

UNIVERSITÉ VIRTUELLE DU SÉNÉGAL
PÔLE SCIENCE, TECHNOLOGIE ET NUMÉRIQUE
MASTER CALCUL SCIENTIFIQUE

PROJET

STATISTIQUE MULTIVARIEE

Professeur : Dr. Seydou Nourou SYLLA
Specialiste en Big Data et Maching Learning

Mamadou Korka

Diallo

Description du projet

Le but du projet consiste à faire un modèle ACP (Analyse Composante Principale) et un modèle AFC (Analyse Composante Factorielle) d'une base de données c'est à dire un dataset. Nous avons choisie un dataset de la pollution de l'air atmosphérique mondiale.

La pollution atmosphérique fait l'objet d'un grand débat scientifique car la pollution de l'air est la contamination de l'environnement intérieur ou extérieur par tout agent chimique, physique ou biologique qui modifie les caractéristiques naturelles de l'atmosphère.

Les appareils à combustion domestiques, les véhicules à moteur, les installations industrielles et les incendies de forêt sont des sources courantes de pollution de l'air. Les polluants les plus préoccupants pour la santé publique comprennent les particules, le monoxyde de carbone, l'ozone, le dioxyde d'azote et le dioxyde de soufre. La pollution de l'air extérieur et intérieur provoque des maladies respiratoires et autres et est une source importante de morbidité et de mortalité.

Notre jeu de données fournit des informations géolocalisées sur les polluants suivants :

- Dioxyde d'azote [NO₂] : Le dioxyde d'azote est l'un des nombreux oxydes d'azote. Il est introduit dans l'air par des phénomènes naturels comme l'entrée de la stratosphère ou l'éclairage. Au niveau de la surface, cependant, le NO₂ se forme à partir des émissions des voitures, des camions et des autobus, des centrales électriques et des équipements hors route. Une exposition sur de courtes périodes peut aggraver les maladies respiratoires, comme l'asthme. Des expositions plus longues peuvent contribuer au développement de l'asthme et des infections respiratoires. Les personnes asthmatiques, les enfants et les personnes âgées sont plus à risque de subir les effets du NO₂ sur la santé.
- Ozone [O₃] : La molécule d'ozone est nocive pour la qualité de l'air extérieur (si en dehors de la couche d'ozone). Au niveau de la surface, l'ozone est créé par des réactions chimiques entre les oxydes d'azote et les composés organiques volatils (COV). Contrairement au bon ozone situé dans la haute atmosphère, l'ozone au niveau du sol peut provoquer plusieurs problèmes de santé comme des douleurs à la poitrine, de la toux, une irritation de la gorge et une inflammation des voies respiratoires. De plus, il peut réduire la fonction pulmonaire et aggraver la bronchite, l'emphysème et l'asthme. L'ozone affecte également la végétation et les écosystèmes. En particulier, il endommage la végétation sensible pendant la saison de croissance.
- Monoxyde de carbone [CO] : Le monoxyde de carbone est un gaz incolore et inodore. À l'extérieur, il est surtout émis dans l'air par les voitures, camions et autres véhicules ou engins qui brûlent des combustibles fossiles. Des articles tels que les appareils de chauffage au kérosène et au gaz, les cuisinières à gaz libèrent également du CO affectant la qualité de l'air intérieur. Respirer de l'air avec une forte concentration de CO réduit la quantité d'oxygène qui peut être transportée dans le sang vers des organes critiques comme le cœur et le cerveau. À des niveaux très élevés, qui ne sont pas susceptibles de se produire à l'extérieur mais qui sont possibles dans des environnements clos. Le CO peut causer des étourdissements, de la confusion, une perte de conscience et la mort.
- Particules [PM_{2.5}] : Les particules atmosphériques, également connues sous le nom de particules d'aérosols atmosphériques, sont des mélanges complexes de petites matières solides et liquides qui pénètrent dans l'air. S'ils sont inhalés, ils peuvent causer de graves problèmes cardiaques et pulmonaires. Ils ont été classés cancérogènes de groupe 1 par l'Agence Internationale de Recherche sur le Cancer (IARC). PM₁₀ fait référence aux particules d'un diamètre de 10 micromètres ou moins. PM_{2.5} fait référence aux particules d'un diamètre

de 2,5 micromètres ou moins.

L'ensemble de données sur la pollution atmosphérique mondiale est :

- Pays : Nom du pays
- Ville : Nom de la ville
- Valeur IQA : Valeur IQA globale de la ville
- Catégorie AQI : Catégorie globale AQI de la ville
- Valeur CO AQI : valeur AQI du monoxyde de carbone de la ville
- Catégorie CO AQI : Catégorie AQI du monoxyde de carbone de la ville
- Valeur AQI de l'ozone : Valeur IQA de l'ozone de la ville
- Catégorie Ozone AQI : Catégorie AQI de l'ozone de la ville
- Valeur IQA NO2 : valeur IQA du dioxyde d'azote de la ville
- Catégorie NO2 AQI : Catégorie AQI du dioxyde d'azote de la ville
- PM2.5 Valeur AQI : valeur AQI des particules d'un diamètre de 2,5 micromètres ou moins de la ville .

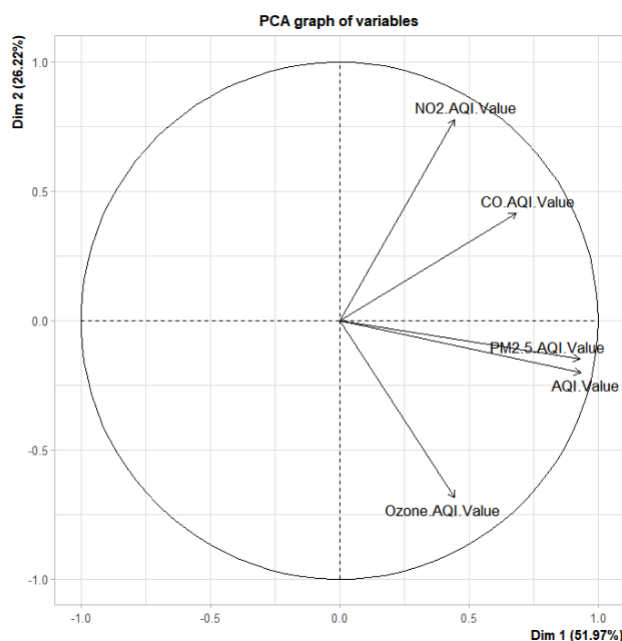
NB :Notre dataset comporte 12 variables et 23463 observations dont 5 variables sont quantitatives et 7 sont qualitatives.

Le Modèle ACP

Le Modèle ACP(Analyse Composante Principale) est une methode ou une technique d'analyse factorielle qui permet de mettre en évidence et décrire des sociations entre deux ou plusieurs variables quantitatives.

De ce fait, notre dataset a cinq(5) variables quantitatives. Donc notre ACP va se concentrer sur les variables quantitaves qui sont : Catégorie PM2.5 AQI,Catégorie NO2 AQI, Valeur AQI de l'ozone,Valeur CO AQI,Valeur CO AQI.

- Observons notre modèle ACP :



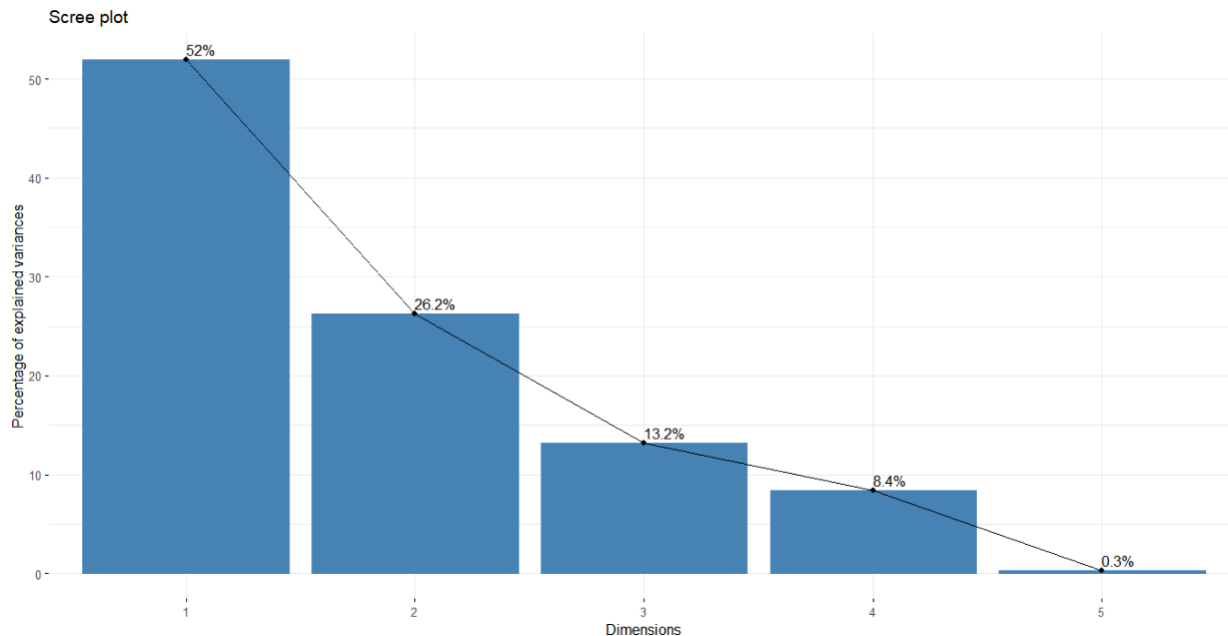
Le graphe nous montre que toutes nos variables sont centrées à droite par rapport à l'axe 1 .

```

> # Recuperer les valeurs Propres et leurs Pourcentages
> model_ACP$eig
      eigenvalue percentage of variance cumulative percentage of variance
comp 1 2.59858659      51.9717318      51.97173
comp 2 1.31100182      26.2200365      78.19177
comp 3 0.65877715      13.1755430      91.36731
comp 4 0.41852976       8.3705952      99.73791
comp 5 0.01310467       0.2620935     100.00000

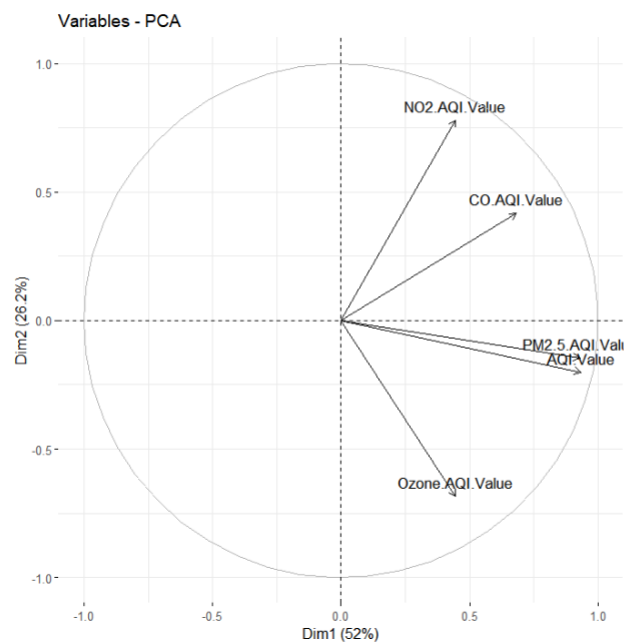
```

on peut voir aussi que la première composante a une valeur propre plus élevée (2.59) que celle des autres suivie de la deuxième composante (1.31) ainsi de suite.



l'histogramme des pourcentages des valeurs propres nous montre que la Catégorie AQI du dioxyde d'azote a une pourcentage plus élevée suivie la Catégorie AQI du monoxyde de carbone ainsi de suite.

— L'interprétation des variables par rapport aux axes.

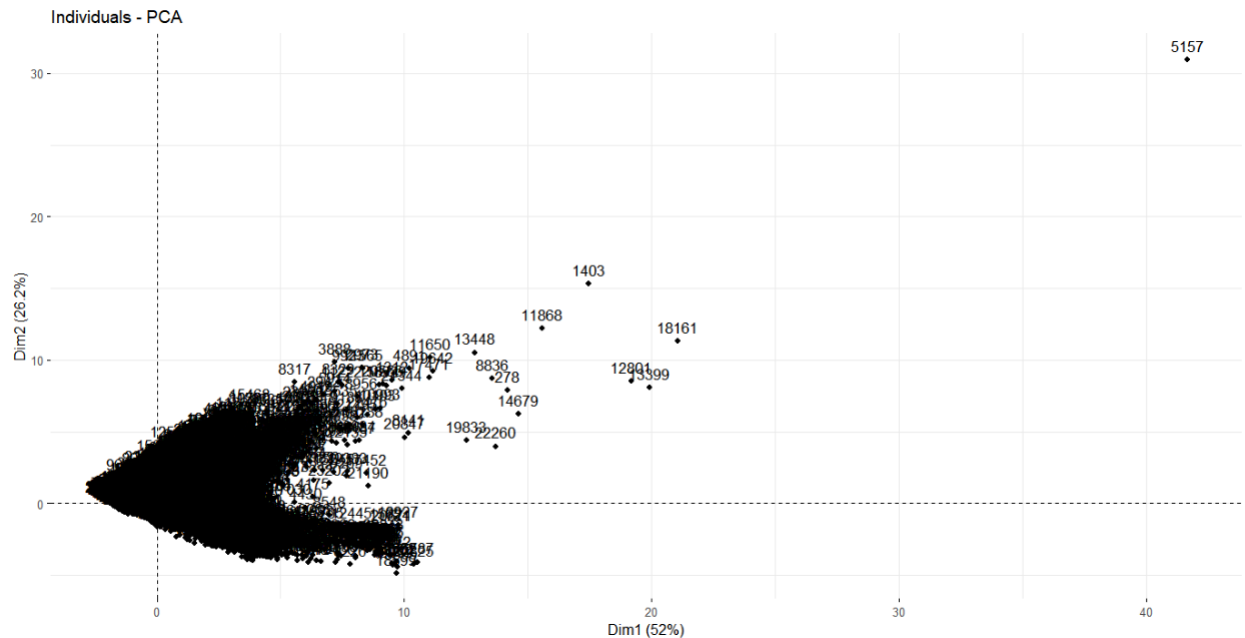


Le graphe du cercle des corrélations nous montre que les variables sont corrélées à l'axe 1 au facteur 1.

On a une corrélation qui est positive pour toutes les variables.

On a même des corrélations qui sont très proches aux coordonnées de l'axe 1 (PM2.5 Valeur AQI, Valeur AQI) cela veut dire qu'ils créent une pollution de l'air faible.

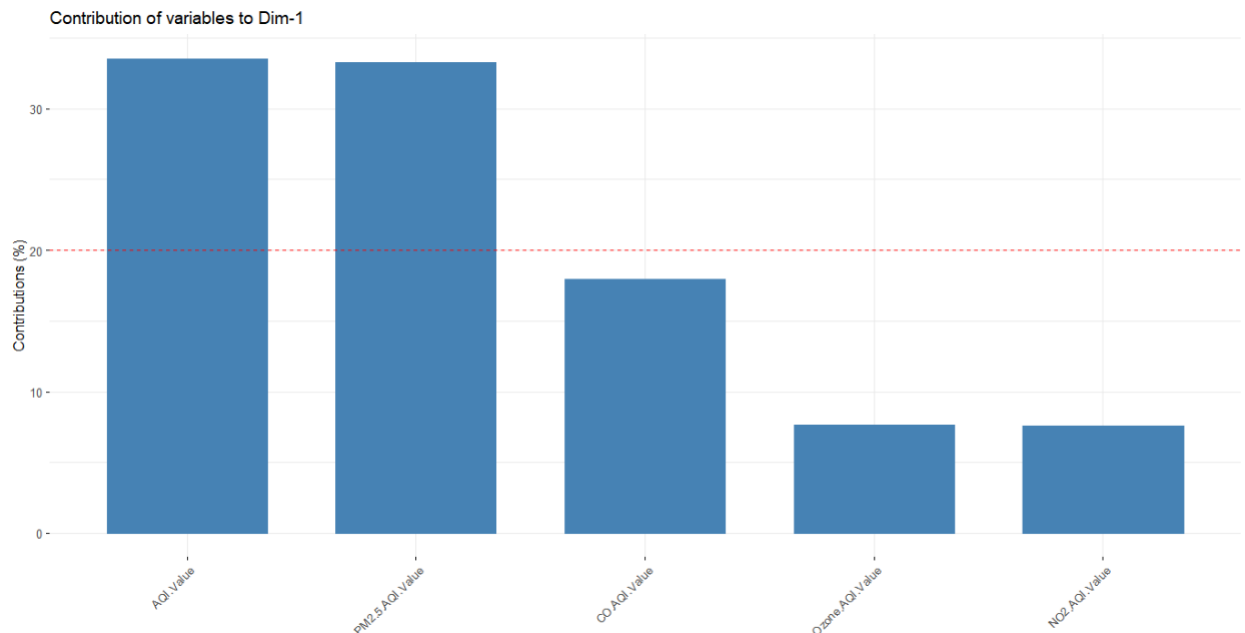
Les villes qui sont à droite avec une forte coordonnée par rapport à l'axe 1 dégagent beaucoup de NO2 AQI et Valeur CO AQI et causent une forte pollution de l'air atmosphérique. Les villes qui sont en gauche avec une faible coordonnée par rapport à l'axe 1 dégagent de OzoneAQS.value et causent moyen de pollution de l'air. L'axe 1 a toutes les variables et à droite du graphe l'on a toutes les villes qui entraînent une forte pollution de l'air.



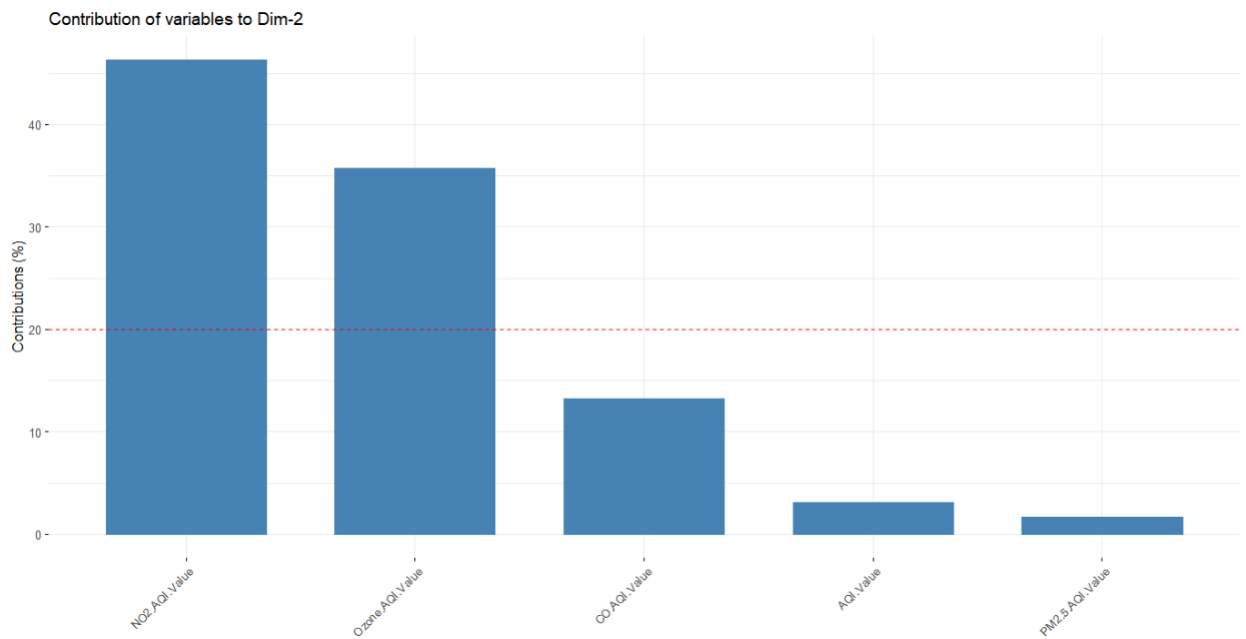
Le graphique nous confirme que toutes les villes qui polluent se trouvent à droite.

```
# % Voyons la contributions de chaque variable
> model_ACP$var$contrib
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
AQI.Value	33.554240	3.096501	12.2502866	0.01736121	51.081610781
CO.AQI.Value	17.944221	13.191071	30.001717	38.66312703	0.001409757
Ozone.AQI.Value	7.653797	35.723066	40.3166608	16.02803237	0.280243612
PM2.5.AQI.Value	7.599280	46.34462	0.9519757	45.11240460	0.001877665
PM2.5.AQI.Value	33.244862	1.654900	16.2809051	0.1862748	63.0358185



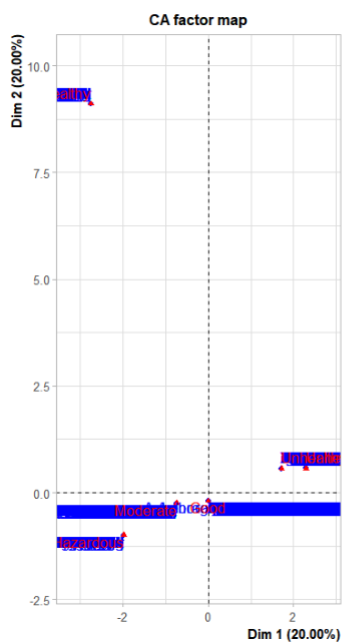
on voit que la variable AQI.Value a une contribution plus forte dans les toutes dimensions 1,2,3, etc... suivie de PM2.5 Valeur AQI,Ozone.AQI.Value,NO2.AQI.Value,PM2.5.AQI.Value

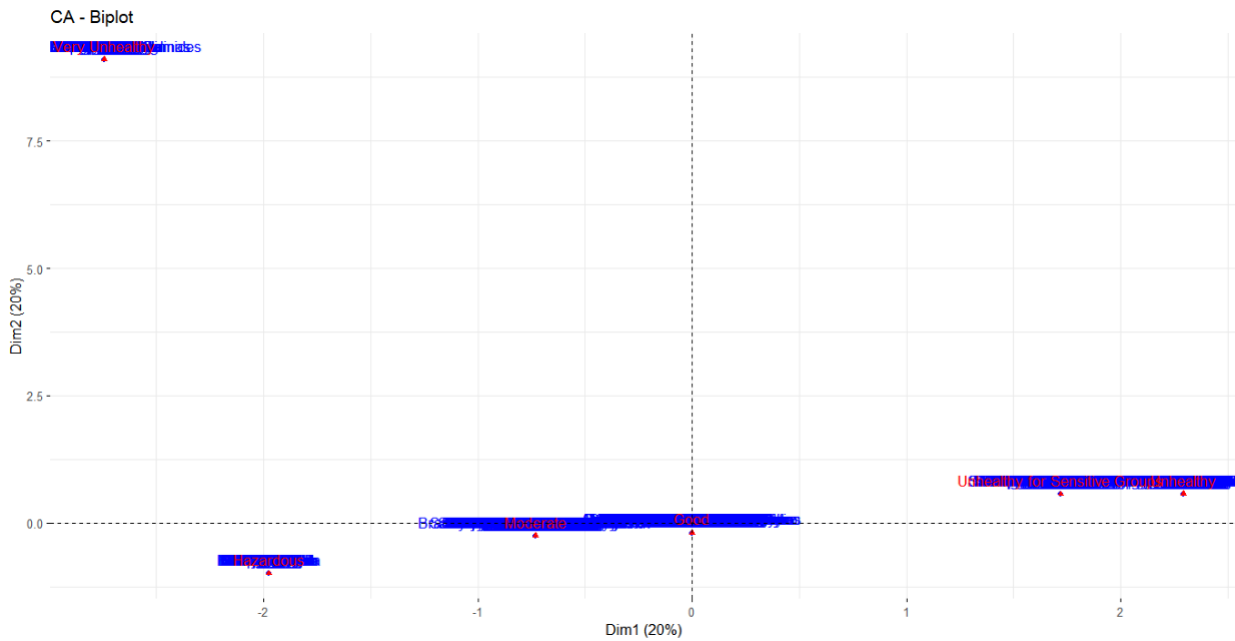


On voit que la contribution du variable NO2 AQI est plus important de la dimension 2 suivie de Ozone.AQI.Value ainsi de suite.

Le Modèle AFC

Le Modèle AFC (Analyse Composante Factorielle) est une methode ou une technique d'analyse factorielle qui permet de mettre en évidence et décrire des sociations entre deux ou plusieurs variables qualitatives.





Le graphique nous montre que les villes ayant une pollution de l'air sont très proches par rapport à l'axe 1 et sont à droite. Par contre les villes qui sont opposées ont un comportement différent par rapport à la pollution de l'air.

Les villes qui sont à gauche ont un comportement faible par rapport à la pollution de l'air.