A statistical analysis on factors that contributed/slowed down the spread of COVID-19

Project for the exam: Machine learning, statistical learning, deep learning and artificial intelligence - Unsupervised Learning

Marzio De Corato

October 26, 2020

Overview

FINAL GOAL: Survey the statistical correlations between the COVID-19 cases/deaths and a selected set of attributes for different clusters via the principal component analysis

Cluster features

Provices (IT)

- Unemployment '19
- Private Transport '12
- Air Quality '19
- Public Transport '12
- Density '19
- Mean income '19

Regions (IT)

- Mortality
- Pop. for GP
- Mean Income '19
- Number of Visits for Pop. '17
- Public Transport
- Density '19
- Mean income '19
- Tests
- LEA '17
- Number of visit '17
- Public structures '17

Countries

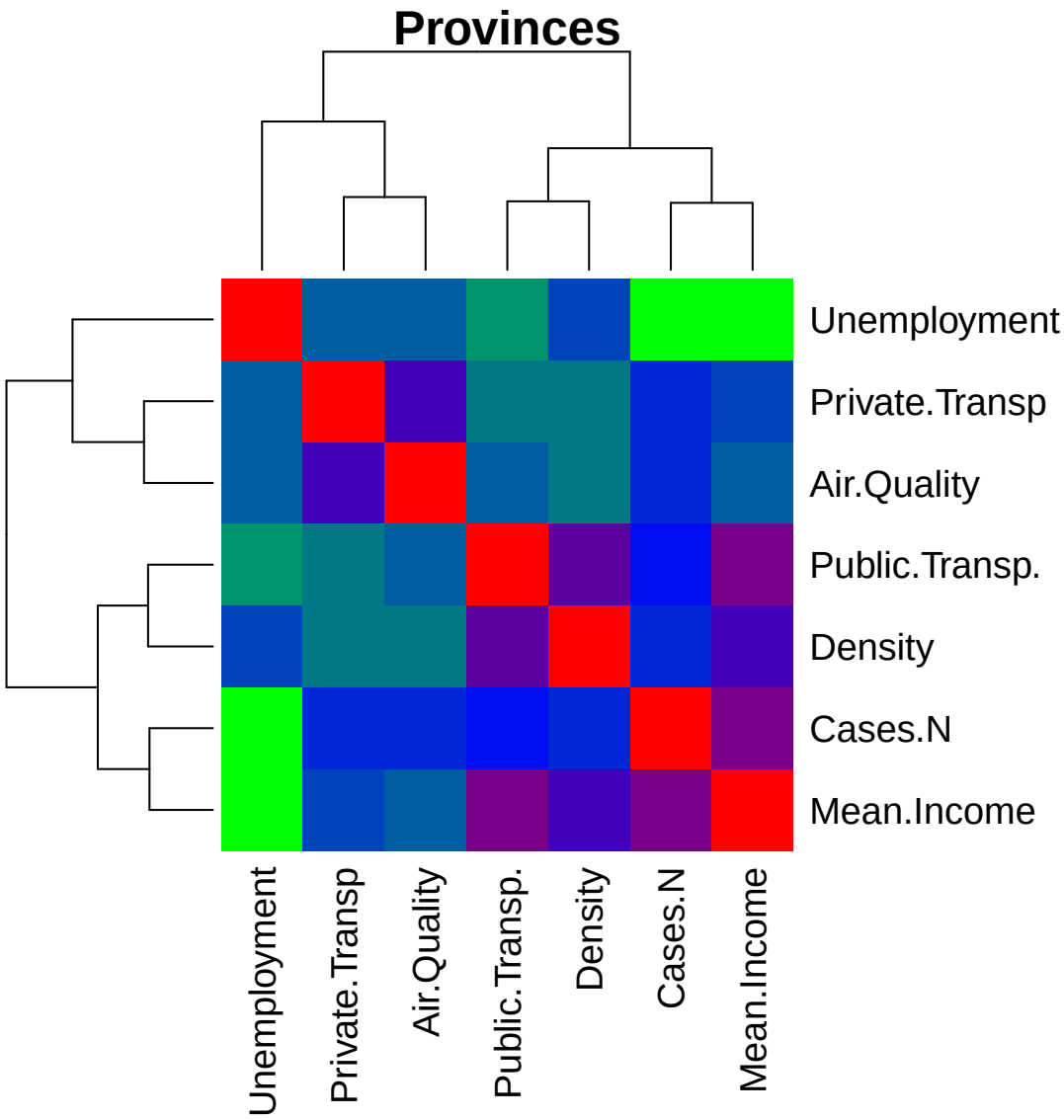
- $PM_{2.5}$ '16
- Traffic Mortality '16
- Pollution Mortality '16
- GDP pro-capita '19
- Health expenditure '17
- UHC '17

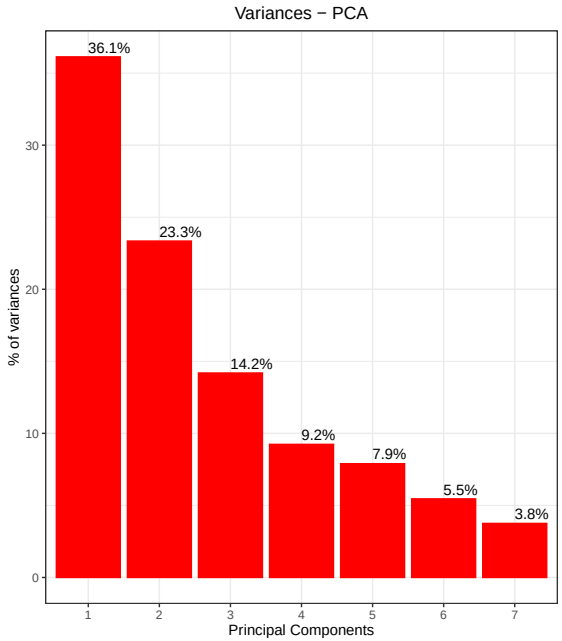
Datasets features

TEMPORAL INTERVAL: Provinces up to 24/08/2020; Regions up to 24/08/2020; Countries up to 27/08/2020

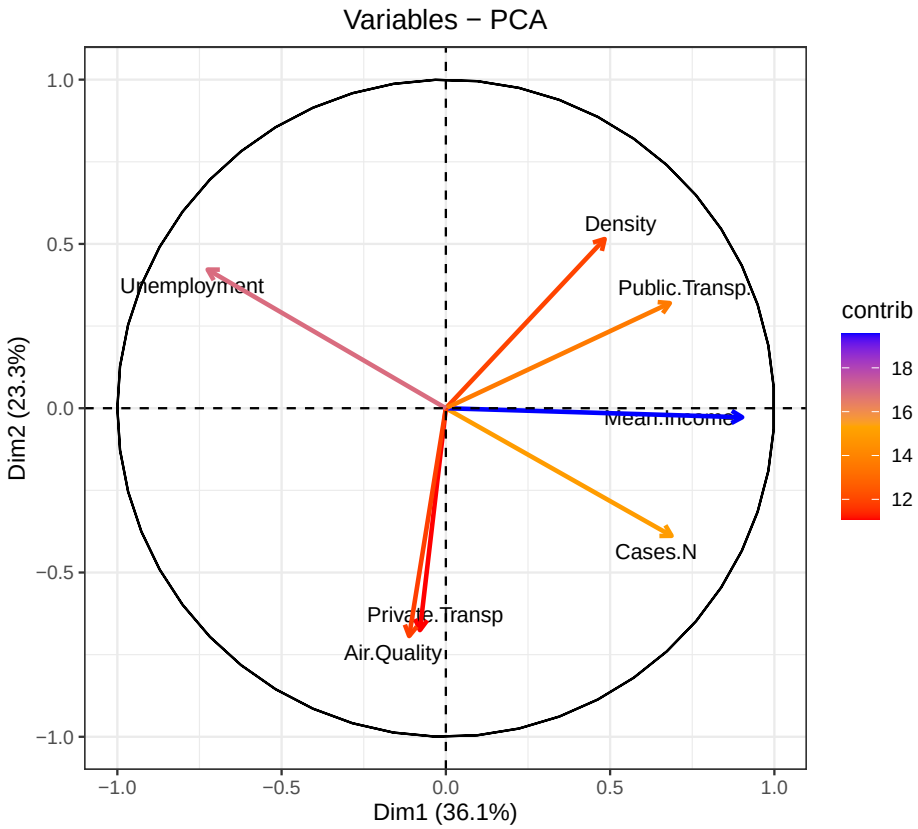
SOURCES: Protezione Civile, Istituto Italiano di Statistica, Ministro delle Finanze, Ministro della Salute, World Health Organization and World Bank

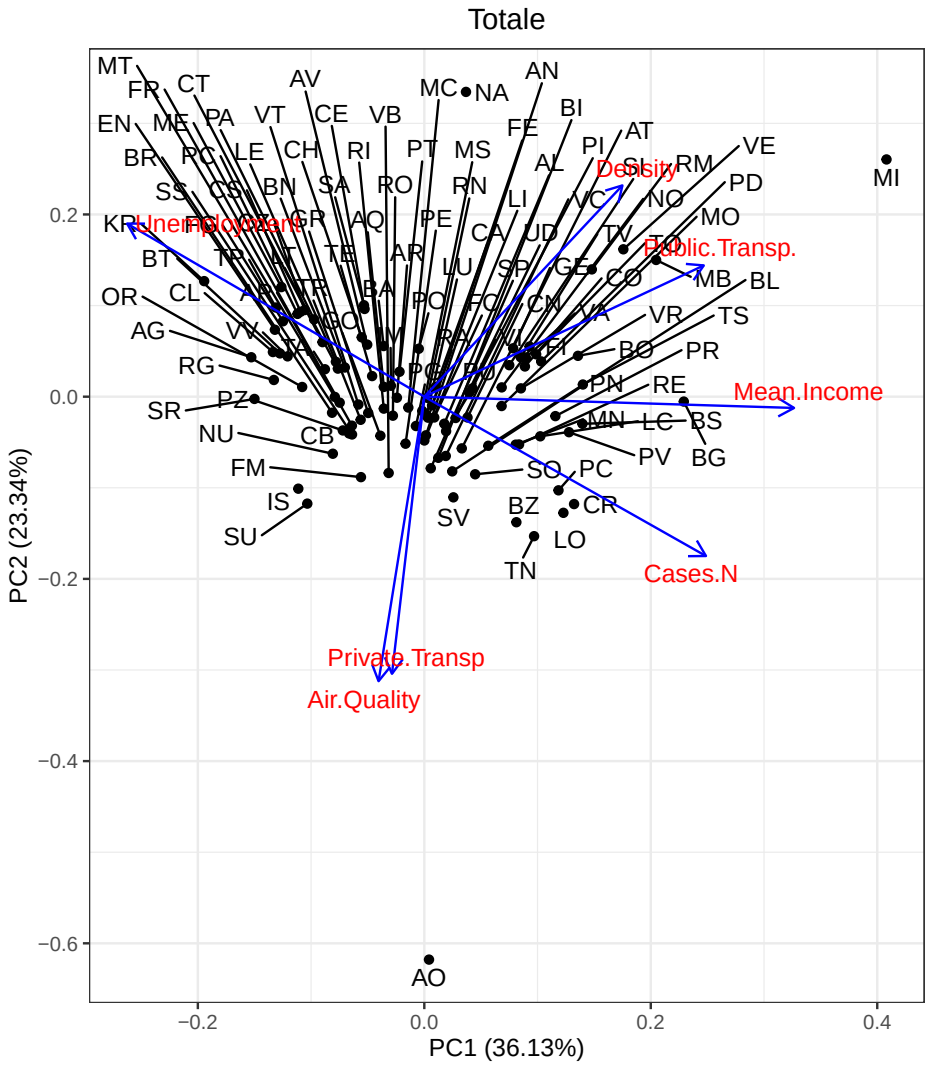
Provinces


Provinces
Unemployment
Private.Transp
Air.Quality
Public.Transp.
Density
Cases.N
Mean.Income

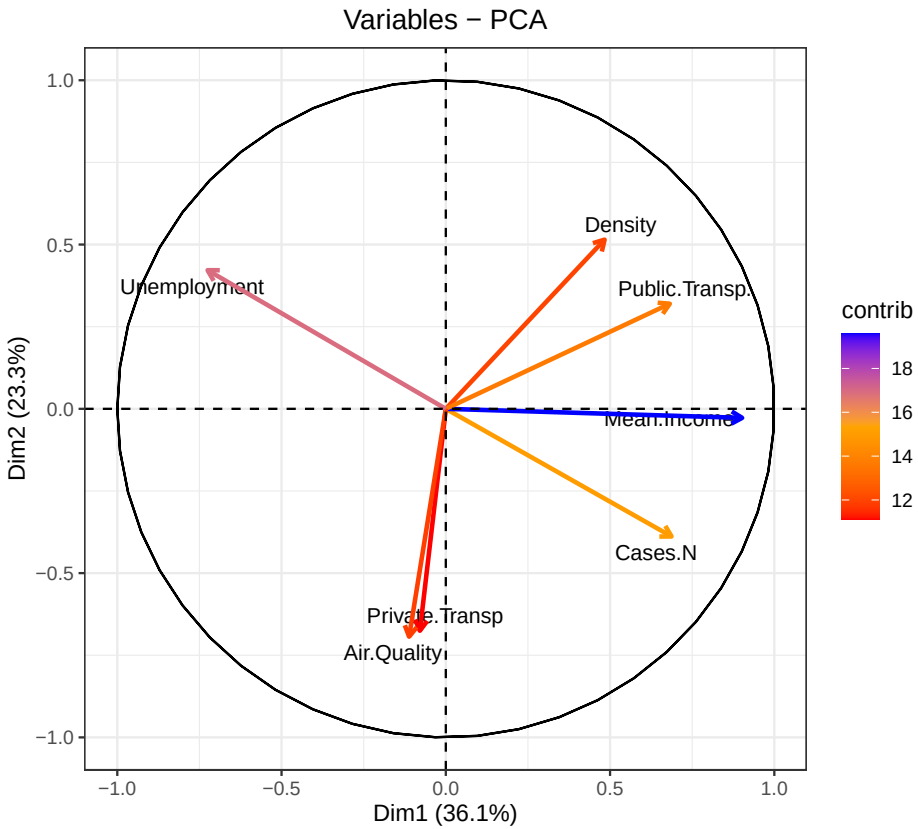

Variances − PCA
36.1%
23.3%
14.2%
9.2%
7.9%
5.5%
3.8%
% of variances
Principal Components

Provinces


Variables − PCA
Density
Unemployment
Public.Transp.
Mean.Income
Cases.N
Private.Transp
Air.Quality
contrib
18
16
14
12
Dim2 (23.3%)
Dim1 (36.1%)
1.0
0.5
0.0
−0.5
−1.0

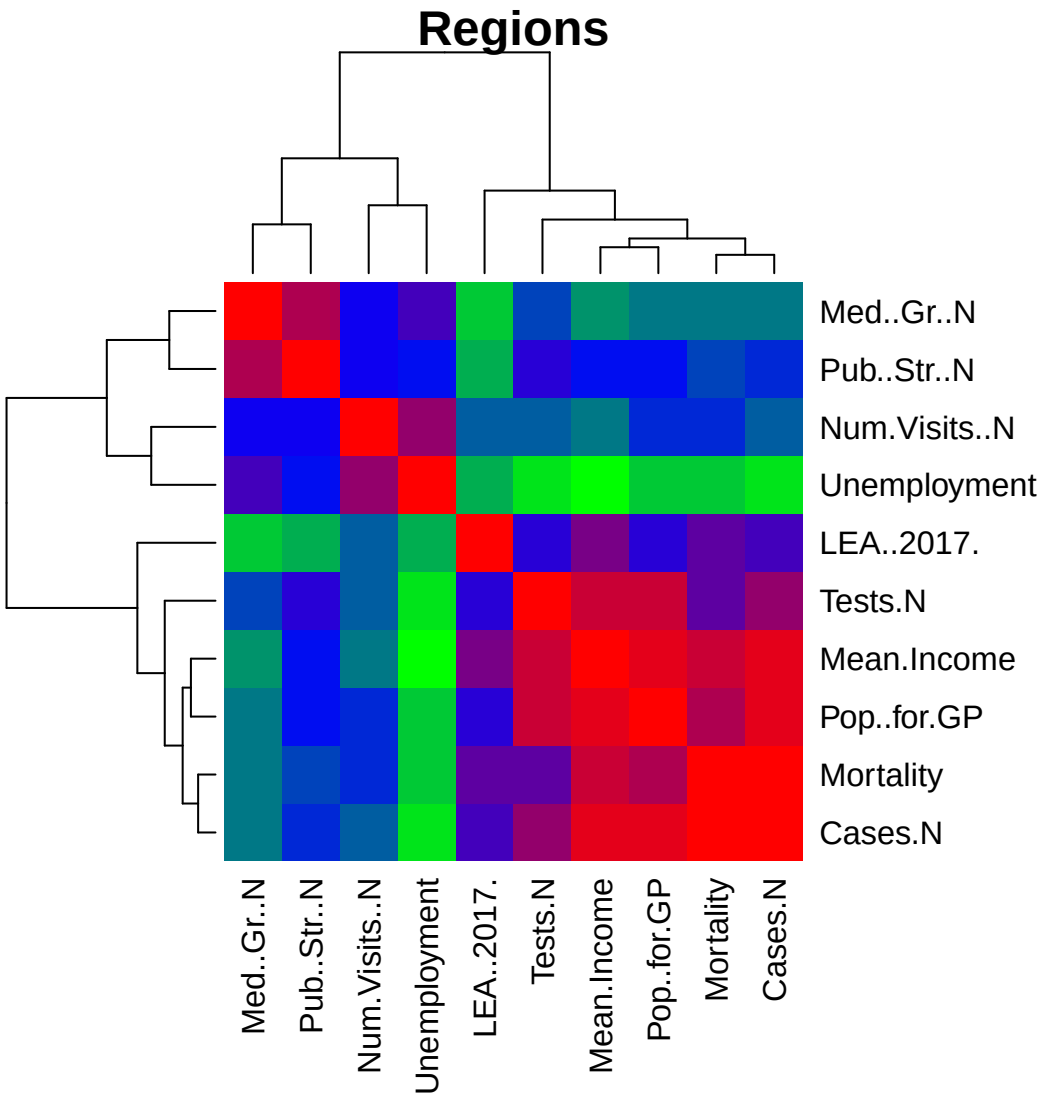

Totale
Density
Public.Transp.
Mean.Income
Cases.N
Private.Transp
Air.Quality
MI
AO
PC2 (23.34%)
PC1 (36.13%)
0.2
0.0
−0.2
−0.4
−0.6
0.4

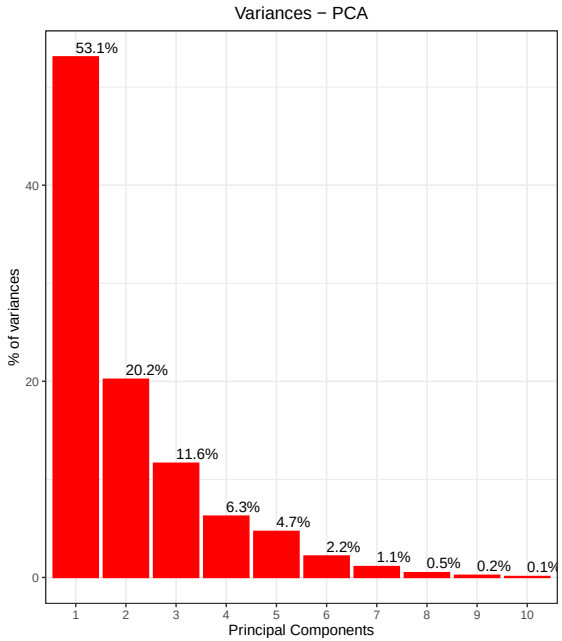
Provinces



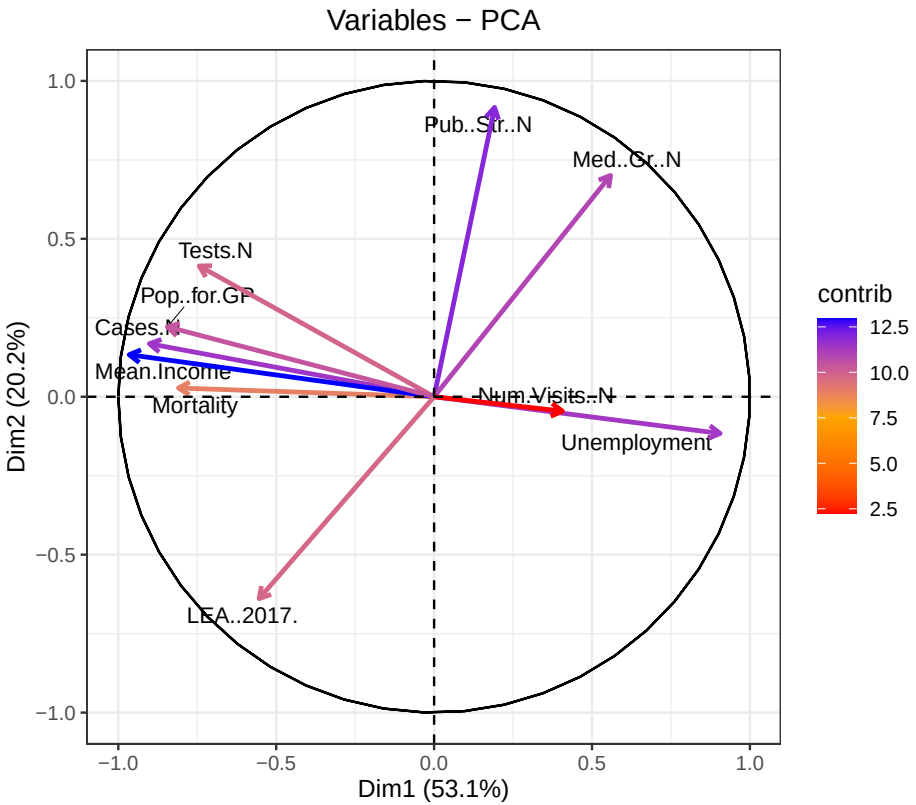
- The public transport, the density, the normalized cumulative cases and the mean income are positively correlated
- This can be explained by the fact that a higher density, a higher public transport demand and a higher income increase the rate of contact between the individuals
- Rich people can spend more money for social events or perhaps in to travels thus they increase their connectivity (the opposite happens for the unemployment)

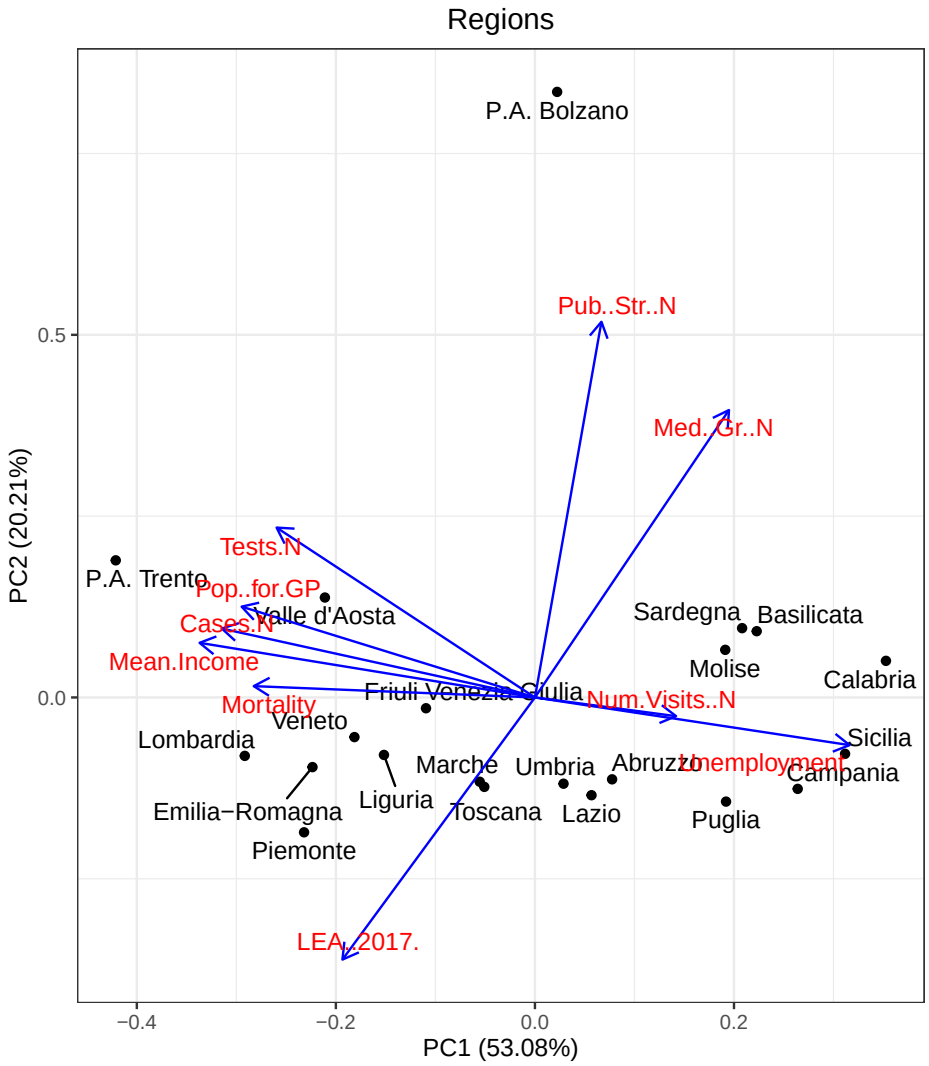
Regions


Regions
Med..Gr..N
Pub..Str..N
Num.Visits..N
Unemployment
LEA..2017.
Tests.N
Mean.Income
Pop..for.GP
Mortality
Cases.N

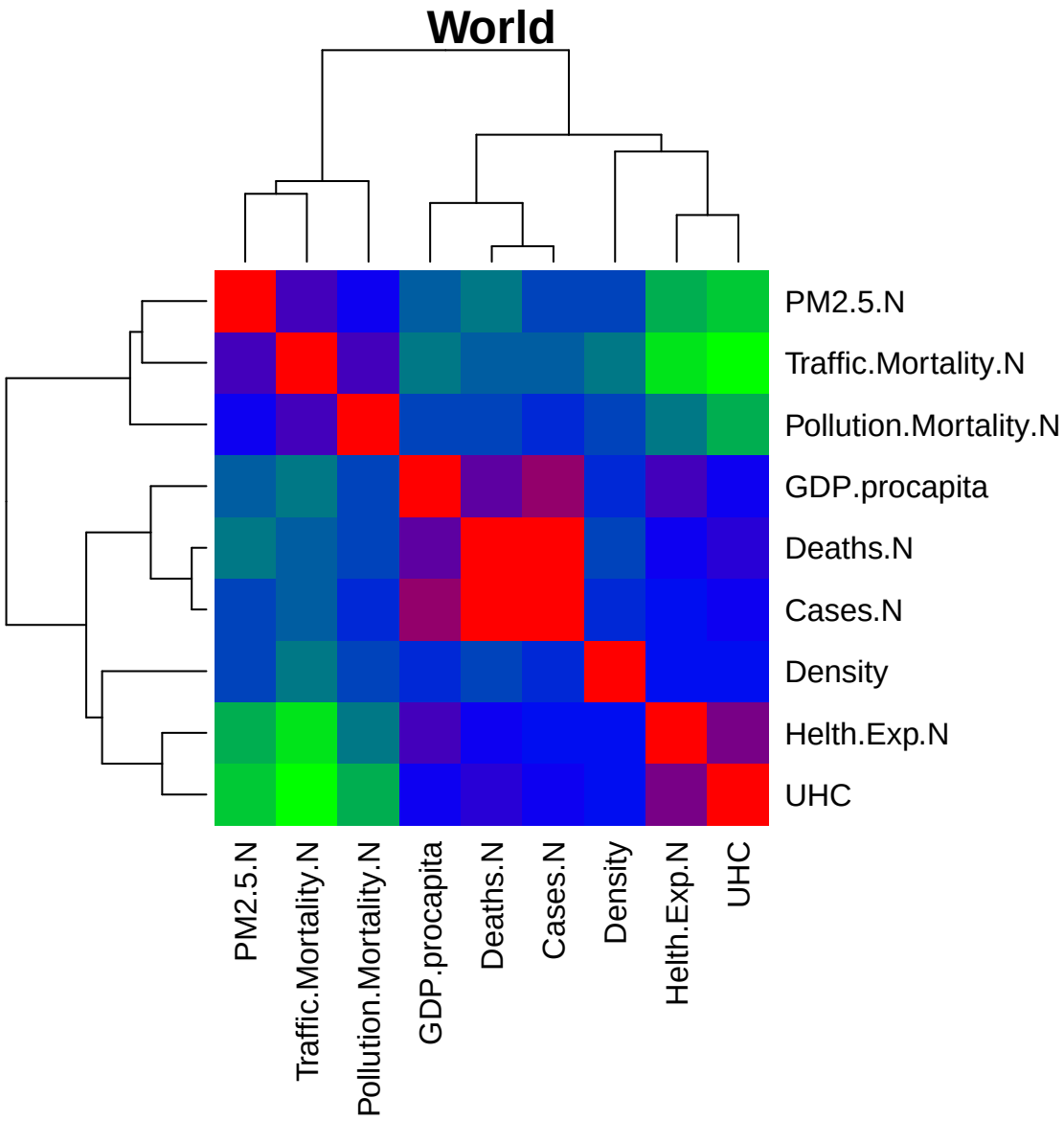

Variances − PCA
% of variances
Principal Components
53.1%
20.2%
11.6%
6.3%
4.7%
2.2%
1.1%
0.5%
0.2%
0.1%

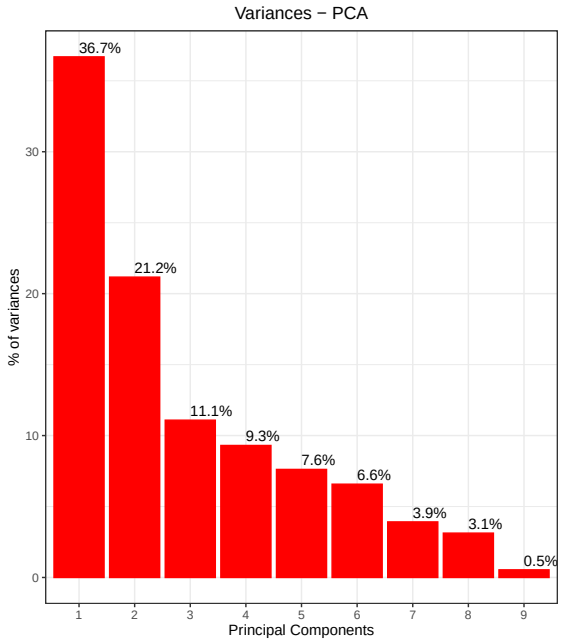
Regions


Variables − PCA
Pub..Str..N
Med..Gr..N
Tests.N
Pop..for.GP
Cases.N
Mean.Income
Mortality
Num.Visits..N
Unemployment
LEA..2017.
contrib
12.5
10.0
7.5
5.0
2.5
Dim2 (20.2%)
Dim1 (53.1%)
1.0
0.5
0.0
−0.5
−1.0

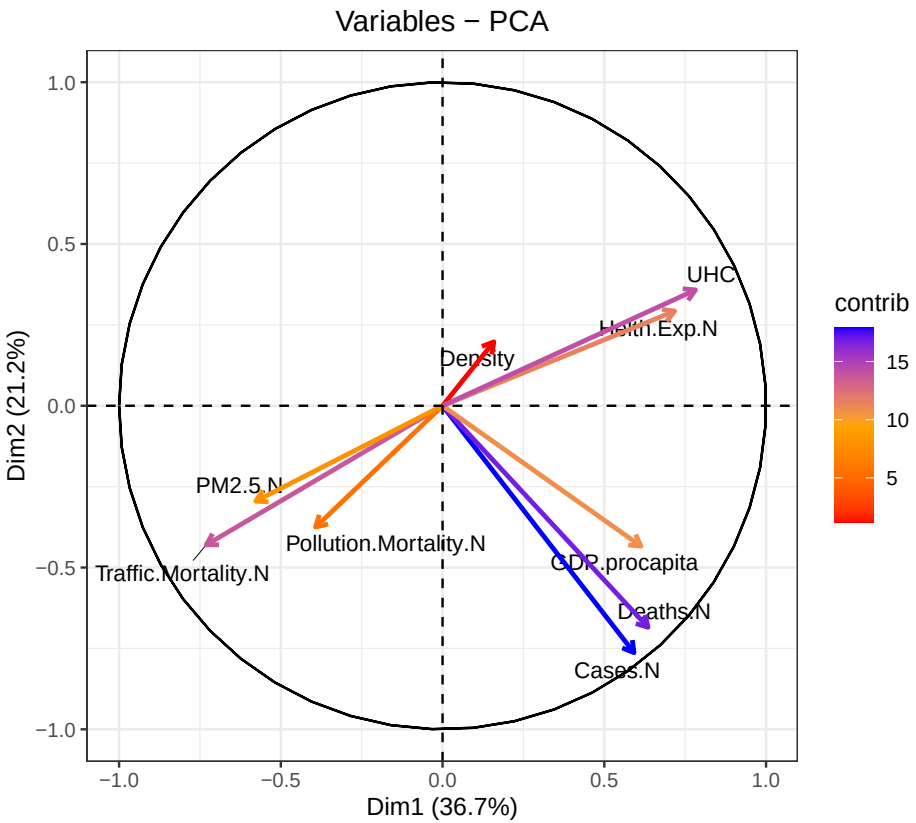

Regions
P.A. Bolzano
Pub..Str..N
Med..Gr..N
Tests.N
P.A. Trento
Pop..for.GP
Valle d'Aosta
Cases.N
Sardegna
Basilicata
Mean.Income
Molise
Calabria
Friuli Venezia Giulia
Num.Visits..N
Mortality
Veneto
Sicilia
Lombardia
Marche
Umbria
Abruzzo
Unemployment
Campania
Liguria
Toscana
Lazio
Emilia−Romagna
Puglia
Piemonte
LEA..2017.
PC2 (20.21%)
PC1 (53.08%)
0.5
0.0
−0.4
−0.2
0.0
0.2

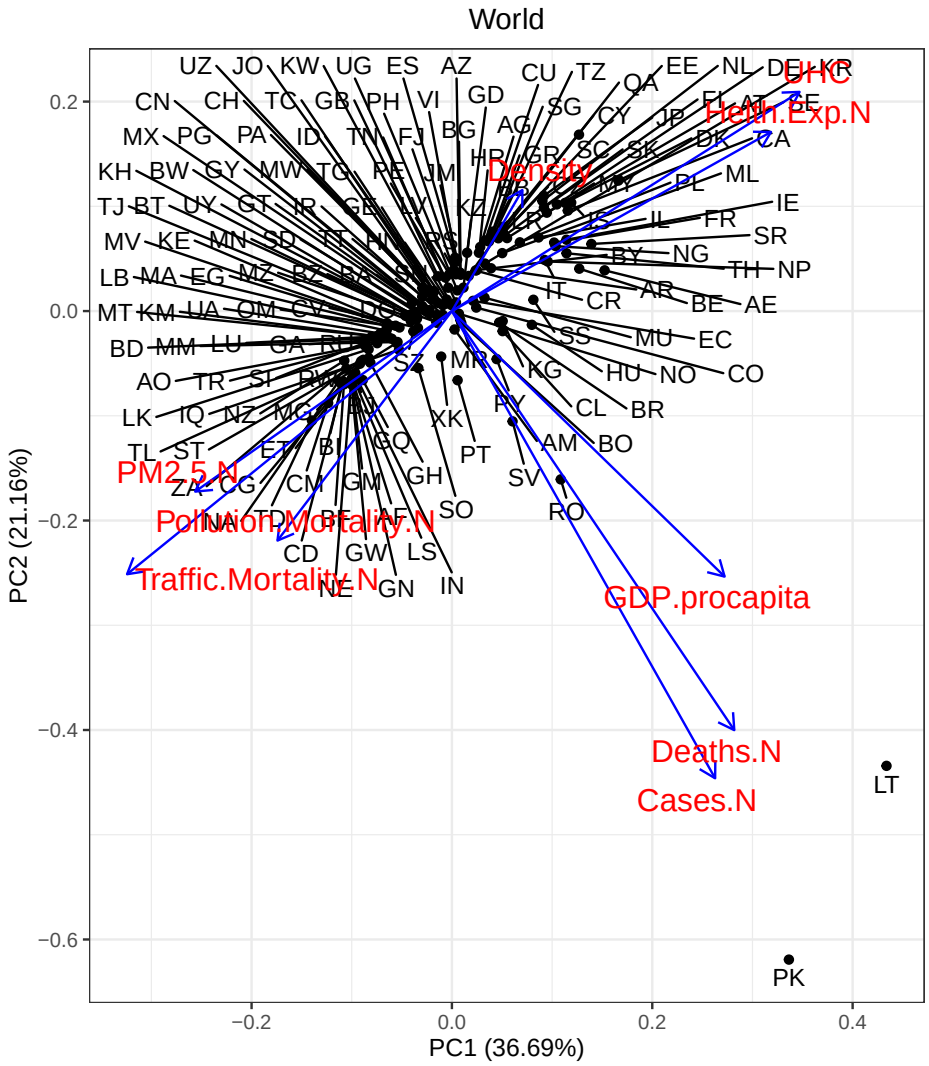
Countries


World
PM2.5.N
Traffic.Mortality.N
Pollution.Mortality.N
GDP.procapita
Deaths.N
Cases.N
Density
Helth.Exp.N
UHC


Variances − PCA
% of variances
Principal Components
36.7%
21.2%
11.1%
9.3%
7.6%
6.6%
3.9%
3.1%
0.5%

Countries


Variables − PCA
Dim2 (21.2%)
Dim1 (36.7%)
contrib
15
10
5
UHC
Health.Exp.N
Density
PM2.5.N
Pollution.Mortality.N
Traffic.Mortality.N
GDP.procapita
Deaths.N
Cases.N


World
PC2 (21.16%)
PC1 (36.69%)
UHC
Health.Exp.N
Density
PM2.5.N
Pollution.Mortality.N
Traffic.Mortality.N
GDP.procapita
Deaths.N
Cases.N
LT
PK