Exposure to air pollution and COVID-19 expansion in Western-Europe

17 July 2020

Mini-Project of the class
Statistical Signal Processing

(COM-500)

Tim Tuuva

Darja Nonaca

Summary

- Introduction
- Presentation of the PCA tool and test on synthetic data
- Results & Discussion
- PCA on Harvard dataset
- Conclusion

Introduction

- Many of the pre-existing conditions that increase the risk of death in those with COVID-19 are the same diseases that are affected by long-term exposure to air pollution

- Recent studies of Harvard University showed that an increase of 1µg/m3 in PM2.5 concentration, is associated with an 8% increase in the COVID-19 death rate in the United States

- In this work we investigated whether a similar correlation could be found in Western-Europe

Test of the PCA tool on the synthetic data (1)

Synthetic data generation

We distribute the particles 1 and 2 within the dataset according to Gaussian distribution as follows:

- c_{m1} (age) is i.i.d. in the interval [0-99]
- c_{m2} (particle 1) ~ $N(\mu_1, \sigma^2)$
- c_{m3} (particle 2) ~ $N(\mu_1, \mu_2, \sigma^2)$

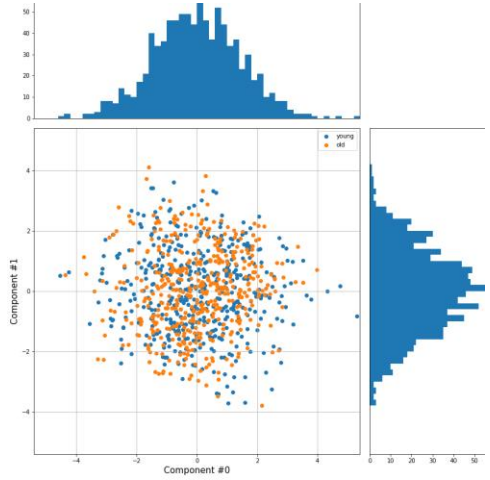
Then we correlate the diseases 1 and 2 to the first three features (c_{m1} , c_{m2} , c_{m3}) as follows:

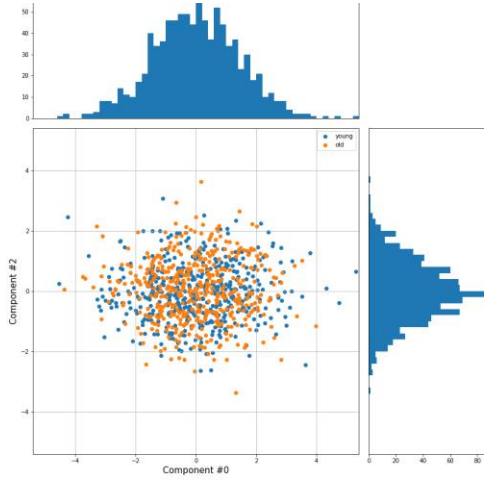
- c_{m4} (desease 1) = $\alpha\, c_{m2} + \beta\, c_{m1}$
- c_{m5} (desease 2) = $\gamma\, c_{m3}$

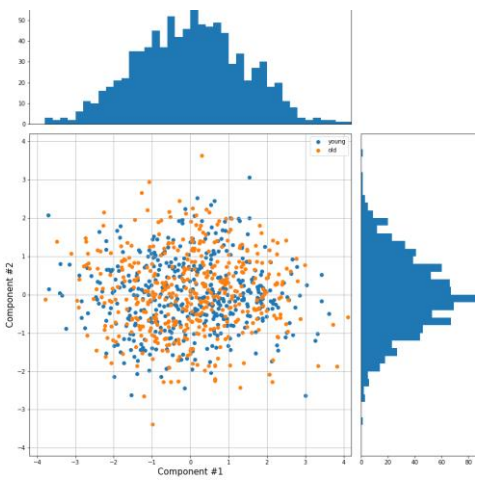
Singular values of the PCA

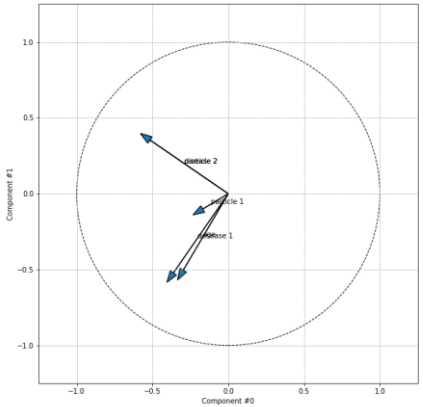
λ_1	λ_2	λ_3	λ_4	λ_5
4.26086984e+01	4.21239011e+01	3.00844774e+01	2.44462381e−14	9.54265059e−15

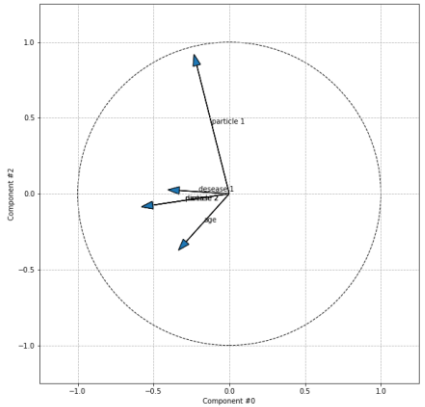
Test of the PCA tool on the synthetic data (2)

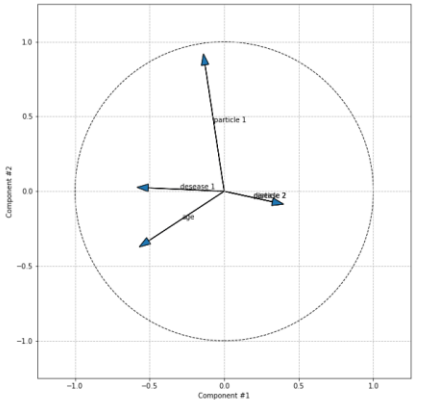












PCA on real data

Dataset

Information on the dataset	
France	Ile-de-France Grand Est Provence-Alpes-Côte d'Azur Auvergne-Rhône-Alpes
UK	All territory

Features

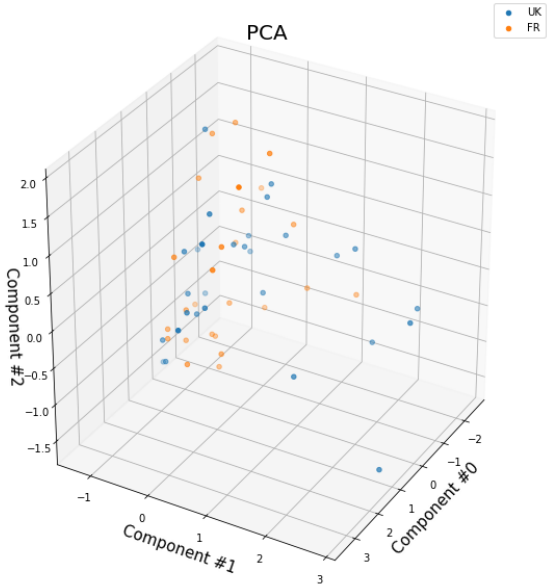
NO_2 $[\mu g/m^3]$	PM10 $[\mu g/m^3]$	PM2.5 $[\mu g/m^3]$	$covid\_death\_rate$/population

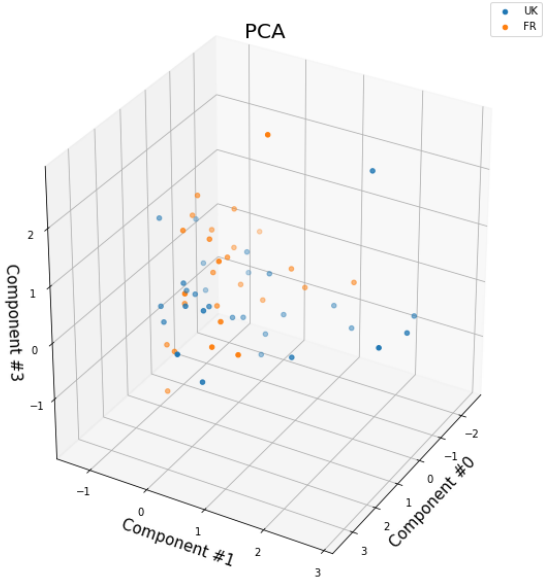
Cleaning procedure

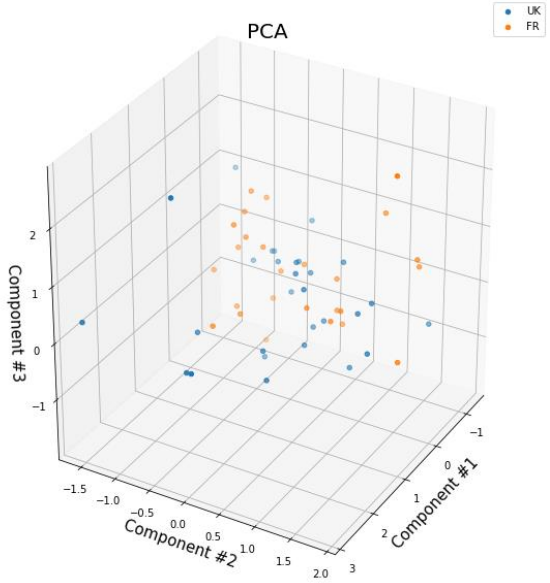
- The dataset distribution is centered
- Outliers are removed using the Interquartile range (IRQ)
- Final dataset is composed of 55 points

Results (1)

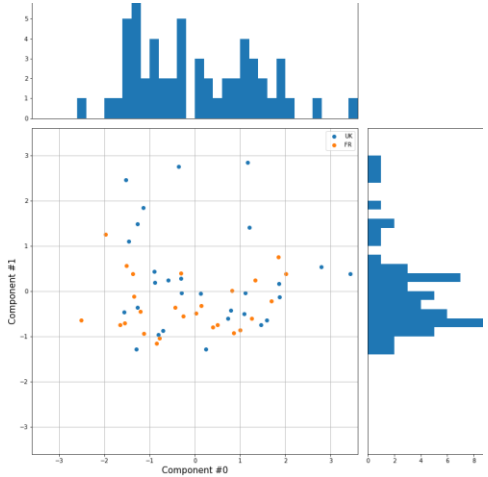
λ_1	λ_2	λ_3	λ_4
9.82809116	7.09382197	6.24462426	5.48552476

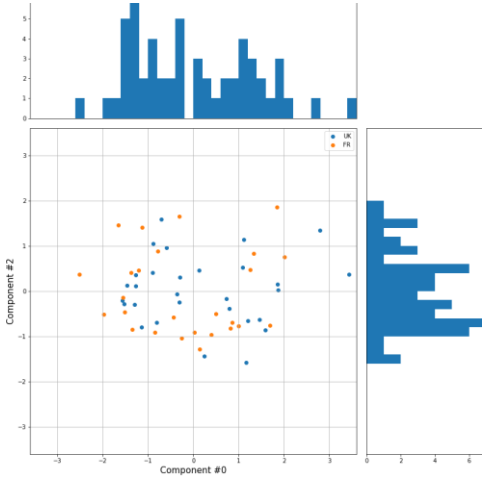


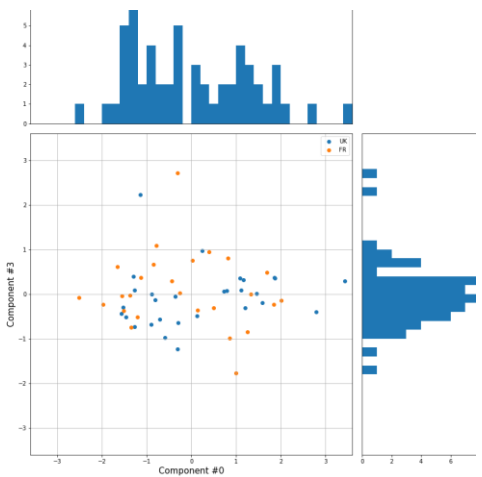


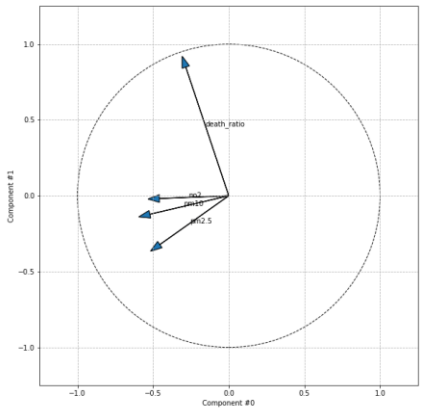


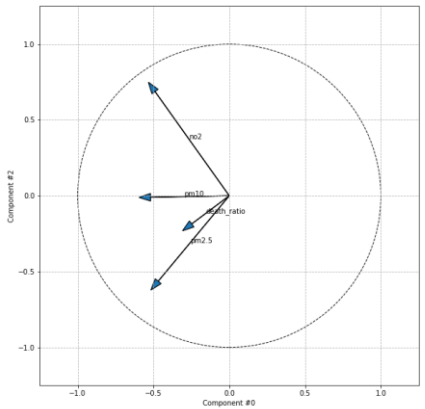
Results (2)

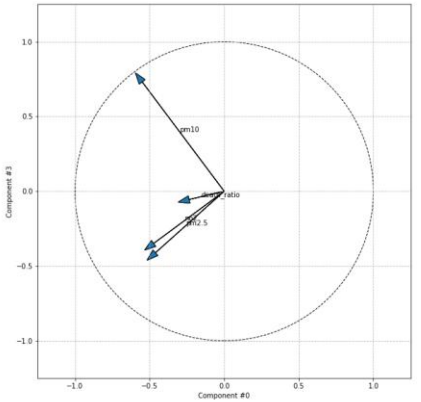

UK
FR
Component #1
Component #0


UK
FR
Component #2
Component #0


UK
FR
Component #3
Component #0


death_ratio
no2
pm10
pm2.5
Component #1
Component #0

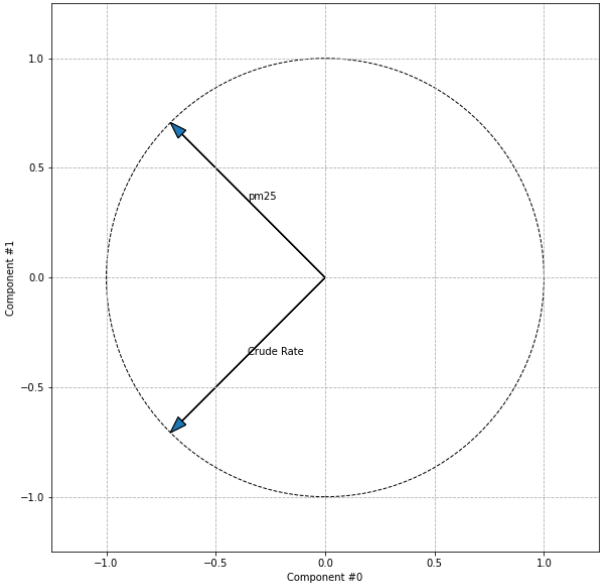

no2
pm10
death_ratio
pm2.5
Component #2
Component #0

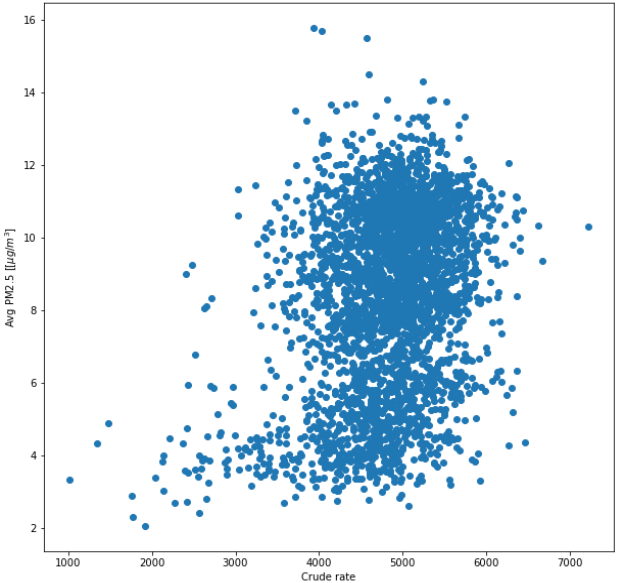

pm10
death_ratio
pm2.5
Component #3
Component #0

Harvard data set

County	Crude Rate	pm25
AL	4868.1	11.712587
AL	4266.9	10.077723
AL	5136.8	10.981967
AL	5219.9	11.998715
AL	4842.1	11.793023
...	...	...
WY	4657.3	4.896375
WY	2690.5	3.867683
WY	4357.0	5.088060
WY	4671.9	5.025529
WY	4605.7	3.424542

3096 rows × 2 columns





Conclusion

- Difficult to find the European information on the COVID-19 mortality per ”county”
- Furthermore, by performing our analysis on the data assembled by Harvard University, we found that PM2.5 and death rate are uncorrelated which is contradictory to their article
- This make us conclude that the way we do the correlation analysis should be rethought...