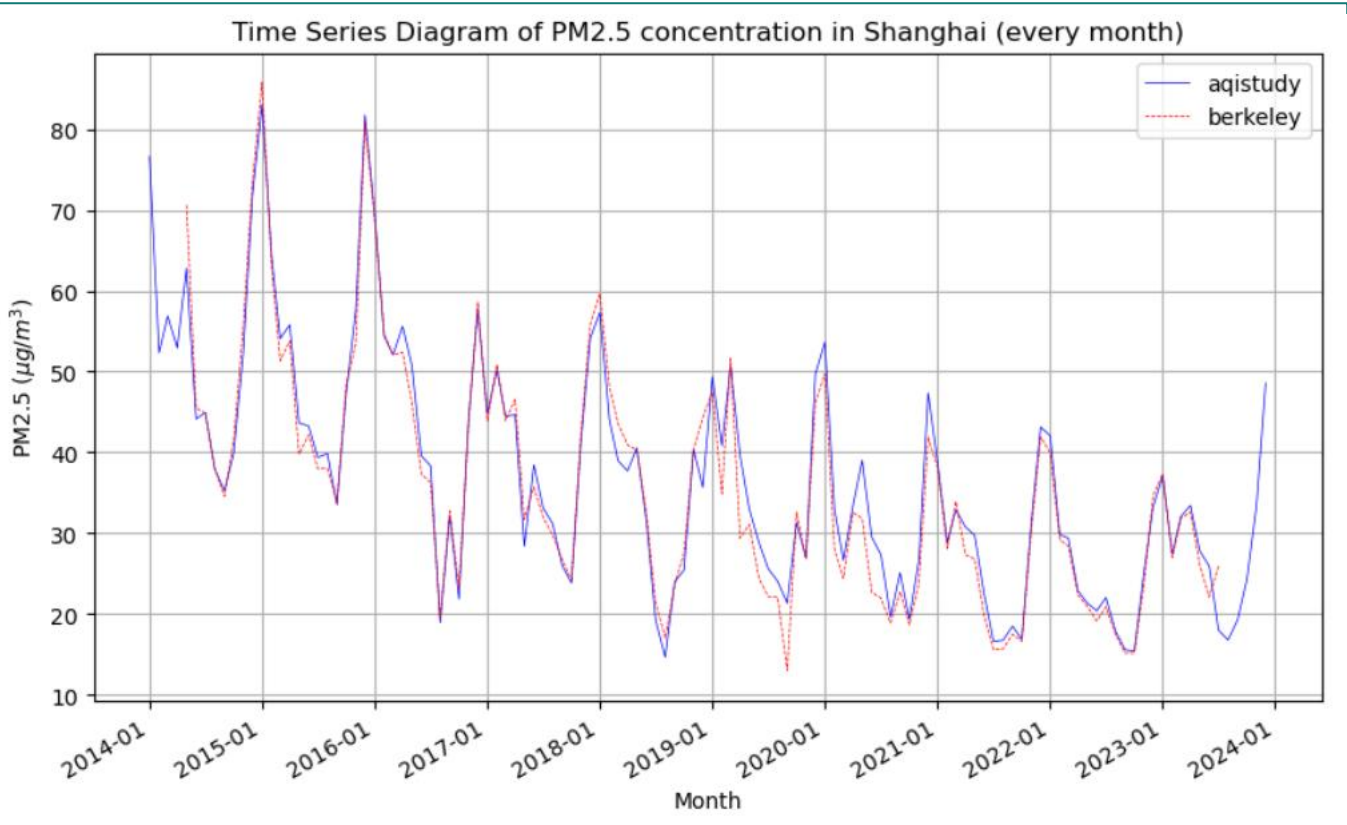
- polluants & qualité de l'air
- maladies respiratoire

Source de données:

4. (aqistudy.cn) Base de données des vill



日期	AQI	质量等级	PM2.5					
2014-01-01	195	中度污染	147	181	63	1.7	99	61
2014-01-02	147	轻度污染	113	131	37	1.6	95	60
2014-01-03	189	中度污染	142	163	56	1.4	96	45



Partie 1: Acquisition des données et prétraitement (1)

Recherche de données sur 2 aspects

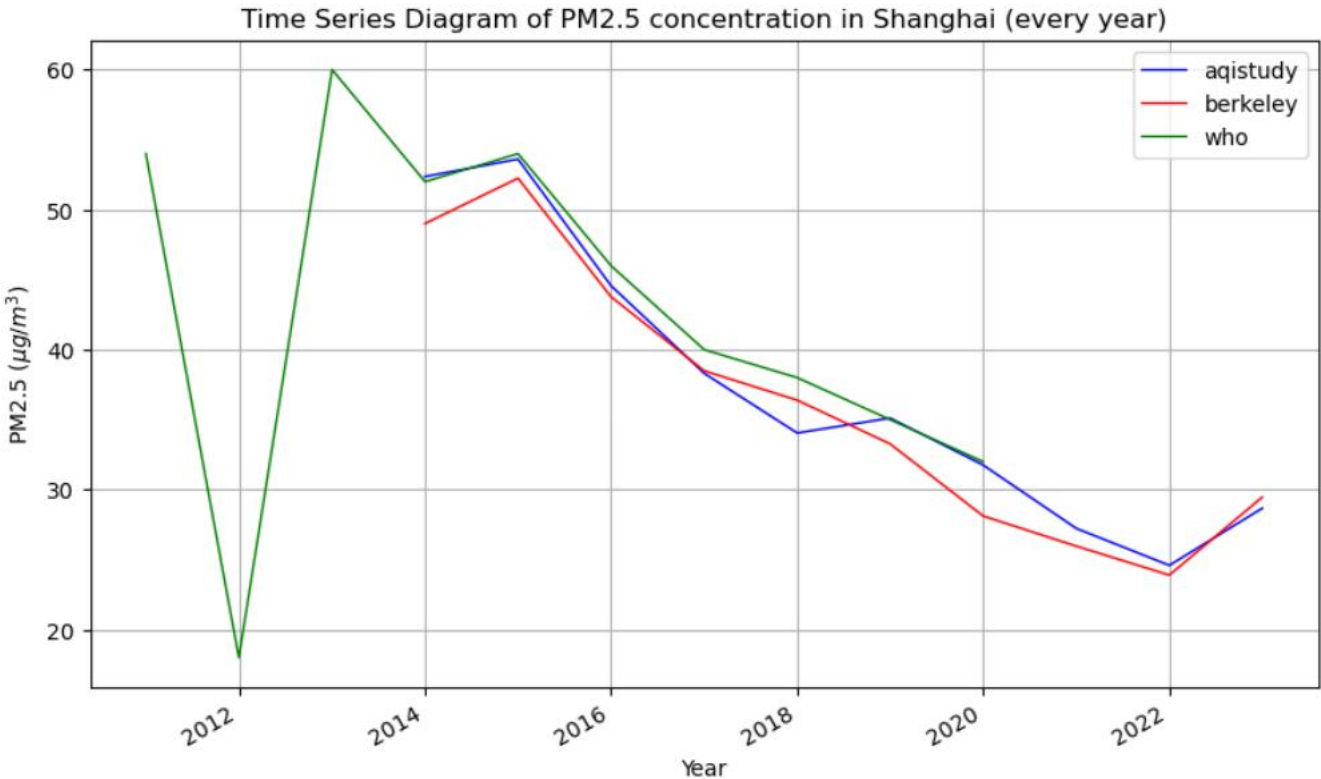
- polluants & qualité de l'air
- maladies respiratoire

Source de données:

4. (aqistudy.cn) Base de données des vill



日期	AQI	质量等级	PM2.5					
2014-01-01	195	中度污染	147	181	63	1.7	99	61
2014-01-02	147	轻度污染	113	131	37	1.6	95	60
2014-01-03	189	中度污染	142	163	56	1.4	96	45



空气质量 分指数 (IAQI)	污染物项目浓度限值									
	二氧化硫 (SO ₂) 24 小时 平均/ (μg/m ³)	二氧化硫 (SO ₂) 1 小时 平均/ (μg/m ³) ⁽¹⁾	二氧化氮 (NO ₂) 24 小时 平均/ (μg/m ³)	二氧化氮 (NO ₂) 1 小时 平均/ (μg/m ³) ⁽¹⁾	颗粒物 (粒径小 于等于 10μm) 24 小时 平均/ (μg/m ³)	一氧化碳 (CO) 24 小时 平均/ (mg/m ³)	一氧化碳 (CO) 1 小时 平均/ (mg/m ³) ⁽¹⁾	臭氧 (O ₃) 1 小时 平均/ (μg/m ³)	臭氧 (O ₃) 8 小时滑 动平均/ (μg/m ³)	颗粒物 (粒径小 于等于 2.5μm) 24 小时 平均/ (μg/m ³)
0	0	0	0	0	0	0	0	0	0	0
50	50	150	40	100	50	2	5	160	100	35
100	150	500	80	200	150	4	10	200	160	75
150	475	650	180	700	250	14	35	300	215	115
200	800	800	280	1 200	350	24	60	400	265	150
300	1 600	⁽²⁾	565	2 340	420	36	90	800	800	250
400	2 100	⁽²⁾	750	3 090	500	48	120	1 000	⁽³⁾	350
500	2 620	⁽²⁾	940	3 840	600	60	150	1 200	⁽³⁾	500

These Breakpoints...							...equal this AQI	...and this category
O ₃ (ppm) 8-hour	O ₃ (ppm) 1-hour ¹	PM _{2.5} (μg/m ³) 24-hour	PM ₁₀ (μg/m ³) 24-hour	CO (ppm) 8-hour	SO ₂ (ppb) 1-hour	NO ₂ (ppb) 1-hour	AQI	
0.000 - 0.054	-	0.0 – 12.0	0 - 54	0.0 - 4.4	0 - 35	0 - 53	0 - 50	Good
0.055 - 0.070	-	12.1 – 35.4	55 - 154	4.5 - 9.4	36 - 75	54 - 100	51 - 100	Moderate
0.071 - 0.085	0.125 - 0.164	35.5 – 55.4	155 - 254	9.5 - 12.4	76 - 185	101 - 360	101 - 150	Unhealthy for Sensitive Groups
0.086 - 0.105	0.165 - 0.204	(55.5 - 150.4) ³	255 - 354	12.5 - 15.4	(186 - 304) ⁴	361 - 649	151 - 200	Unhealthy
0.106 - 0.200	0.205 - 0.404	(150.5 - 250.4) ³	355 - 424	15.5 - 30.4	(305 - 604) ⁴	650 - 1249	201 - 300	Very unhealthy
⁽²⁾	0.405 - 0.504	(250.5 - 350.4) ³	425 - 504	30.5 - 40.4	(605 - 804) ⁴	1250 - 1649	301 - 400	Hazardous
⁽²⁾	0.505 - 0.604	(350.5 - 500.4) ³	505 - 604	40.5 - 50.4	(805 - 1004) ⁴	1650 - 2049	401 - 500	Hazardous

Problème Rencontré

Recherche de données sur 2 aspects:

- polluants & qualité de l'air
- maladies respiratoire


Problème:

1. PAS de données sur le nombre de personnes diagnostiquées
2. PAS de données quoditiennes ou mensuelles
3. PAS de données en fonction des villes

Partie 1: Acquisition des données et prétraitement

1. IHME: Institute for Health Metrics and Evaluation

location	cause	year	Death Rate
Guyana	Respiratory infec...	1990	48.524773827237134
Guinea-Bissau	Respiratory infec...	1990	260.93392193693967
Brunei Darussalam	Chronic respirato...	1990	28.047355314816443
Honduras	Respiratory infec...	1990	44.64436203593287
Kuwait	Chronic respirato...	1990	4.6259092778190425

 IHME | GHDx

Search

Explore results from the 2019 Global Burden of Disease (GBD) study. For more info, refer to the [About](#) section.

Measure ⓘ
Deaths ×

Metric ⓘ
Rate ×

Cause ⓘ
Respiratory infections and tuberculosis ×
Chronic respiratory diseases ×

Location
Armenia × Azerbaijan × Georgia ×
Kazakhstan × +200 more

Partie 1: Acquisition des données et prétraitement

2. <https://ourworldindata.org/explorers/air-pollution> (source: Zenodo)



['Entity' 'Year'
'Nitrogen oxide (NO_x)'
'Sulphur dioxide (SO₂)'
'Carbon monoxide (CO)'
'Organic carbon (OC)'
'NMVOCs'
'Black carbon (BC)'
'Ammonia (NH₃)'
'Nitrogen oxide (NO_x).1'
'Sulphur dioxide (SO₂).1'
'Carbon monoxide (CO).1'
'Organic carbon (OC).1'
'NMVOCs.1'
'Black carbon (BC).1'
'Ammonia (NH₃).1']

Partie 1: Acquisition des données et prétraitement (1)

3. (The World Bank) World Development Indicators

```
df_wdi.columns.values
```

```
array(['Country Name', 'Country Code', 'Indicator Name', 'Indicator Code',  
      '1960', '1961', '1962', '1963', '1964', '1965', '1966', '1967',  
      '1968', '1969', '1970', '1971', '1972', '1973', '1974', '1975',  
      '1976', '1977', '1978', '1979', '1980', '1981', '1982', '1983',  
      '1984', '1985', '1986', '1987', '1988', '1989', '1990', '1991',  
      '1992', '1993', '1994', '1995', '1996', '1997', '1998', '1999',  
      '2000', '2001', '2002', '2003', '2004', '2005', '2006', '2007',  
      '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015',
```

Les indicateurs qu'on veut:

```
selected_indicators = {  
    "NY.GDP.MKTP.CD": "GDP",           # GDP  
    "EN.ATM.PM25.MC.M3": "PM2.5",     # PM2.5  
    "SP.POP.TOTL": "Population",       # Population, total  
    "AG.SRF.TOTL.K2": "Surface",       # Surface area (sq. km)  
}
```

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	...	2014	2015	2016
0	Africa Eastern and Southern	AFE	Access to clean fuels and technologies for coo...	EG.CFT.ACCS.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	17.392349	17.892005	18.359993
1	Africa Eastern and Southern	AFE	Access to clean fuels and technologies for coo...	EG.CFT.ACCS.RU.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	6.720331	7.015917	7.281390

Partie 1: Acquisition des données et prétraitement

3. (The World Bank) World Development Indicators

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962
0	Africa Eastern and Southern	AFE	Access to clean fuels and technologies for coo...	EG.CFT.ACCS.ZS	NaN	NaN	NaN
1	Africa Eastern and Southern	AFE	Access to clean fuels and technologies for coo...	EG.CFT.ACCS.RU.ZS	NaN	NaN	NaN

→
pivot

	Country Name	Country Code	1990			
Indicator Code			AG.SRF.TOTL.K2	EN.ATM.PM25.MC.M3	NY.GDP.MKTP.CD	SP.POP.TOTL
0	Afghanistan	AFG	6.528600e+05	49.282398	NaN	1.069480e+07
1	Africa Eastern and Southern	AFE	1.510674e+07	30.132449	2.546735e+11	3.098907e+08
2	Africa Western and Central	AFW	9.166270e+06	64.258847	1.218036e+11	2.067390e+08
3	Algeria	ALD	2.275222e+06	21.217122	2.222551e+10	2.222551e+08

→
stack

Indicator Code	Country Name	Country Code	level_2	AG.SRF.TOTL.K2	EN.ATM.PM25.MC.M3	NY.GDP.MKTP.CD	SP.POP.TOTL
0	Afghanistan	AFG	1990	652860.0	49.282398	NaN	10694796.0
1	Afghanistan	AFG	1991	652860.0	NaN	NaN	10745167.0
2	Afghanistan	AFG	1992	652860.0	NaN	NaN	12057433.0
3	Afghanistan	AFG	1993	652860.0	NaN	NaN	14003760.0
4	Afghanistan	AFG	1994	652860.0	NaN	NaN	15455555.0
...

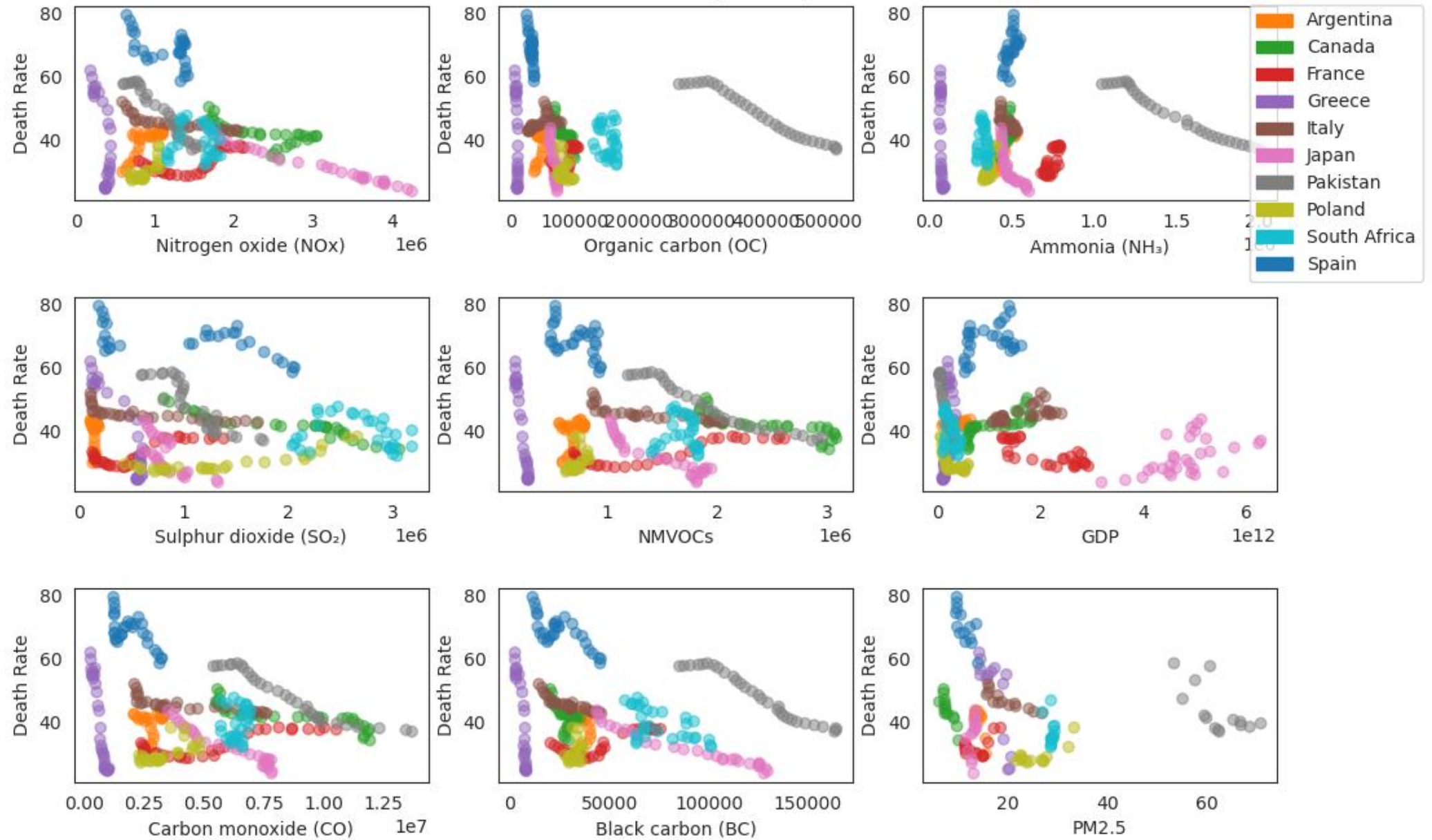
Partie 2: Analyse et traitement des données

- Premières étape:

Join==> deux dataframe (merged_CR et merged_RI, RI=Respiratory Infectious, CR= Chronic Respiratory Diseases)

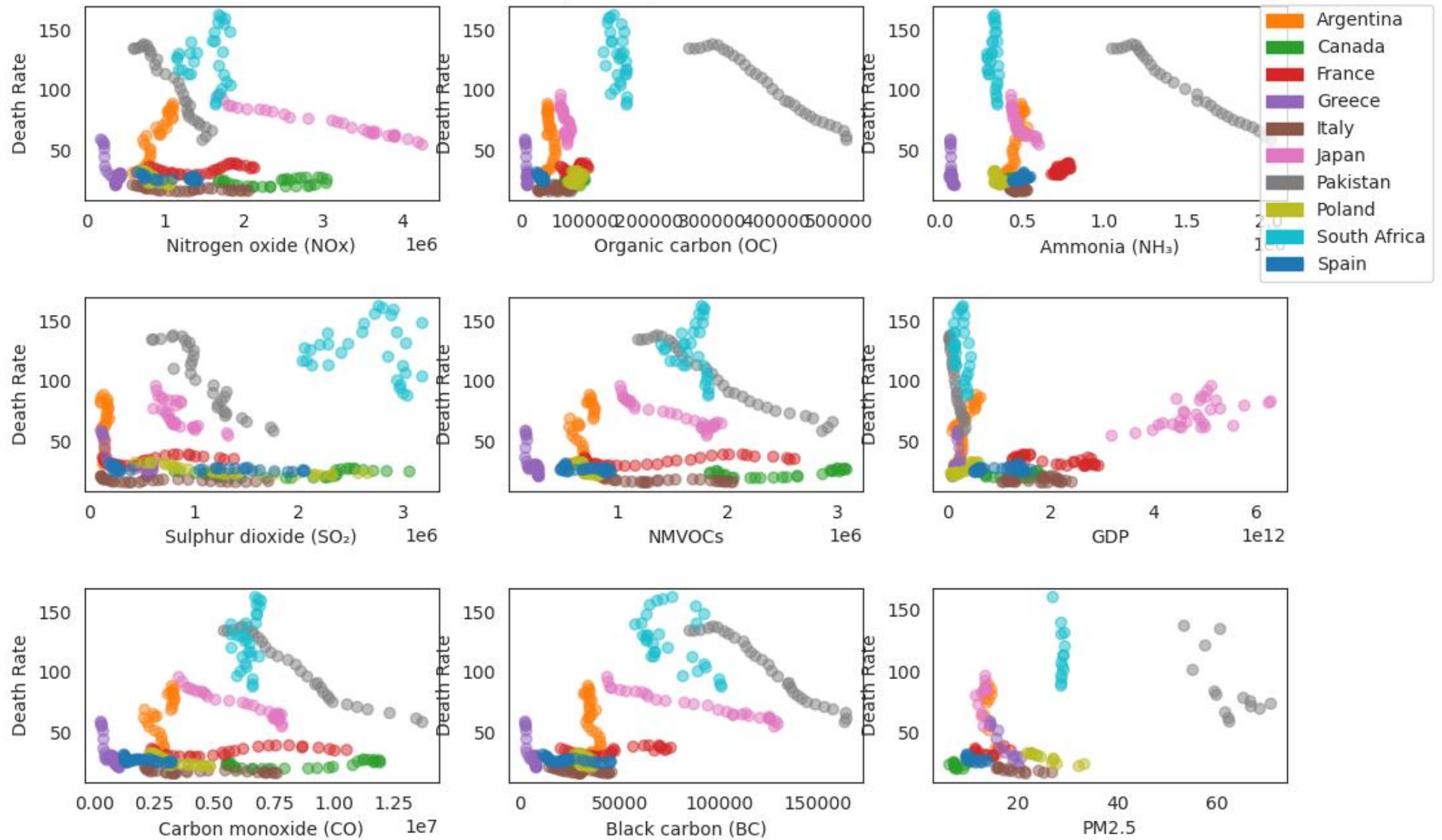
- Scatter Diagram of Death Data and other attributes:

Death Rate for Chronic respiratory diseases

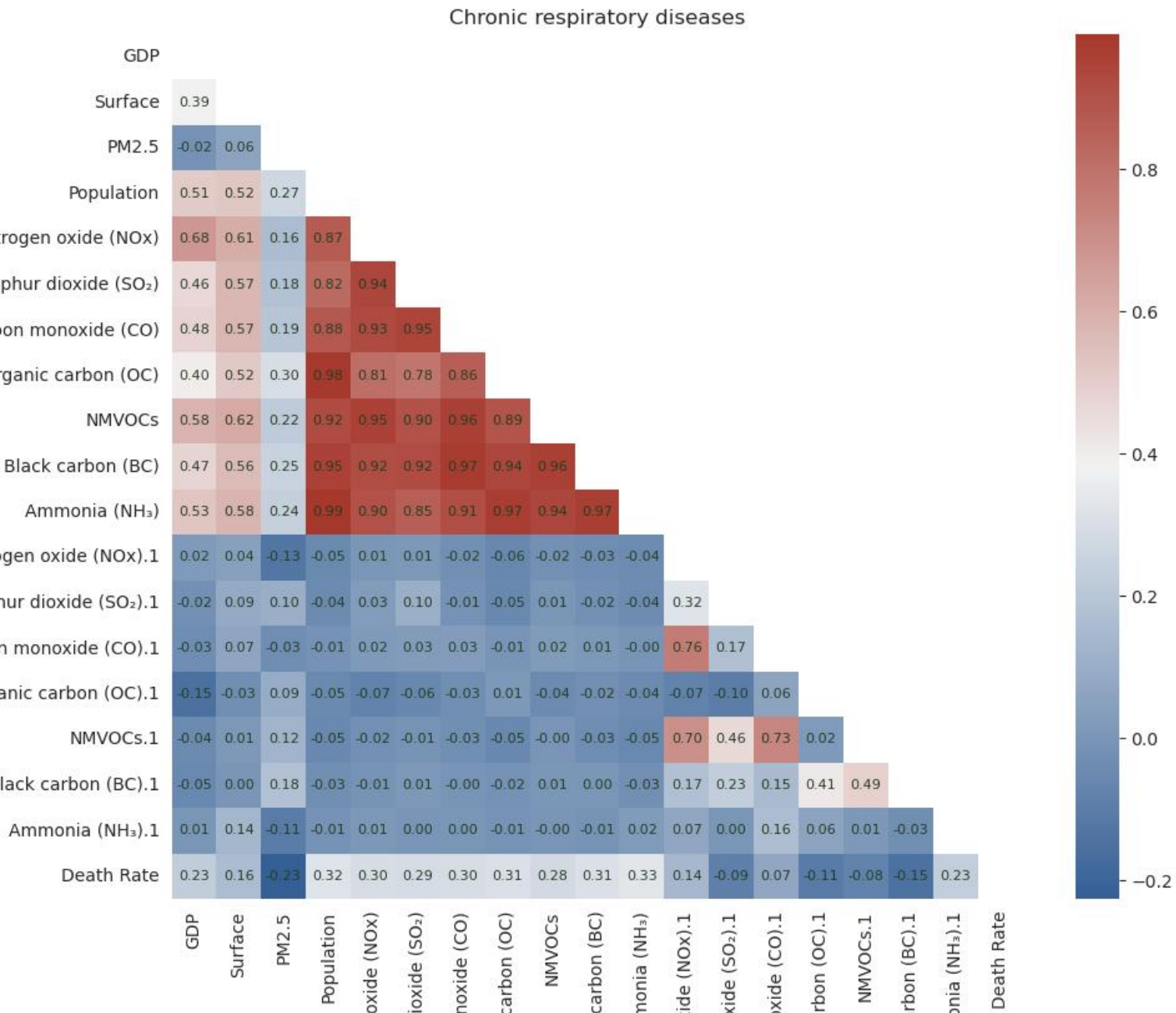


maladie respiratoire chronique

Death Rate for Respiratory infections and tuberculosis



infections respiratoires

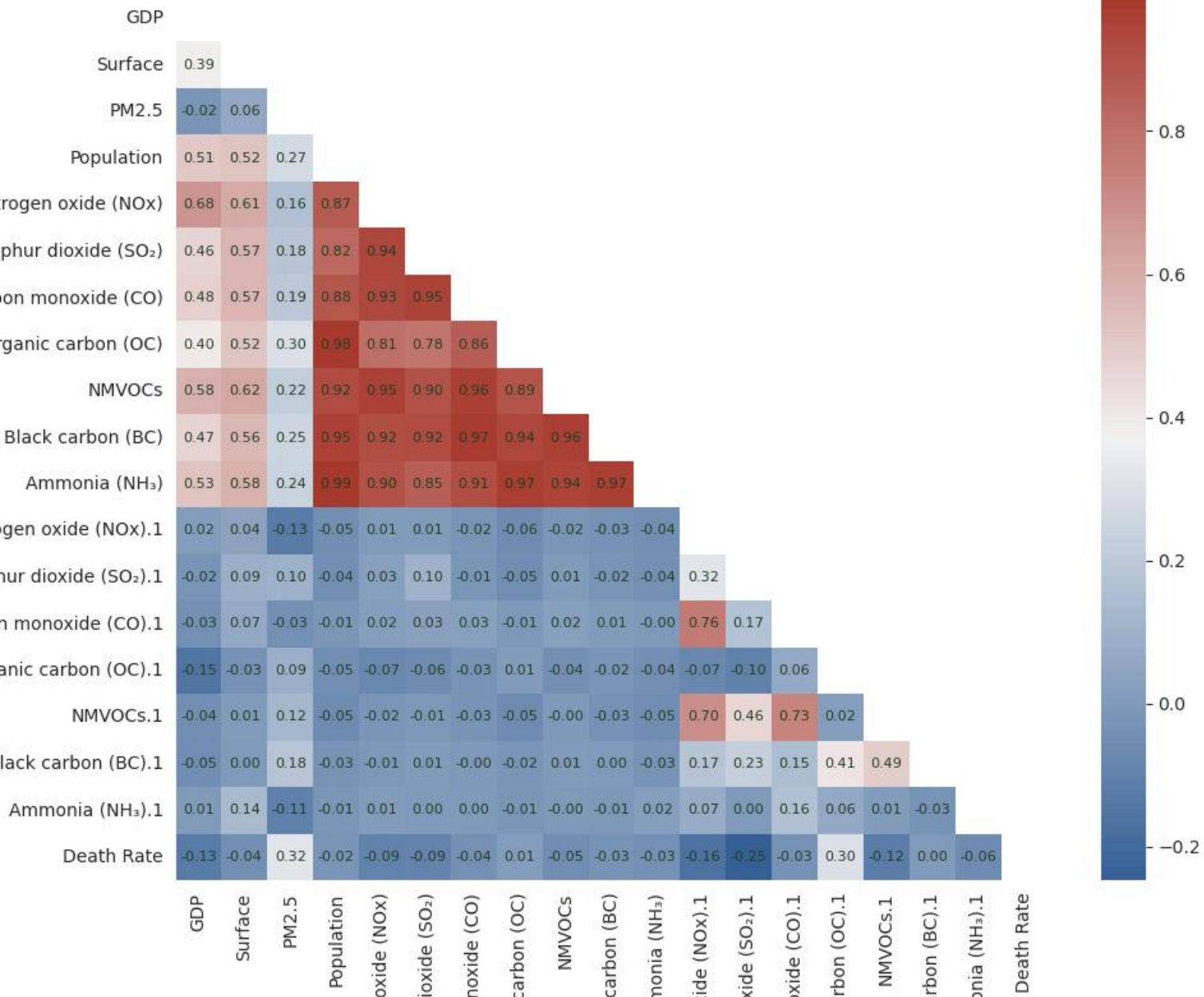


problème de colinéarité

maladie respiratoire chronique



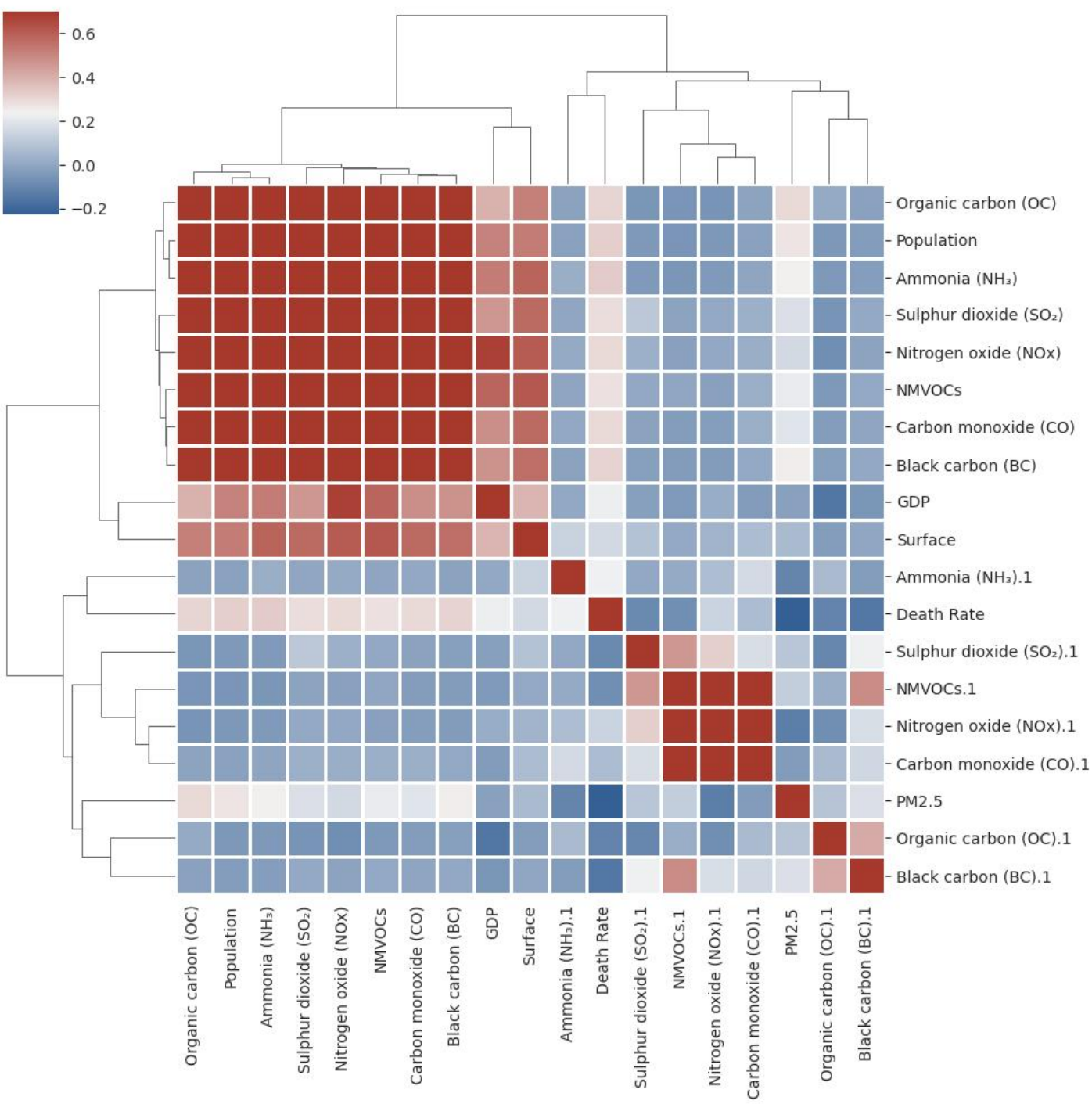
Respiratory infections and tuberculosis



problème de colinéarité

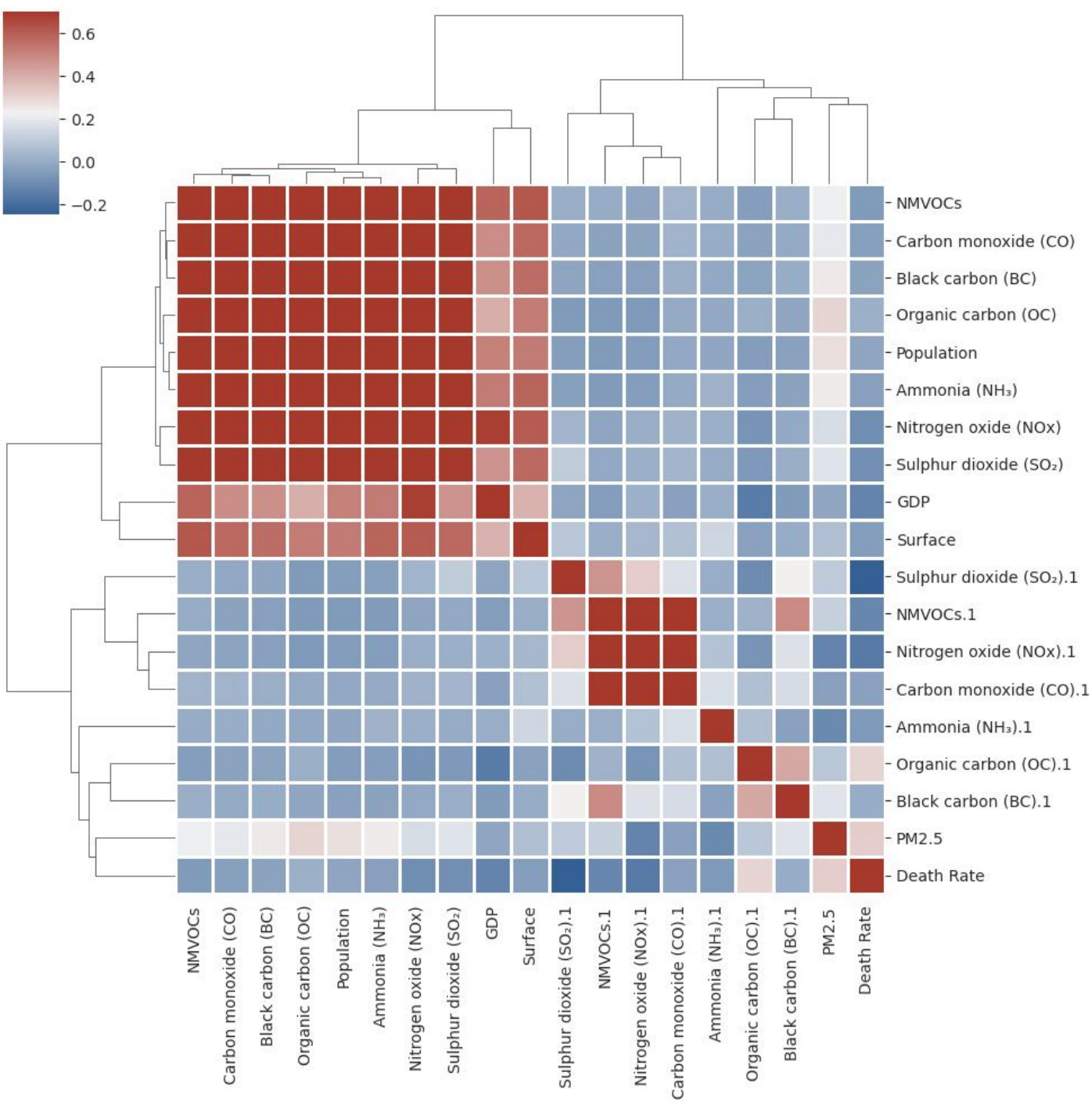
infections respiratoires





maladie respiratoire chronique

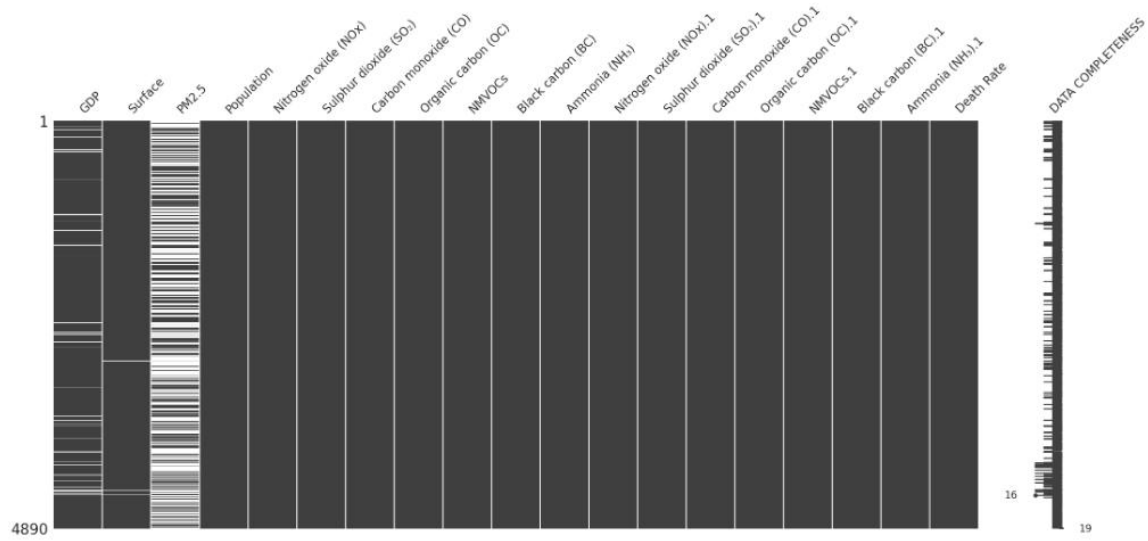




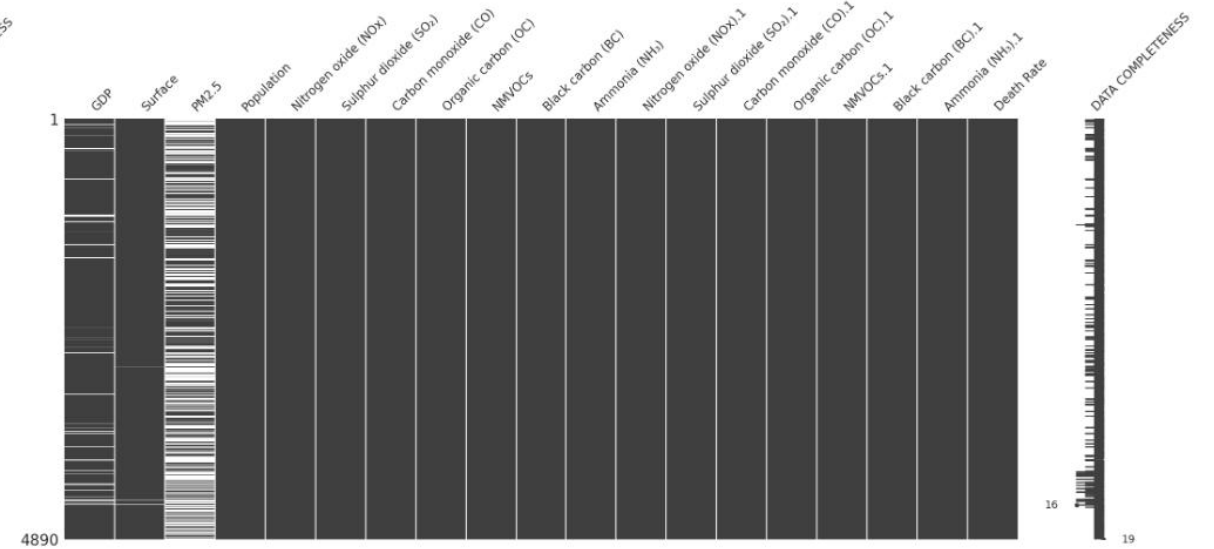
infections respiratoires



Partie 3: Machine Learning



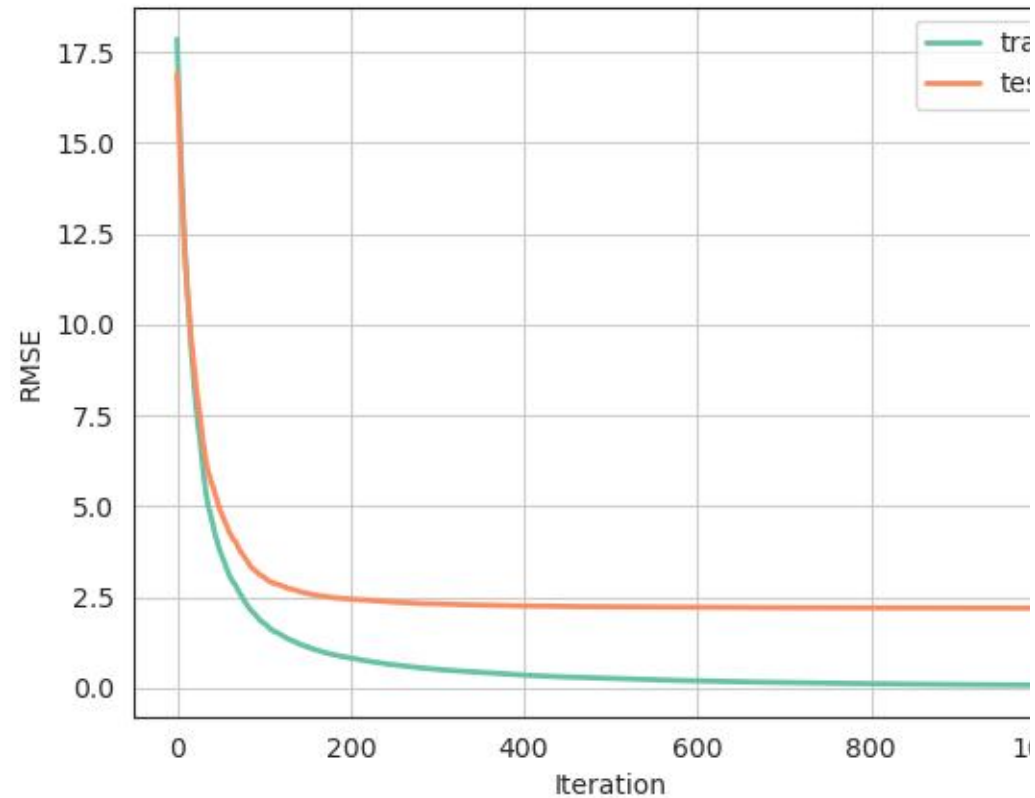
maladie respiratoire chronique



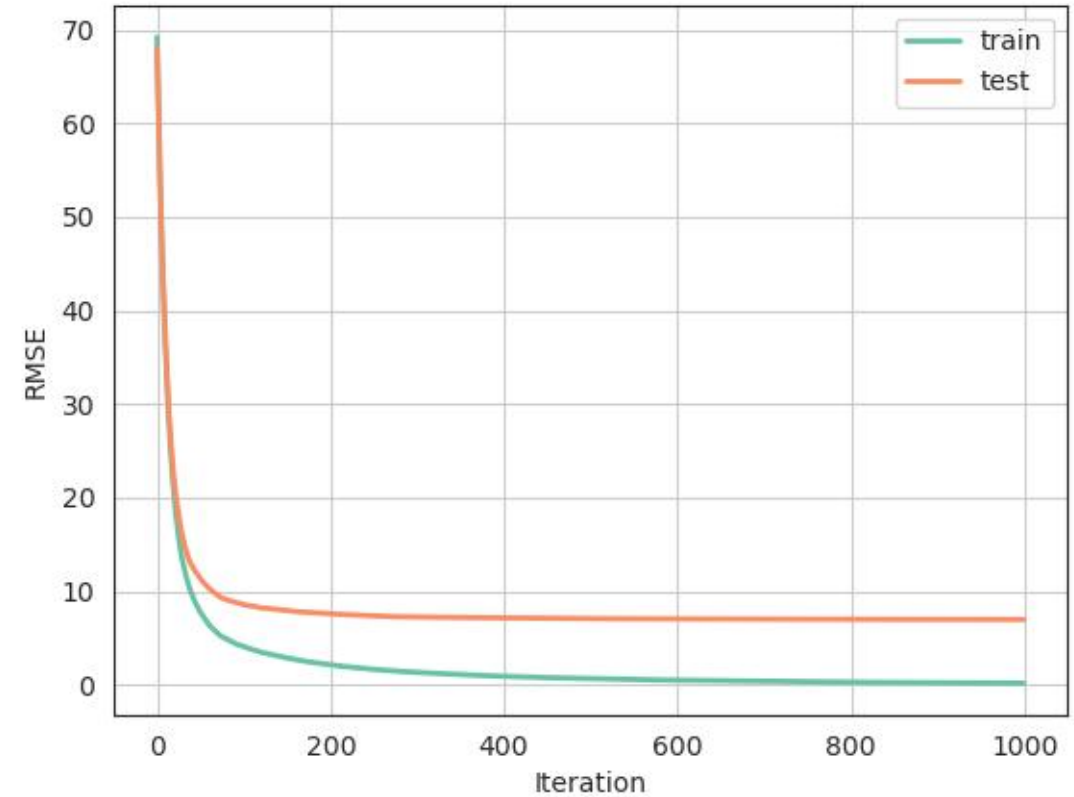
infections respiratoires

Partie 3: Machine Learning

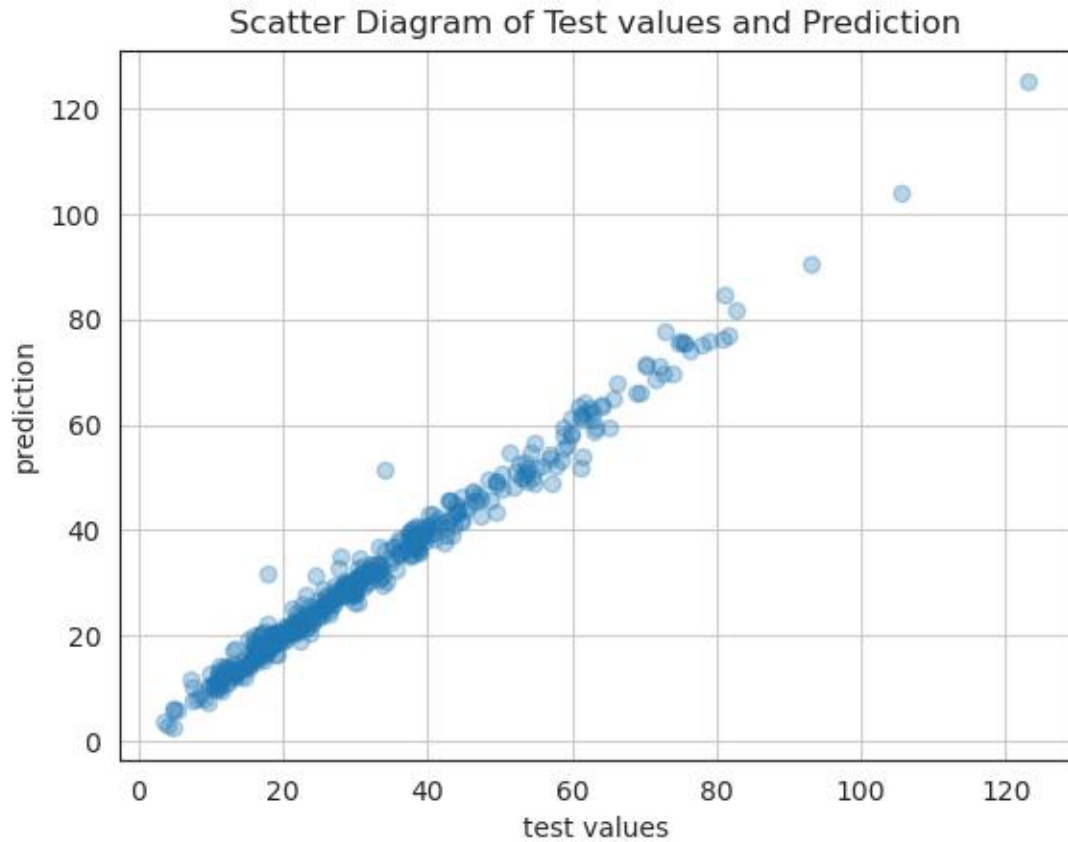
maladie respiratoire chronique



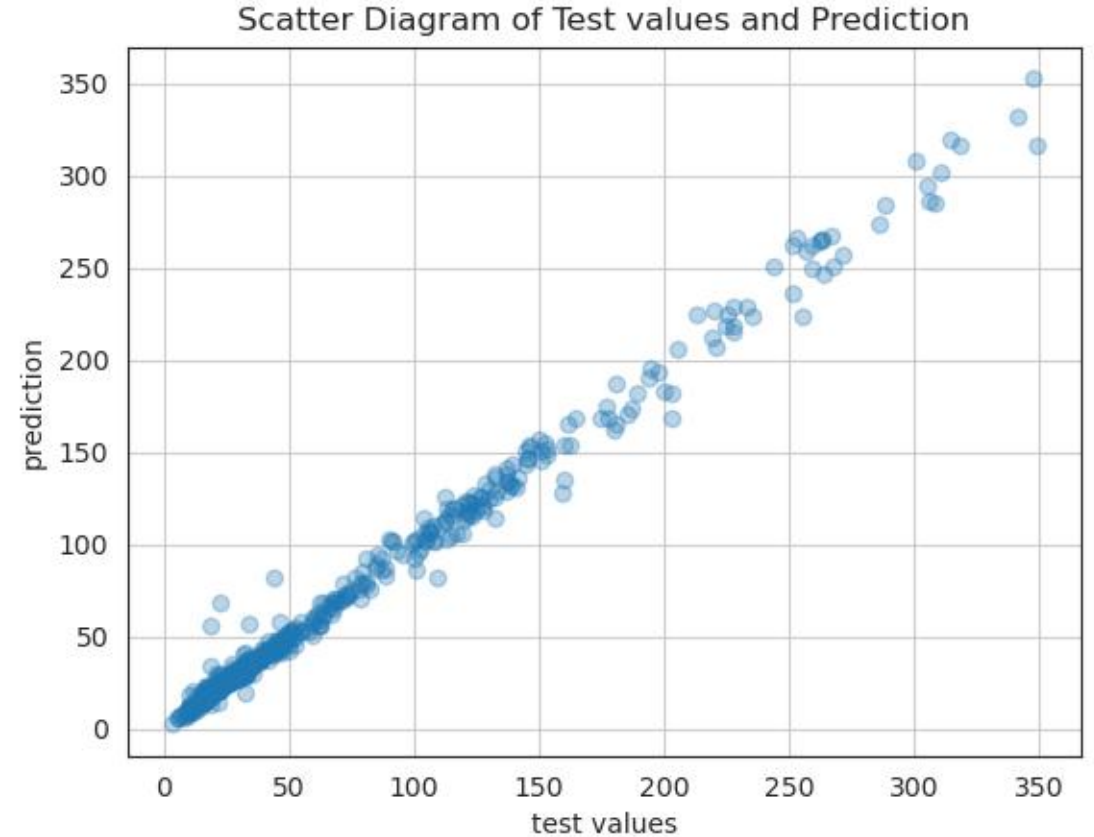
infections respiratoires



Partie 3: Machine Learning

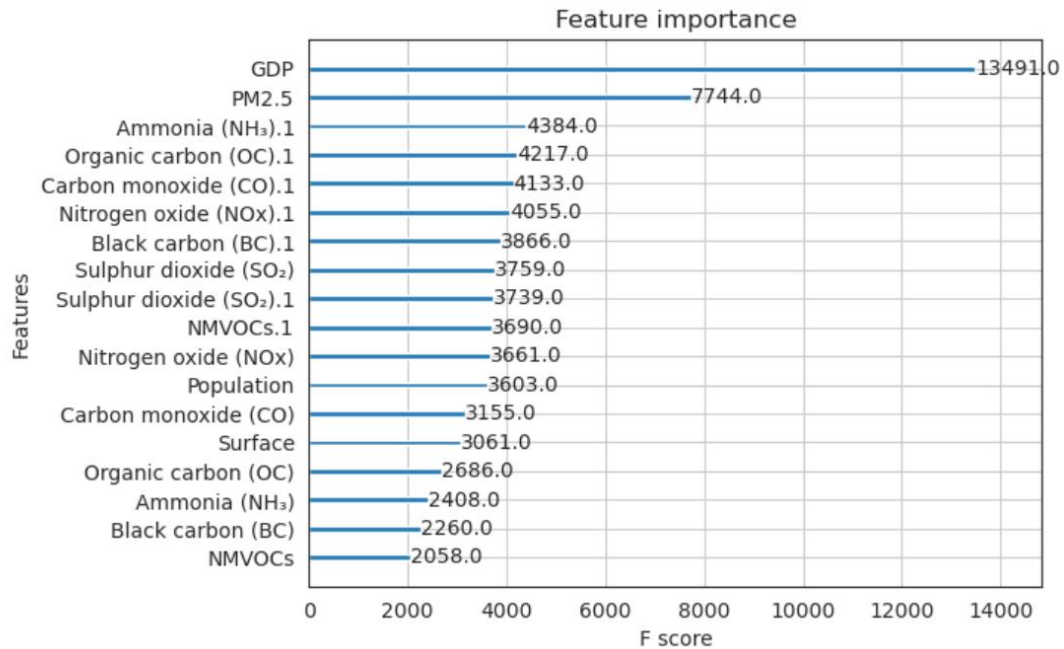


maladie respiratoire chronique

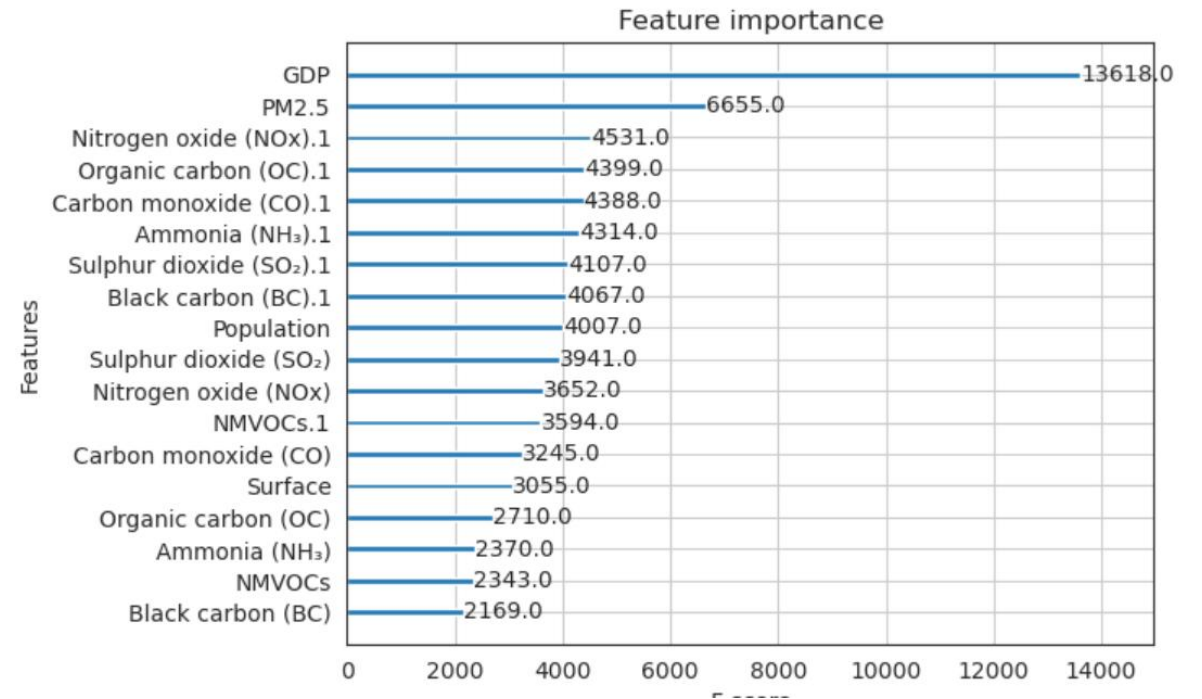


infections respiratoires

Partie 3: Machine Learning



maladie respiratoire chronique



infections respiratoires

