

# TP2-ADM

Guillaume Bernard-Reymond et Lorenzo Gaggini

Décembre 2023

Dans ce TP, nous avons à notre disposition un jeu données concernant 100 villes françaises dont on connaît 54 indicateurs répartis en quatre grands thèmes : Economie, Risques, Nature et Culture.

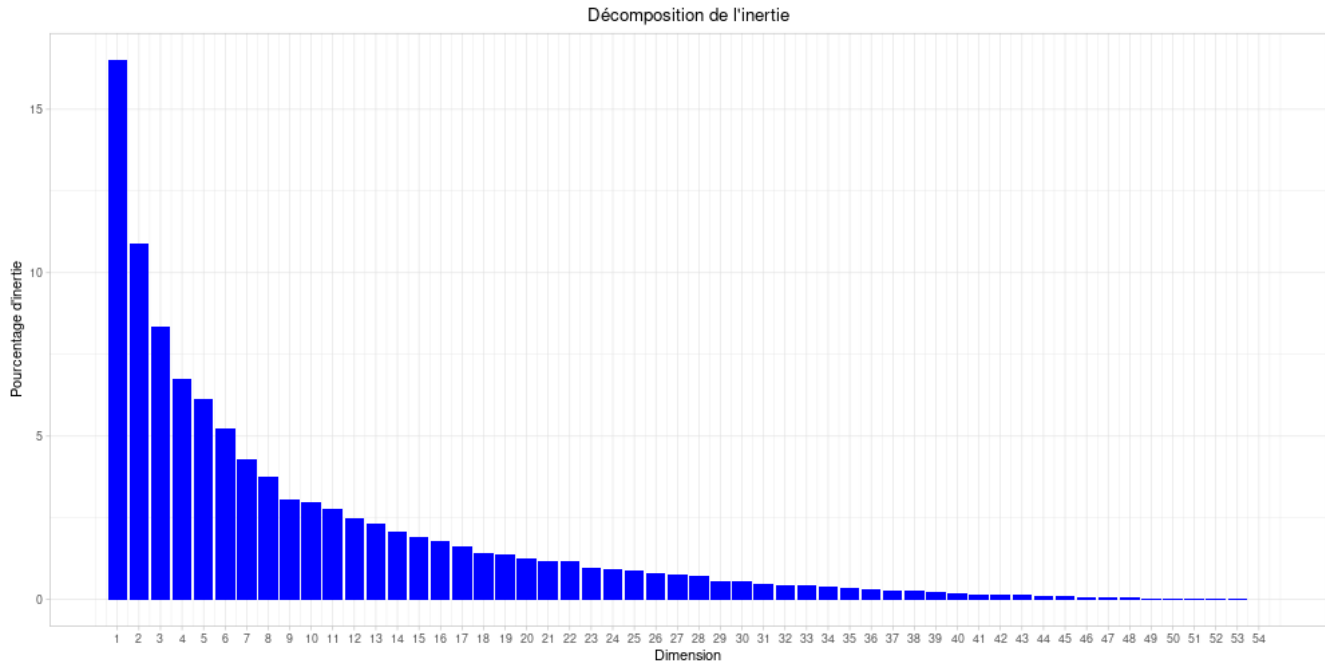
Nous avons décidé d'utiliser la bibliothèque Factoshiny pour son confort d'utilisation.

## 1 ACP normé sur l'ensemble des variables

### 1.1 Etude de la variance cumulée

Dans un premier temps, nous nous intéressons à la variance cumulée et notamment les grands décrochements. Nous allons devoir faire un choix qui ne sera malheureusement pas sans perte d'informations.

Voici le graphique des valeurs des valeurs propres :



Voici le tableau des variances cumulées :

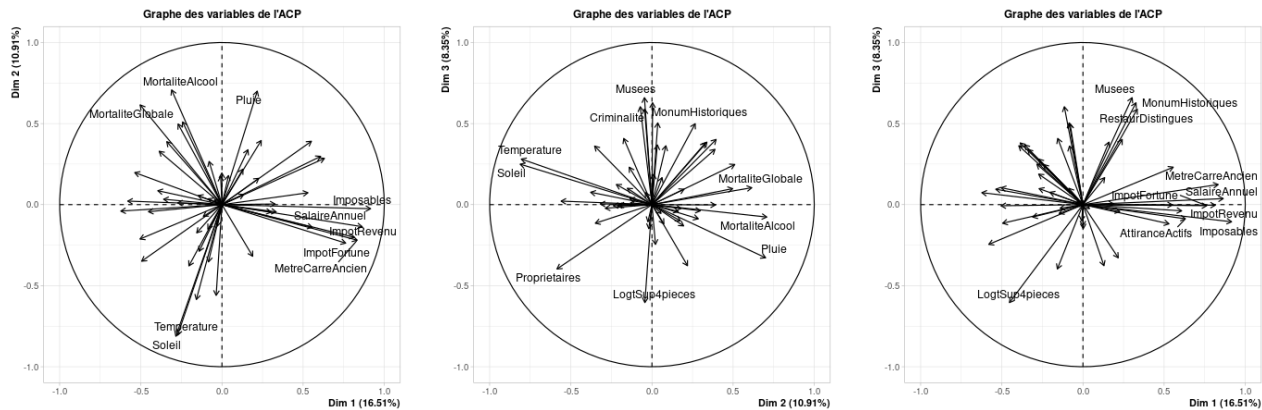
Valeurs propres	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
Variance	8.914	5.889	4.511	3.646	3.306	2.820	2.326	2.041
% of var.	16.508	10.906	8.353	6.752	6.121	5.222	4.307	3.780
Cumulative % of var.	16.508	27.414	35.767	42.519	48.641	53.862	58.169	61.949

Le nombre d'axes choisis pour mener l'étude de l'ACP est le même pour les individus que les variables car les valeurs propres des opérateurs d'inertie directe et duale sont les mêmes à un certain nombre de valeurs propres nulles près. Ainsi dans toute la suite du rapport, nous ne distinguerons plus, sauf mention explicite du contraire, la variance cumulée des individus et celle des variables.

On peut observer un décrochement après la troisième valeur propre. Nous nous contenterons des trois premiers axes avec une inertie cumulée de seulement 35.767%. Pour être plus exhaustif, on pourrait prendre davantage de valeurs propres. Toutefois, ne pouvant observer de véritables décrochements dans les valeurs propres, nous serions obligés d'en garder environ 8 car la différence entre deux valeurs successives est faible. Le temps de traitement en deviendrait trop long.

## 1.2 Etude des variables

Sur les graphiques suivants sont marqués les 10 variables les plus contributives à la fabrication des axes :



### Axe 1

- Contribution**

Tout d'abord, observons que  $\frac{1}{54} \approx 2\%$  ce qui nous pousse à ne considérer que les variables dont la contribution est très supérieure à 2%. On trouve alors : *Imposables* (9,41%), *SalairesAnnuels* (8,38%), *MetreCarreAncien* (7,75%), *ImpotRevenu* (7,46%) et *ImpotFortune* (6,52%). En additionnant, leur contribution on obtient : 39,52% ce qui est très supérieur à  $5 \times 2\% = 10\%$ . Ce sont donc des variables économiques qui contribuent en une part importante à la fabrication de l'axe 1. On voit d'ailleurs sur le plan (1,2), ce groupe de variables allant dans la même direction.

- Originalité : COS2**

On retrouve ici, les mêmes variables que précédemment et dont l'originalité est principalement décrite par l'axe 1 : *Imposables* (0,84), *SalairesAnnuels* (0,75), *ImpotFortune* (0,66), *MetreCarreAncien* (0,65).

On peut en conclure que cet axe 1 décrit des marqueurs de richesse.

### Axe 2

- Contribution** Parmi les variables dont la contribution est très supérieure à 2%, il y a : *Soleil* (11,20%), *Température* (11,01%), *MortaliteAlcool* (8,50%), *Pluie* (8,33%) et *MortaliteGlobale* (6,43%). De manière cumulée, on obtient 45,47%. Sur la représentation du plan (1,2), on voit ces variables s'opposer : *Soleil* et *Température* d'un côté, les trois restantes de l'autre.
- Originalité : COS2** L'originalité des variables *Soleil* et *Température* est capturée à hauteur de 0,66 et 0,65 par l'axe 2. Si on leur ajoute leurs valeurs sur les axes 1 et 3, on trouve *Soleil* : 0,8 et *Température* : 0,8 aussi. Pour la *MortaliteAlcool*, on a 0,50 sur l'axe 2 et 0,61 si l'on considère les trois axes. De même, le COS2 de la mortalité globale est de 0,38 le long de l'axe 2 et 0,64 si l'on considère les 3 axes. Enfin pour la pluie, on trouve 0,49 sur l'axe 2 et 0,65 pour les 3 axes.

Ici on commence à avoir une idée de ce que semble représenter cet axe : une opposition entre le Nord et le Sud de la France. L'étude ultérieure des individus nous permettra d'en dire certainement plus.

### Axe 3

- **Contribution**

Voici un tableau des variables les plus contributives à la fabrication de l'axe 3 :

Variables	Musées	MonumentHistorique	LogtSup4Pieces	Criminalite	RestauDistingues	Total
Contribution	9,65	8,15	8,10	8,08	7,70	41,68

- **Originalité : COS2**

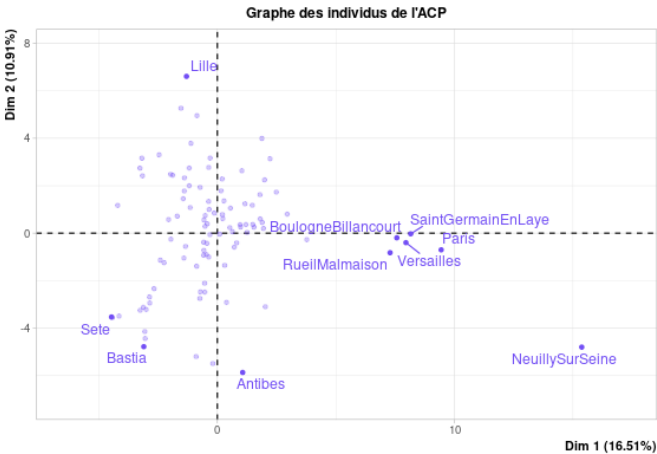
Pour ce qui est de l'originalité, on ne retrouve rien de très significatif. Celle-ci semble être diluée dans davantage d'axes que les 3 considérés.

Variables	Musées	MonumentHistorique	LogtSup4Pieces	Criminalite	RestauDistingues
Originalité sur l'axe 3	0,44	0,39	0,37	0,36	0,35
Originalité sur les 3 axes	0,53	0,5	0,58	0,38	0,46

Si l'on observe les plans (2,3) et (1,3), on peut voir un faisceau entre des variables culturelles *Musées*, *MonumentHistorique* et *RestauDistingues*, la variable *LogtSup4Pieces* étant elle négativement corrélée aux trois précédentes. L'originalité de la variable criminalité n'étant finalement pas suffisamment décrite par l'axe 3, on peut supposer que l'axe 3 reflète plutôt une tendance culturelle et qui va opposer grandes et petites villes.

### 1.3 Etude des individus

Voici le graphique du plan (1,2) avec les 10 individus les plus contributifs :



Pour ce qui concerne la contribution, nous comparerons les valeurs avec la contribution moyenne  $\frac{1}{100} = 0,01$ .

### Axe 1

- **Contribution** Six villes ont une contribution forte pour l'axe 1 dont une plus particulièrement :

Individus	Neuilly-Sur-Seine	Paris	Saint-Germain-en-Layes	Versailles	Boulogne-Billancourt	Rueil-Malmaison	Total
Contribution	26,51	10,0	7,46	7,10	6,43	5,96	63,46

La contribution de ces six villes est très largement supérieure à la contribution moyenne de 6%.

- **Originalité : COS2**

Les villes dont l'originalité est largement supportée par l'axe 1 sont les mêmes que précédemment à l'exception notable de Paris :

Individus	Versailles	Saint-Germain-en-Layes	Neuilly-Sur-Seine	Boulogne-Billancourt	Rueil-Malmaison
Originalité	0,73	0,67	0,65	0,61	0,57

L'originalité de Paris le long de cet axe est seulement de 0,25. Ceci semble indiquer que Paris n'est pas seulement originale par sa dimension économique mais que son originalité est multiple.

Ce que l'on peut conclure de cette étude de l'axe 1, c'est que ce dernier représente effectivement des variables économiques et que quelques villes de l'Ouest parisien contribuent fortement à son orientation. Si l'on met côte à côte le plan (1,2) du graphe de l'ACP des variables et celui des individus, on remarque évidemment une direction similaire entre les variables étudiées et les villes contributives.

## Axe 2

- **Contribution**

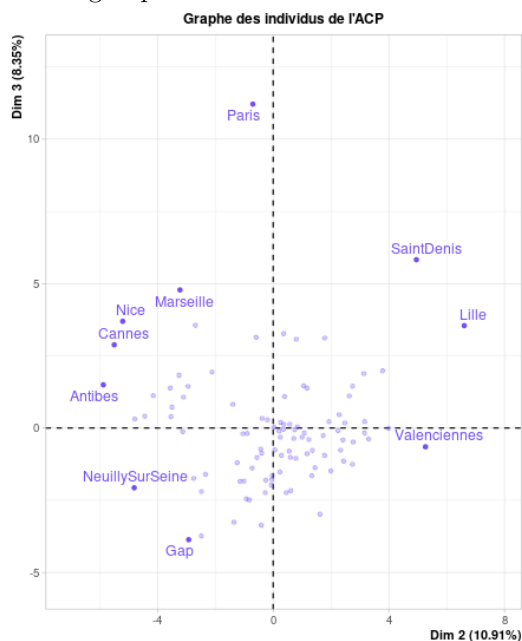
Les principales contributions à la construction de l'axe 2 sont données par 6 villes : *Lille* (7,42%), *Antibes* (5,85%), *Cannes* (5,13%), *Valenciennes* (4,71%), *Nice* (4,60%) et *Saint-Denis* (4,17%). En additionnant leur contribution, on obtient 31,88%.

- **Originalité : COS2**

Les cinq villes dont l'originalité est décrite partiellement par cet axe 2 sont : *Lille* (0,47), *Antibes* (0,44), *Troyes* (0,42), *Nice* (0,40) et *Valenciennes* (0,38).

De cet axe 2 se dégage un axe Nord-Sud. Au Nord, des villes pluvieuses avec une mortalité globale supérieure et au Sud des villes chaudes et ensoleillées.

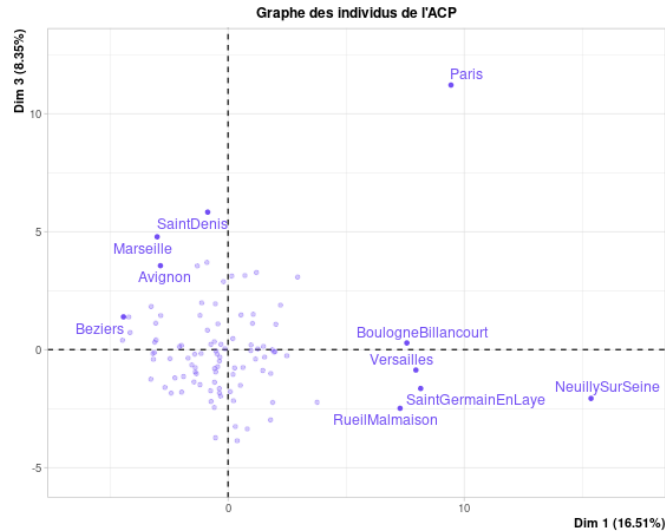
L'édition du plan (2,3) nous montre un regroupement de villes du Sud :



Ce regroupement se trouve être justement dans la même direction que celui des variables *Soleil* et *Températures*.

## Axe 3

Voici le plan (1,3) des individus :



#### • Contribution

On trouve ici quelques contributeurs importants et notamment une ville singulière *Paris* contribuant à hauteur de 27,86%. Les deux autres villes sont *Saint-Denis* (7,53%) et *Marseille* (5,06%). Dans une moindre mesure : *Gap* (3,31%) et *Cholet* (3,09%) contribue à la construction de cet axe.

#### • Originalité COS2

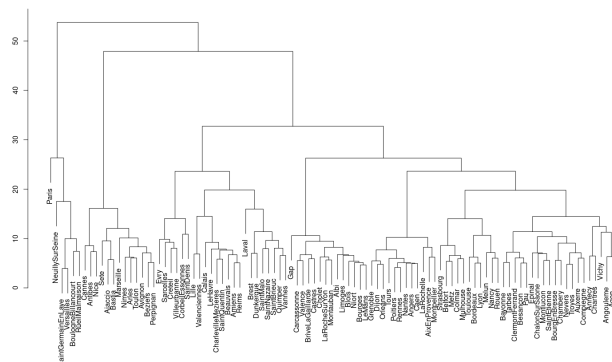
Aucune ville ne semble être majoritairement décrite par cet axe 3. Toutefois on peut remarquer un fait notable. Nous avons dit que *Paris*, bien que contributeur de l'axe 1, ne voyait pas son originalité y être décrite. *Paris* se trouve être plus original selon cet axe 3 (0,35) auquel il participe grandement à sa construction. Les villes contributrices tirent une part de leur originalité de cet axe : *Marseille* (0,29), *Cholet* (0,28) *Saint-Denis* (0,26) et *Gap* (0,22).

Enfin *La-Roche-Sur-Yon* a un COS2 égale à 0,31.

En superposant les plans où se trouvent l'axe 3, on remarque évidemment que les variables contributives à sa construction la position de *Paris* ne sont pas étrangères. La direction et la même et cela est dû au fait que *Paris* a une contribution majeure à cet axe. Elle en tire d'ailleurs son originalité principale. Cet axe semble être celui des marqueurs culturels.

## 1.4 Conclusion

On retrouve dans cet ACP, les résultats obtenus dans le TP2 où le dendrogramme donnait une impression visuelle de former 3 classes :



On retrouve dans cette ACP, 3 axes principaux :

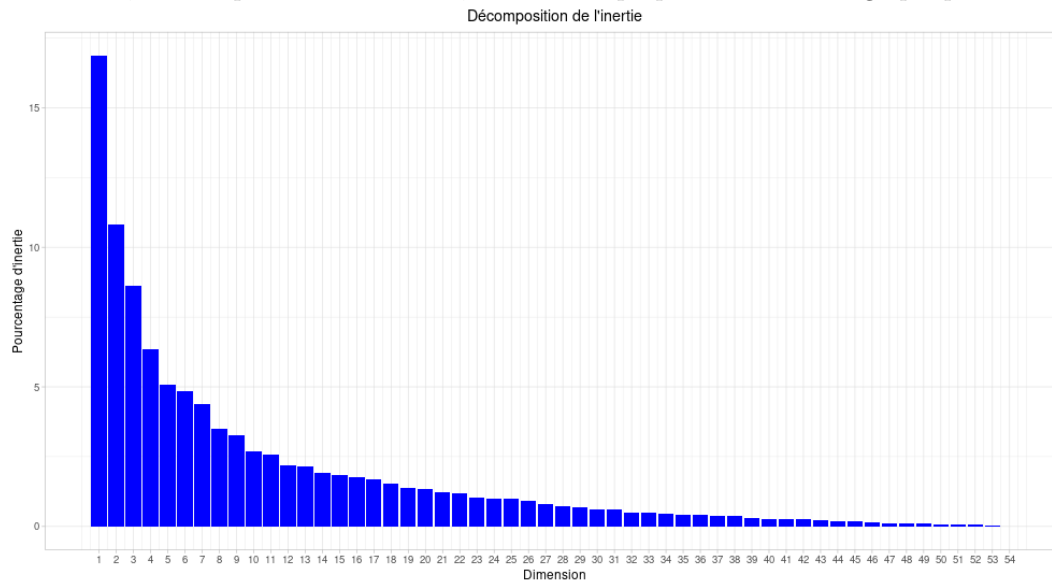
- **Axe 1** : axe à dominante économique avec de gros contributeurs se caractérisant par une position géographique bien particulière : l'Ouest parisien.
- **Axe 2** : celui-ci oppose le Sud ensoleillé et chaud au Nord pluvieux dont certaines variables au caractère plus morbide sont contributrices.
- **Axe 3** : Cette fois ci c'est la culture qui se distingue mais la culture parisienne. En effet, il s'agit d'un contributeur majeur isolé à sa fabrication.

## 2 ACP de rang sur l'ensemble des variables

Après avoir obtenu la matrice des rangs, nous lançons l'ACP avec Factoshiny.

### 2.1 Etude de la variance cumulée

Nous nous intéressons, comme pour l'ACP normée, aux valeurs propres dont voici le graphique :



et le tableau de valeurs :

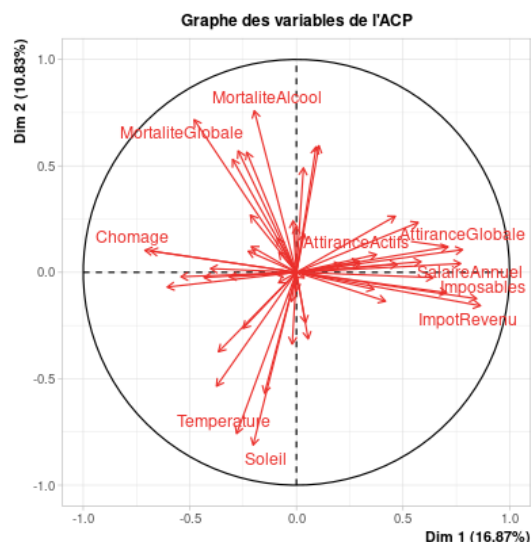
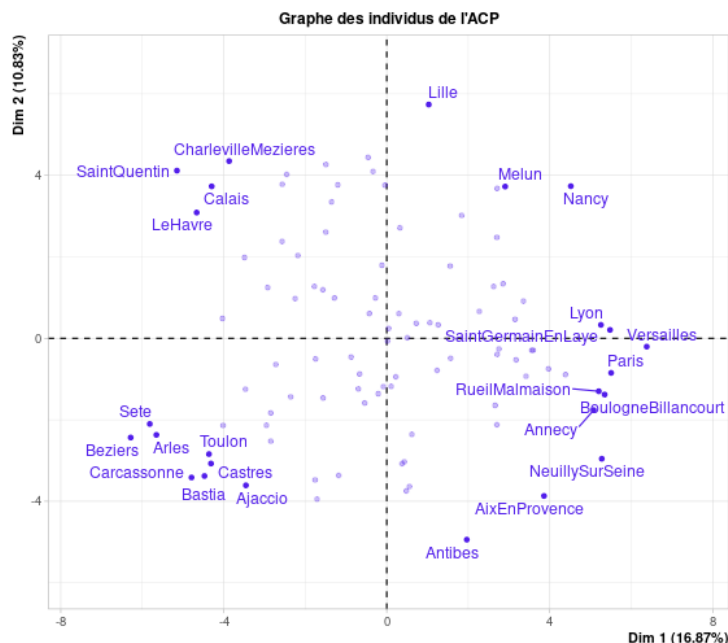
Valeurs propres	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
Variance	9.110	4.511	4.662	3.422	2.752	2.616	2.358	1.879
% of var.	16.871	10.825	8.634	6.338	5.096	4.844	4.366	3.480
Cumulative % of var.	16.871	27.696	36.330	42.668	47.764	52.608	56.974	60.455

On a un décrochement entre la 4-ème et la 5-ème valeurs ce qui nous pousse à nous intéresser aux quatre premiers axes. C'est déjà une différence avec l'ACP normée.

### 2.2 Etude des différents plans

#### Plan (1,2)

Côte à côte voici les graphiques des individus et celui des variables avec les 25 plus gros contributeurs pour les individus et les 10 plus gros contributeurs pour les variables :



On retrouve le même faisceau des variables de richesse que sur l'ACP normée, auxquelles viennent se rajouter *AtiranceActifs* et *AttiranceGlobale*. Ces variables sont à mettre en opposition à la variable chômage qui décrit plus la pauvreté. Dans la même direction se trouvent les villes trouvées lors de l'ACP normée, celle de l'Ouest parisien. Des villes réputées riches comme *Lyon*, *Annecy* ou bien *Aix-en-Provence* viennent compléter la liste.

**Tableau des variables les plus contributrices pour l'axe 1 :**

Côté -				Côté +			
Variables	Coordonnées	CTR	COS2	COS2	CTR	Coordonnées	Variables
Chômage	-0.71	5.55	0.51	0.75	8.19	0.86	ImpotRevenu
				0.71	7.85	0.85	Imposables
				0.61	6.67	0.78	AttiranceGlobale

**Tableau des individus les plus contributeurs pour l'axe 1 :**

Côté -					Côté +		
Villes	Coordonnées	CTR	COS2	COS2	CTR	Coordonnées	Villes
Beziers	-6.28	4.33	0.57	0.58	4.47	6.38	Versailles
Sète	-5.81	3.71	0.41	0.44	3.33	5.51	Paris
				0.40	3.30	5.48	Saint-Germain-En-Laye

Sur cet axe 1, il y a plusieurs points contributifs, mais peu sont illustratifs car seuls les COS2 de *ImpotRevenu* et *Imposables* sont importants.

L'axe 2 met en opposition des villes du Sud et du Nord avec deux groupes bien distincts. Ceci correspond à une opposition entre les variables *Température*, *Soleil* et les variables *MortaliteAlcool* et *MortaliteGlobale*. Ceci est clairement visible sur les deux tableaux suivants :

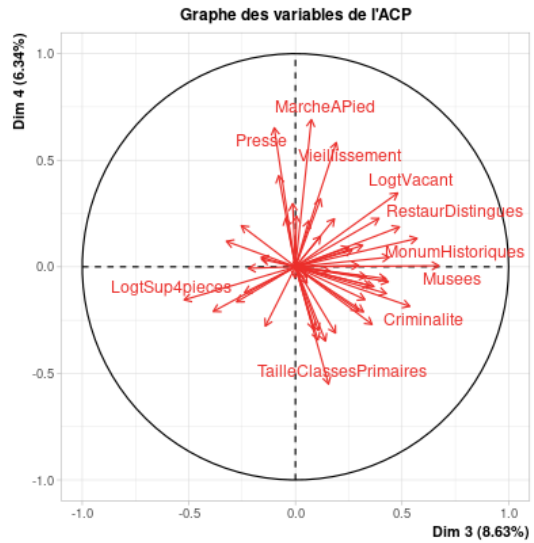
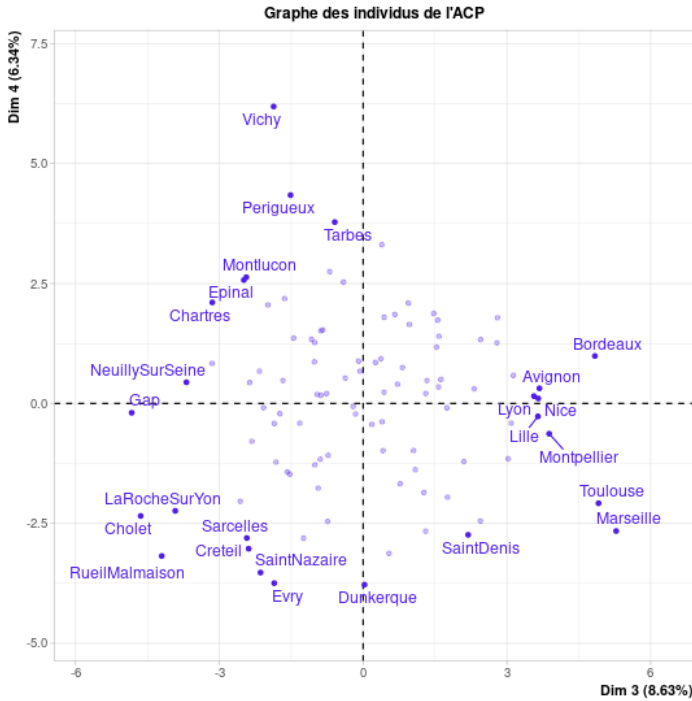
**Tableau des variables les plus contributrices pour l'axe 2 :**

Côté -					Côté +		
Variables	Coordonnées	CTR	COS2	COS2	CTR	Coordonnées	Variables
Soleil	-0.81	11.27	0.66	0.58	9.85	0.76	MortaliteAlcool
Temperature	-0.76	9.83	0.57	0.51	8.80	0.72	MortaliteGlobale
Viellissement	-0.57	5.52	0.32				

Tableau des individus les plus contributeurs pour l'axe 2 :

Côté -				Côté +			
Villes	Coordonnées	CTR	COS2	COS2	CTR	Coordonnées	Villes
Antibes	-4.94	4.18	0.33	0.52	5.64	5.74	Lille
				0.33	3.38	4.44	Valenciennes
				0.22	3.12	4.27	Saint-Denis

Plan (3,4)



En observant ces deux graphiques, il ressort que les variables les mieux représentées pour l'axe 3 sur le plan sont *Musées*, *MonumHistoriques*, *Criminalité* et *RestauDistingue*. Ces variables sont corrélées négativement à *LogtSup4Pieces*. En regardant de plus près les individus contributeurs, on trouve des grandes villes françaises à droite qui font la part belle à la culture mais où la criminalité est importante. Par criminalité, il faut entendre le nombre d'infractions et délits commis. Les infractions relatives au code de la route y sont forts nombreuses dans les grandes villes.

Tableau des variables les plus contributrices pour l'axe 3 :

Côté -				Côté +			
Variables	Coordonnées	CTR	COS2	COS2	CTR	Coordonnées	Variables
LogtSup4Pieces	-0.52	5.85	0.27	0.45	9.72	0.67	Musées
				0.33	6.99	0.57	MonumHistoriques
				0.29	6.18	0.54	Criminalité
				0.32	5.11	0.49	RestauDistingues

Tableau des individus les plus contributeurs pour l'axe 3 :

Côté -				Côté +			
Villes	Coordonnées	CTR	COS2	COS2	CTR	Coordonnées	Villes
Gap	-4.83	5	0.33	0.44	5.98	5.28	Marseille
Cholet	-4.64	4.62	0.33	0.43	5.17	4.91	Toulouse
				0.40	5.02	4.84	Bordeaux

La principale différence avec l'ACP normée c'est que l'on ne retrouve pas Paris sur cet axe alors qu'il en était le principal contributeur précédemment. Le fait de faire une ACP de rang à éliminer une grande partie de la singularité



de Paris et les variables culturelles qui la décrivait. Ici les COS2 sont plutôt faibles et l'originalité des villes est partiellement mis en évidence ici.

Sur l'axe 4, trois variables ont une forte contribution : *MarcheAPied* (13.95), *Presse* (12.43) et *Vieillessement* (9.94). Pour ce qui est des villes, deux sortent du lot *Vichy* (11.20) et *Tarbes* (4.17). L'originalité de Vichy est tirée à 50% par ces variables et à 31% pour Tarbes. Concernant les variables, *MarcheaPied* tire 48% de son originalité le long de cet axe, 43% pour *Presse* et 34% pour *Vieillessement*.

Ces trois variables semblent décrire des variables liées à une population de retraités. Il faudrait regarder plus en détail les villes concernées pour en avoir le cœur net.

### **3 ACP normé pour les différents thèmes**

### **4 ACP de rang pour les différents thèmes**