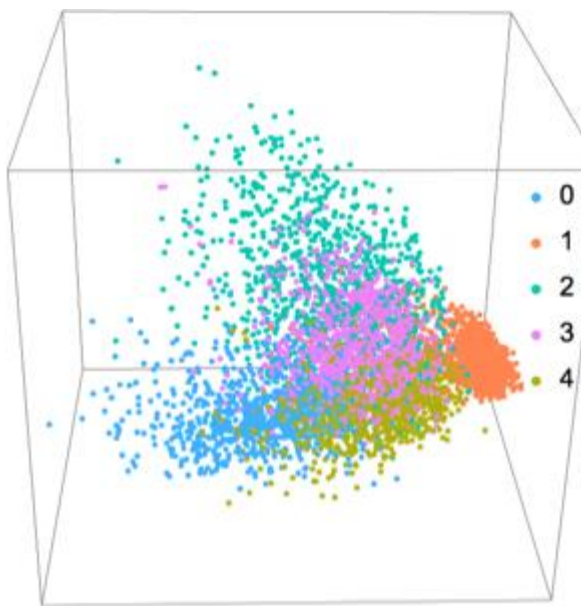


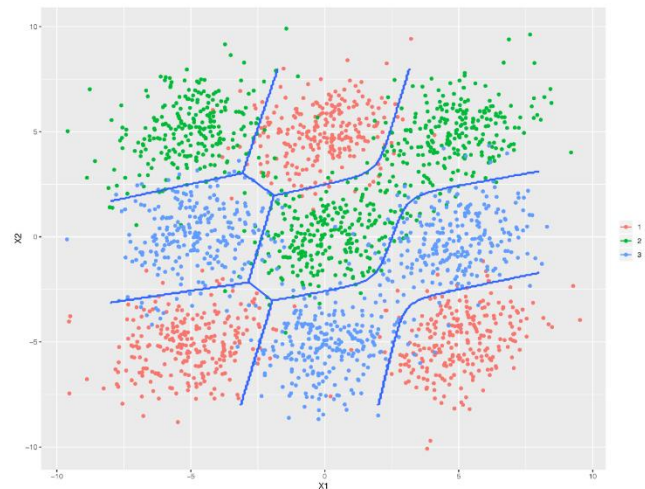
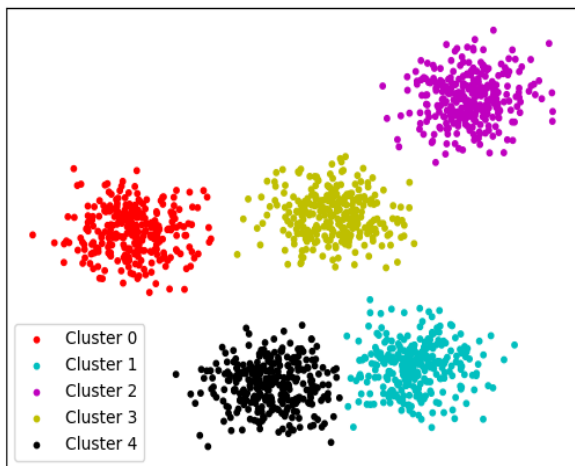
## Projet S3 : Méthodes de classifications

- *K-means*
- *Classification ascendante hiérarchique (CAH)*



### Réalisé par

- *AGOSSOUVO Bernice*
- *CHELIGHEM Nedjma*



## Table de matière

<b>I. Introduction .....</b>	<b>3</b>
<b>II. Statistiques .....</b>	<b>4</b>
II.1 Présentation de jeu de données .....	4
II.2 Statistiques univariées :.....	4
II.3 Statistique Bivariée .....	5
<b>III. Analyses en Composantes Principales des données (ACP) .....</b>	<b>7</b>
III.1 Choix des axes factoriels.....	7
III.2 Contribution et qualité de représentation des Variables et individus sur les axes factoriels 8	
III.3 Interprétation des résultats de l'ACP .....	12
III.4 Classification des individus .....	12
<b>IV. Méthodes de classifications .....</b>	<b>13</b>
IV.1 Classification de type K-Means.....	13
IV.2 Classification Ascendante Hiérarchique (CAH) .....	17
IV.3 CAH avec la fonction HCPC.....	22
<b>V. Conclusion .....</b>	<b>27</b>
<b>VI. Table d'illustration (Figures).....</b>	<b>28</b>
<b>VII. Table d'illustration (Tableau).....</b>	<b>28</b>
<b>VIII. Annexe .....</b>	<b>29</b>

## I. Introduction

La classification des données est une approche statistique qui consiste à opérer des regroupements en classes homogènes d'un ensemble d'individus. Son objectif est de rechercher une typologie ou segmentation, c'est à dire une répartition des individus dans des classes, sur la base de l'observation. Il existe différentes méthodes de classification dont les méthodes supervisées et non supervisées.

L'objectif de la classification supervisée est principalement de définir des règles permettant de grouper des objets dans des classes à partir de variables qualitatives ou quantitatives caractérisant ces objets. On dispose ainsi d'éléments déjà classés et on cherche à classer de nouveaux éléments. La classification non supervisée quant à elle se réalise sans connaissance de groupes d'appartenance. On dispose d'éléments non classés et on cherche à les regrouper en classes. Ces différentes méthodes s'appuient sur des algorithmes. Ainsi on peut citer des algorithmes tels que : K-means, Hierarchical clustering (Classification hiérarchique), Principal Component Analysis, Neural networks (Réseaux de neurones) pour ne citer que ça.

Dans ce projet, nous utilisons la méthode non supervisée avec les algorithmes K-means et de CAH pour concevoir une classification de nos données.

On commence par décrire notre jeu de données, vient après une analyse statistique sur les différentes variables, valeurs moyennes prises par les individus, leur dispersion. Enfin après une Analyse en Composantes Principales faite sur le jeu de données dans le but de réduire les dimensions des variables , nous utiliserons différentes méthodes de classification pour classer nos données.

## II. Statistiques

### II.1 Présentation de jeu de données

Les données sont tirées d'une façon aléatoire sur le site <https://fr.tutiempo.net/climat>, elles contiennent les informations climatiques de 51 villes de la France sur l'année 2018. Pour chaque ville, nous avons la température moyenne, maximale et minimale annuelles, le nombre total de jours de pluie, de neige, de tempête, de brouillard au cours de l'année, la précipitation totale annuelle de pluie et/ou neige fondue (mm), Vitesse moyenne annuelle du vent (Km/h).

Toutes les variables sont de type numérique (Ci-dessous la liste des variables et leurs diminutifs pour la suite)

T	Température moyenne annuelle
TM	Température maximale moyenne annuelle
Tm	Température minimale moyenne annuelle
PP	Précipitation totale annuelle de pluie et/ou neige fondue (mm)
V	Vitesse moyenne annuelle du vent (Km/h)
RA	Total jours de pluie durant l'année
SN	Total jours de neige durant l'année
TS	Total jours de tempête durant l'année
FG	Total jours de brouillard durant l'année

Tableau 1: Description des variables

### II.2 Statistiques univariées :

La moyenne est l'indicateur le plus simple pour résumer l'information fournie par un ensemble de données statistiques : elle est égale à la somme de ces données divisée par leur nombre. Elle peut donc être calculée en ne connaissant que ces deux éléments, sans connaître toute la distribution. La moyenne d'une distribution n'est pas toujours le meilleur indicateur : la médiane est souvent plus pertinente. Mais son calcul exige de connaître toute la distribution, ou en tout cas sa partie centrale.

Le coefficient de variation (CV) est le rapport de l'écart-type à la moyenne. Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande. Il est généralement exprimé en pourcentage. Sans unité, il permet la comparaison de distributions de valeurs dont les échelles de mesure ne sont pas comparables. Lorsque l'on dispose de valeurs estimées, le CV rapporte l'écart-type de l'estimation à la valeur de cette estimation. Plus la valeur du coefficient de variation est faible, plus l'estimation est précise.

Nous pouvons voir à travers le tableau ci-dessous les différentes variables avec leurs moyennes et leurs coefficients de variation

Variables	Moyenne	Coefficient de variation
<b>T</b>	13.74	0.13
<b>TM</b>	19.04	0.09
<b>Tm</b>	8.19	0.26
<b>PP</b>	777.79	0.19
<b>V</b>	11.76	0.24
<b>RA</b>	151.24	0.16
<b>SN</b>	8.22	0.71
<b>TS</b>	15.41	0.74
<b>FG</b>	44.12	0.66

Tableau 2: Analyses descriptives

Il ressort de ce tableau que les variables T, TM, Tm PP, V, RA sont moins dispersées autour de leurs moyennes contrairement aux variables SN, TS et FG.

En effet, les valeurs de SN, TS et FG varient respectivement entre 1 et 24, 0 et 38 et 0 et 81. Cela s'explique par le fait que le nombre de jours de neige, de tempête et de brouillard varient largement selon les villes. Il peut ne point avoir de neige toute une année à Perpignan alors que dans la ville de Lille il a eu 26 jours de neige. Il s'en va de même pour le nombre de jours de tempête et de brouillard.

## II.3 Statistique Bivariée

### II.3.1 Matrice de corrélation

Une matrice de corrélation est utilisée pour évaluer la dépendance entre plusieurs variables en même temps. Le résultat est une table contenant les coefficients de corrélation entre chaque variable et les autres.

Il existe différentes méthodes de tests de corrélation : Le test de corrélation de Pearson, la corrélation de Kendall et celle de Spearman qui sont des tests basés sur le rang. Nous avons utilisé ici le test de corrélation de Pearson (voir graphe et tableau ci-dessous).

	T	TM	Tm	PP	V	RA	SN	TS	FG
T	1.00								
TM	0.93	1.00							
Tm	0.89	0.70	1.00						
PP	0.26	0.20	0.28	1.00					
V	0.06	-0.18	0.21	-0.12	1.00				
RA	-0.67	-0.57	-0.70	-0.07	-0.07	1.00			
SN	-0.73	-0.67	-0.70	-0.22	-0.06	0.33	1.00		
TS	0.47	0.47	0.22	0.02	0.24	-0.14	-0.11	1.00	
FG	-0.65	-0.55	-0.64	0.09	-0.19	0.62	0.34	-0.36	1.00

Tableau 3: Matrice de corrélation

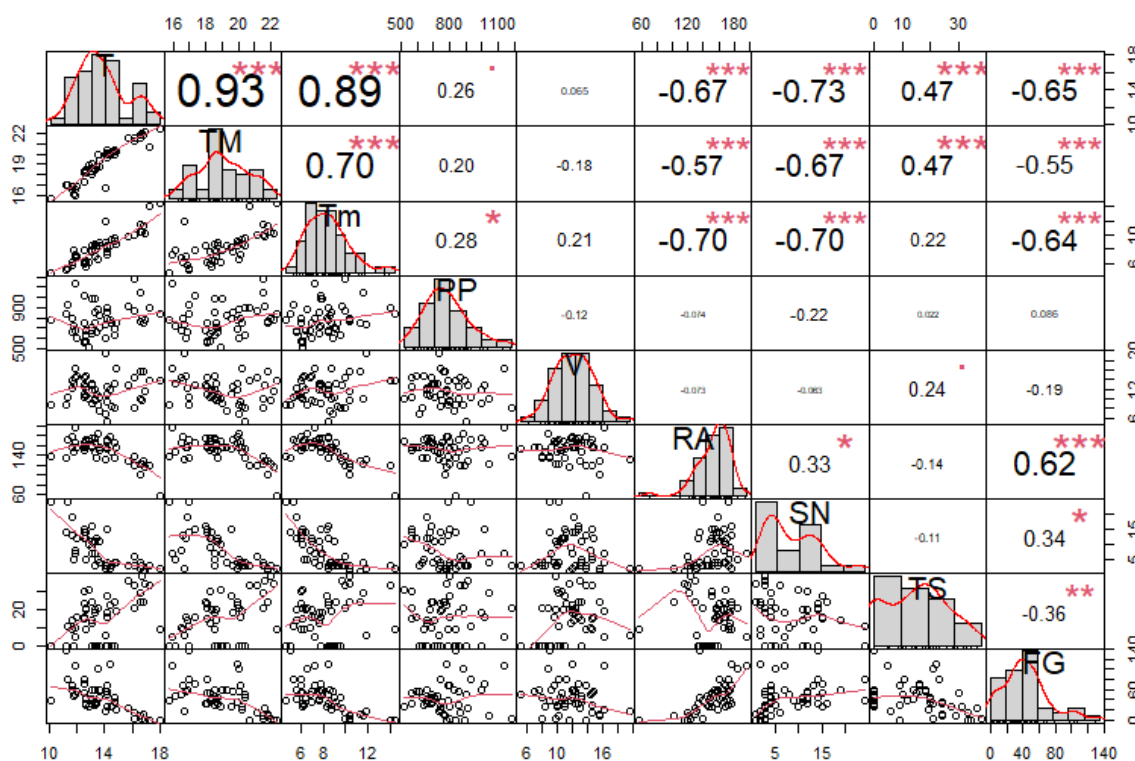


Figure 1: Corrélation des variables

Nous pouvons voir également à travers le tableau que les cases bleues représentent une corrélation linéaire positive (proche de 1), les cases en verts signifient une corrélation linéaire indépendante (proche de 0) et les cases jaunes représentent une corrélation linéaire négative (proche de -1)

### II.3.2 Analyses des corrélations

La température moyenne annuelle (T) et la température maximale moyenne annuelle (TM) atteignent une corrélation de 0.93 très proche de 1 cela signifie une corrélation linéaire positive très forte. Il y a une corrélation linéaire positive marquante également entre la température moyenne annuelle (T) et la température minimale moyenne annuelle (Tm) qui atteint les 0.89. Une autre corrélation linéaire positive marquante est celle entre température maximale moyenne annuelle (TM) et la température minimale moyenne annuelle (Tm) qui atteint les 0.70.

Aussi nous constatons que plusieurs variables sont linéairement corrélées de façon négative. Le premier résultat atteint -0.73 pour la température moyenne annuelle (T) et le total de jours de neige durant l'année (SN). Le second résultat atteint -0.70 (arrondi au dixième) pour la température minimale moyenne annuelle (Tm) et le total de jours de pluie durant l'année (RA), aussi pour la température minimale moyenne annuelle (Tm) et le total de jours de neige durant l'année (SN) ce qui signifie que la température minimale moyenne annuelle est corrélée au total de jours de pluie et de neige durant l'année.

On voit une corrélation non linéaire (proche de 0) entre la précipitation totale annuelle de pluie et/ou neige fondue (PP) avec le total de jours de tempête durant l'année (TS) qui atteint les 0.02 ce qui permet de déduire que ces variables ne sont pas liées linéairement. Pour la température moyenne annuelle (T) et la vitesse moyenne annuelle du vent (V) nous voyons qu'il atteint 0.06 (arrondi au dixième) on peut déduire également qu'ils ne sont pas liés linéairement ces deux variables.

### III. Analyses en Composantes Principales des données (ACP)

#### III.1 Choix des axes factoriels

En se référant à la règle de **Kaiser**, on ne peut retenir que deux axes pour l'ACP alors que ces deux axes nous permettent de prendre en compte seulement 65,1% de l'information, ce qui n'est pas assez parlant. Ainsi on construit le diagramme d'éboulis des valeurs propres ci-dessous pour mieux faire le choix

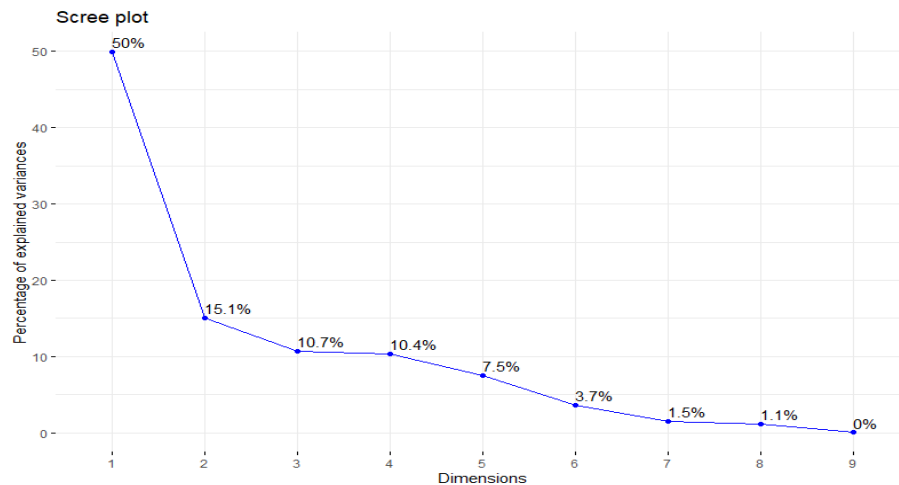


Figure 2: Diagramme d'éboulis des valeurs propres

Le diagramme indique que les valeurs propres commencent à former une ligne droite après la troisième composante principale. Comme 75,76 % représente une part acceptable de la variation expliquée dans les données, nous faisons le choix des trois premiers axes principaux soit des trois premières composantes principales

## III.2 Contribution et qualité de représentation des Variables et individus sur les axes factoriels

### III.2.1 Variables

#### III.2.1.1 Qualité de représentation et contribution

Nous mettons en évidence les variables selon leurs qualités de représentation sur chaque axe ou leurs contributions aux composantes principales. Ensuite, nous décrivons la visualisation des variables et tirons des conclusions concernant leurs corrélations.

Les variables qui sont bien représentées sont celles qui sont proches du cercle de corrélation ou dont le cos 2 est élevé. Le graphe ci-dessous nous donne un aperçu de la qualité de représentation des variables sur les différents axes

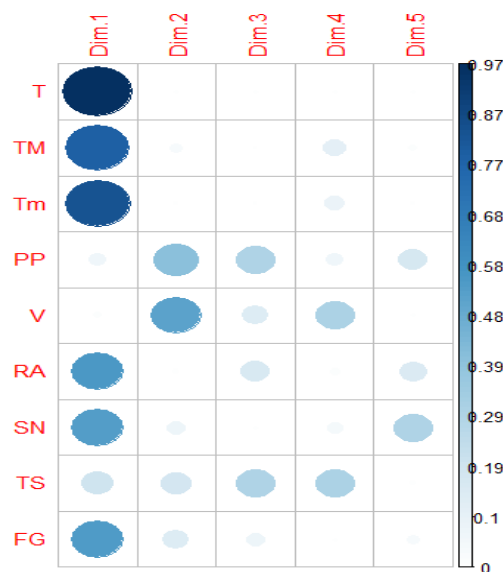


Figure 3: Qualité de représentation des variables sur les différents axes

On peut voir les variables bien représentées, avec un cercle bien grand comparé aux autres et une couleur bleu foncé. Ainsi les variables T, TM, Tm, RA, SN et FG sont respectivement bien représentées sur le **1<sup>er</sup> axe factoriel (Dim 1)** alors que PP et V sont plutôt mieux représentés sur le **2<sup>ème</sup> axe factoriel**. On peut faire le parallèle avec la contribution de chaque variable sur chaque axe factoriel (voir graphiques ci-dessous)



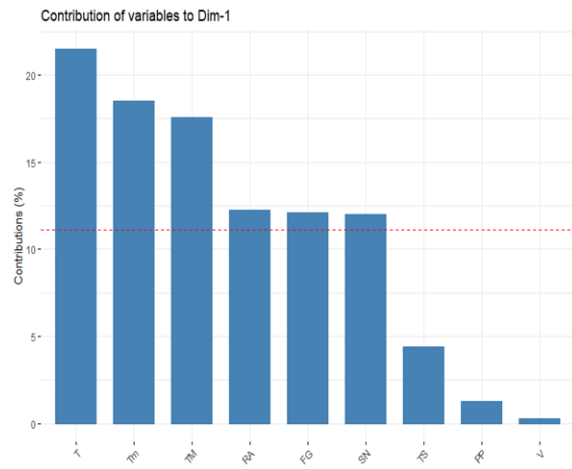


Figure 5: Contribution de chaque variable sur le 1<sup>er</sup> axe factoriel

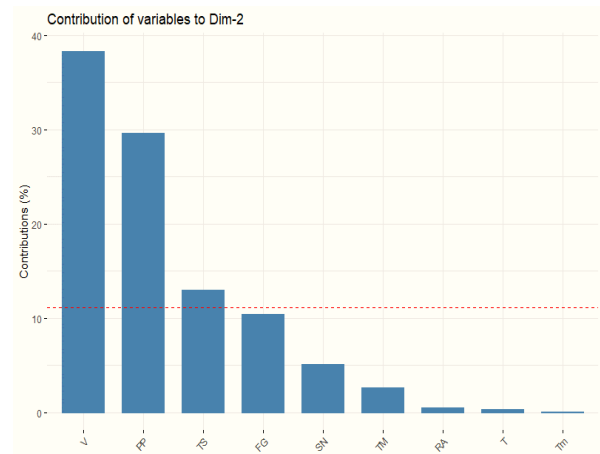


Figure 4: Contribution de chaque variable sur le 2<sup>ème</sup> axe factoriel

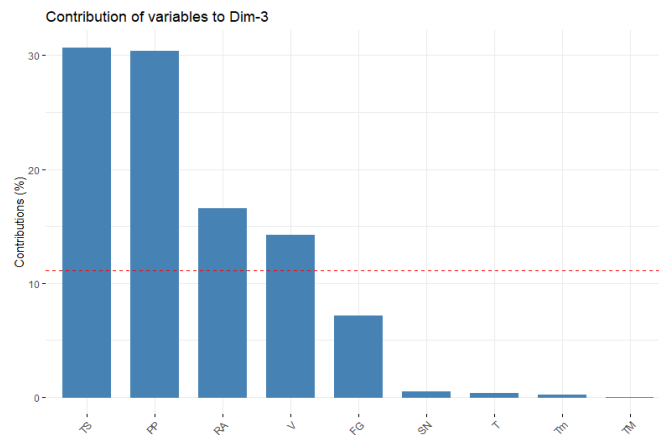
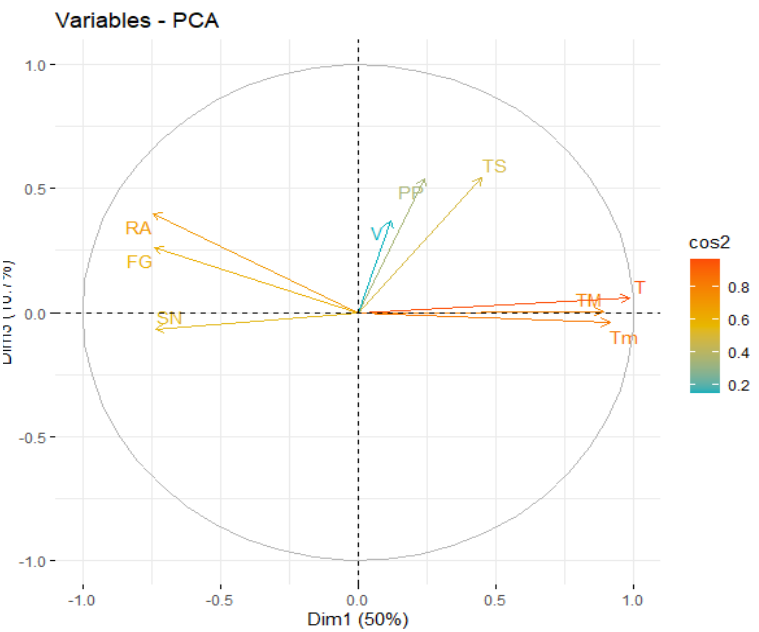
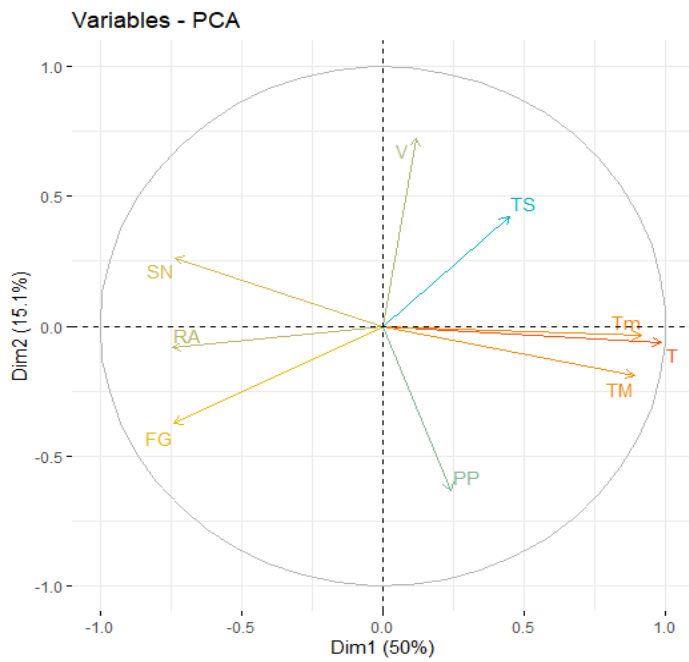


Figure 6: Contribution de chaque variable sur le 3<sup>ème</sup> axe factoriel

### III.2.1.2 Cercle de corrélation

La corrélation entre une variable et une composante principale (PC) est utilisée comme coordonnées de la variable sur la composante principale. La représentation des variables diffère de celle des observations : les observations sont représentées par leurs projections, mais les variables sont représentées par leurs corrélations (Abdi and Williams 2010).



Liaison entre les variables

- ✓ Les variables TM, Tm sont positivement liées avec la variable T
- ✓ Les variables RA, SN et FG sont négativement liées avec les variables T, TM et Tm

### III.2.2 Individus

#### III.2.2.1 Qualité de représentation et contribution

##### Plan factoriel (1,2)

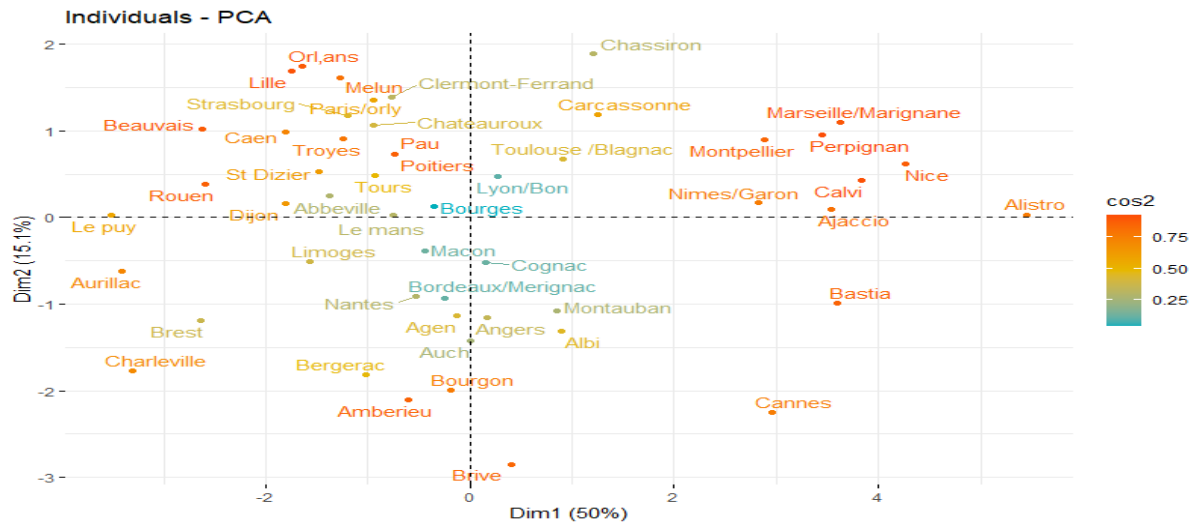


Figure 9: Qualité de représentation des individus sur le plan factoriel (1,2)

Nous pouvons voir sur le plan factoriel que les villes ayant une couleur foncée s'approchant du rouge représentent une bonne qualité de représentation (proche de 1) et les villes ayant une couleur bleue ont une faible qualité de représentation (proche de 0)

Ainsi Alistro, Calvi, Rouen, Perpignan, Marseille, Ajaccio, Beauvais et Bastia sont bien représentés sur le 1<sup>er</sup> axe. Celles bien représentées sur l'axe 2 sont Brive, Bourgon et Amberieu.

### Plan factoriel (1,3)

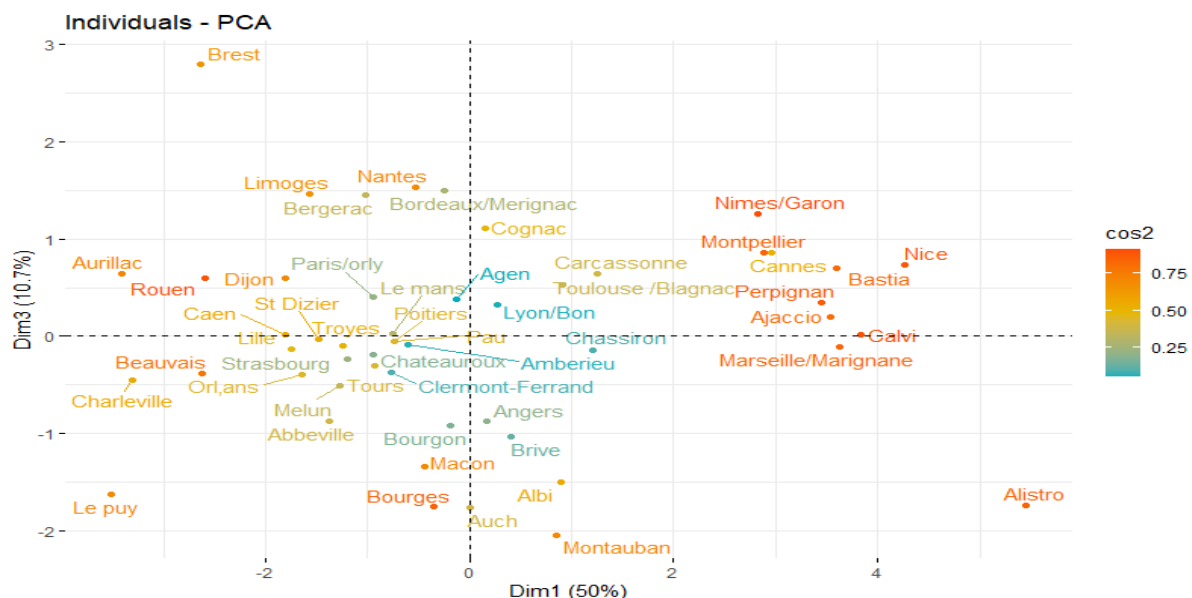


Figure 10: Qualité de représentation des individus sur le plan factoriel (1,3)

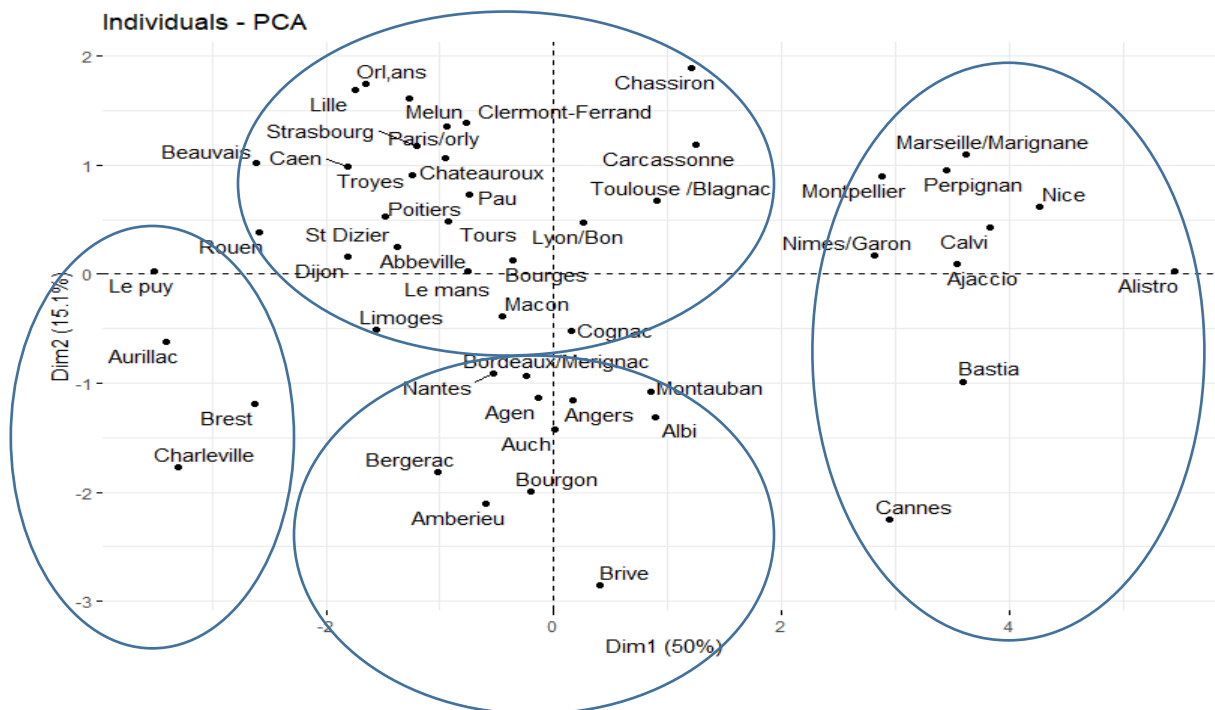
Sur l'axe 3 les individus bien représentés sont Auch, Montauban, Bourges, Brest, Macon et Albi et elle contribue bien à la construction de l'axe 3 (Voir Annexe). On note aussi que les villes ayant une couleur bleue ont une faible qualité de représentation (proche de 0).

### III.3 Interprétation des résultats de l'ACP

Les variables T, Tm et TM sont opposées aux variables RA, SN Et FG sur le premier axe. Les villes Rouen, Beauvais et Dijon par ex sont opposés aux villes Alistro, Nice et Calvi sur cet axe. Les premiers prennent des valeurs élevées pour le nombre de jours de pluie, de neige et de tempête et des valeurs faibles pour les températures moyennes, maximales et minimale alors que les seconds prennent des valeurs faibles pour le nombre de jours de pluie, de neige et de tempêtes et des valeurs élevées pour les températures.

### III.4 Classification des individus

En se référant à l'ACP plus précisément à la projection des individus sur le premier plan factoriel, on voit se dessiner quatre tendances. Voir schéma ci-dessous



- ✓ Les individus dont les températures moyenne, maximale et minimale sont élevées alors que le nombre de jours de pluies, de neiges et de brouillard sont faibles.
- ✓ Les individus qui se caractérisent à l'inverse par des températures faibles et des nombres de jours de pluie de neige et de brouillard élevés.
- ✓ Ceux dont la vitesse moyenne annuelle du vent est élevée et dont les diverses températures sont dans la moyenne
- ✓ Et Ceux ayant la précipitation forte et dont les températures sont dans la moyenne.

## IV. Méthodes de classifications

### IV.1 Classification de type K-Means

K-means est un algorithme non supervisé de clustering non hiérarchique. Il permet de regrouper en K clusters distincts les observations du data set. Ainsi les données similaires se retrouveront dans un même cluster. Par ailleurs, une observation ne peut se retrouver que dans un cluster à la fois (exclusivité d'appartenance). Une même observation, ne pourra donc, appartenir à deux clusters différents.

Pour pouvoir regrouper un jeu de données en K cluster distincts, l'algorithme K-Means a besoin d'un moyen de comparer le degré de similarité entre les différentes observations. Ainsi, deux données qui se ressemblent, auront une distance de dissimilarité réduite, alors que deux objets différents auront une distance de séparation plus grande

#### IV.1.1 Description de l'algorithme des K-means :

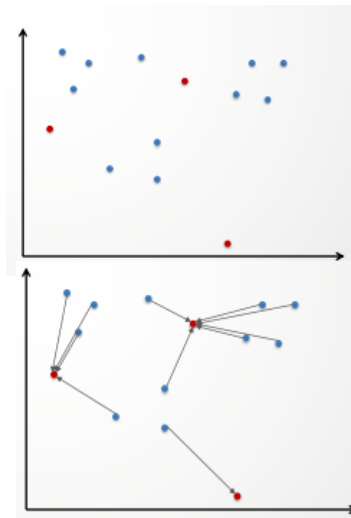
Pour décrire l'algorithme des K-means, on utilisera un exemple en 2D :

Imaginons que nous ayons des points (en bleu) dans un plan, et nous souhaitons faire des groupes.

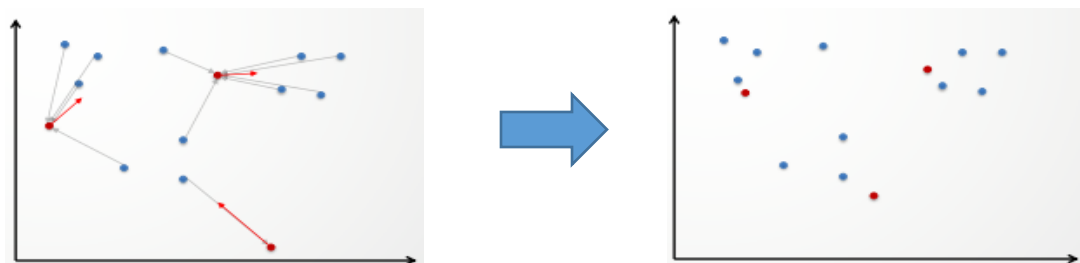
L'utilisateur doit décider la valeur de K, qui sera le nombre de groupes à construire (ici, on choisira K=3).

On commence par placer K points en rouge, au hasard dans le plan.

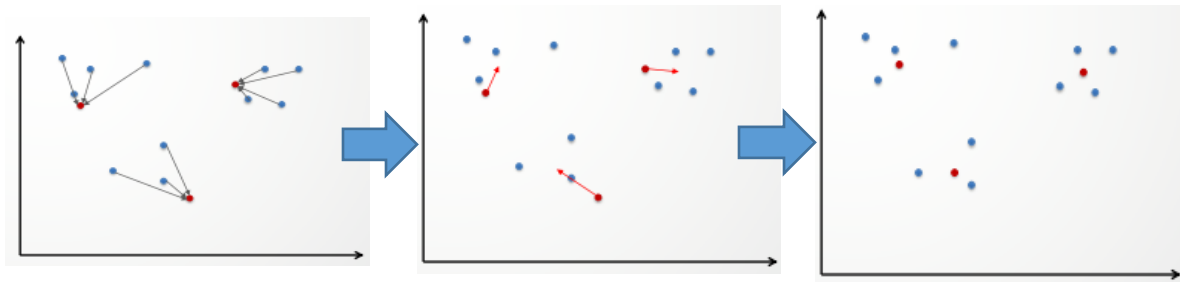
**Phase 1 :** On relie alors chaque points (bleu) de l'image aux points rouge qui lui est le plus proche.



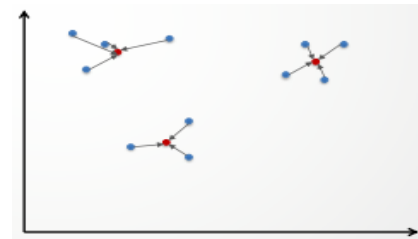
**Phase 2 :** On déplace chaque point rouge vers le barycentre des points avec lesquels elle est reliée.



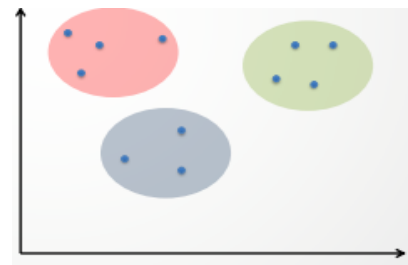
On itère ces deux étapes jusqu'à ce que chaque point rouge ne bouge plus



Une fois que les points rouges ne bougent plus, on relie une dernière fois chaque point au point rouge qui leur est le plus proche.



On regroupe ensemble les points rattachés au même point rouge.



#### IV.1.2 Choix du nombre de Clusters

Le choix du nombre de cluster  $K$  n'est pas forcément intuitif. Spécialement quand le jeu de données est grand et qu'on n'ait pas un a priori ou des hypothèses sur les données. Un nombre  $K$  grand peut conduire à un partitionnement trop fragmenté des données. Ce qui empêchera de découvrir des patterns intéressants dans les données. En revanche, un nombre de clusters trop petit, conduira à avoir, potentiellement, des clusters trop généralistes contenant beaucoup de données.

Dans le cas du projet, on fait le choix de  $K = 4$ , le nombre de classes déduit dans le point précédent

K-means clustering with 4 clusters of sizes 13, 10, 20, 8

Cluster means:

	T	TM	Tm	PP	V	RA	SN	TS	FG
1	0.0889064	0.3081096	0.06205413	0.2549638	-0.90394772	0.3441468	-0.6283096	-0.7113022	0.5051824
2	1.6450977	1.4737896	1.42567223	0.7434610	0.29813327	-1.4894334	-0.9253522	1.1136586	-1.2426222
3	-0.3953042	-0.4365704	-0.33755160	-0.7692619	0.43590326	0.3197564	0.3219636	0.1042926	-0.1985064
4	-1.2125845	-1.2514890	-1.03904925	0.5795124	0.00649031	0.5031622	1.3727844	-0.4969385	1.2286223

Clustering vector:

Dijon	Clermont-Ferrand	Cognac	Bourgon	Le puy	Lille
4	3	1	1	4	3
Lyon/Bon	Macon	Marseille/Marignane	Limoges	Le mans	Montpellier
3	1	2	4	3	2
Nantes	Nice	Nimes/Garon	Montauban	Melun	Orlans
1	2	2	1	3	3
Pau	Toulouse /Blagnac	St Dizier	Tours	Troyes	Strasbourg
3	3	3	3	3	3
Rouen	Poitiers	Perpignan	Paris/orly	Abbeville	Agen
4	3	2	3	3	1
Ajaccio	Albi	Amberieu	Alistro	Angers	Auch
2	1	1	2	1	1
Aurillac	Bastia	Beauvais	Bergerac	Brest	Brive
4	2	4	1	4	1
Bourges	Bordeaux/Merignac	Caen	Cannes	Chassiron	Calvi
3	1	3	2	3	2
Charleville	Carcassonne	Chateauroux			
4	3	3			

within cluster sum of squares by cluster:

[1] 43.50586 38.50786 67.34299 40.36065  
(between\_SS / total\_SS = 57.8 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss" "size" "iter"  
[9] "ifault"

Figure 11: Résultat de la fonction kmeans sous R

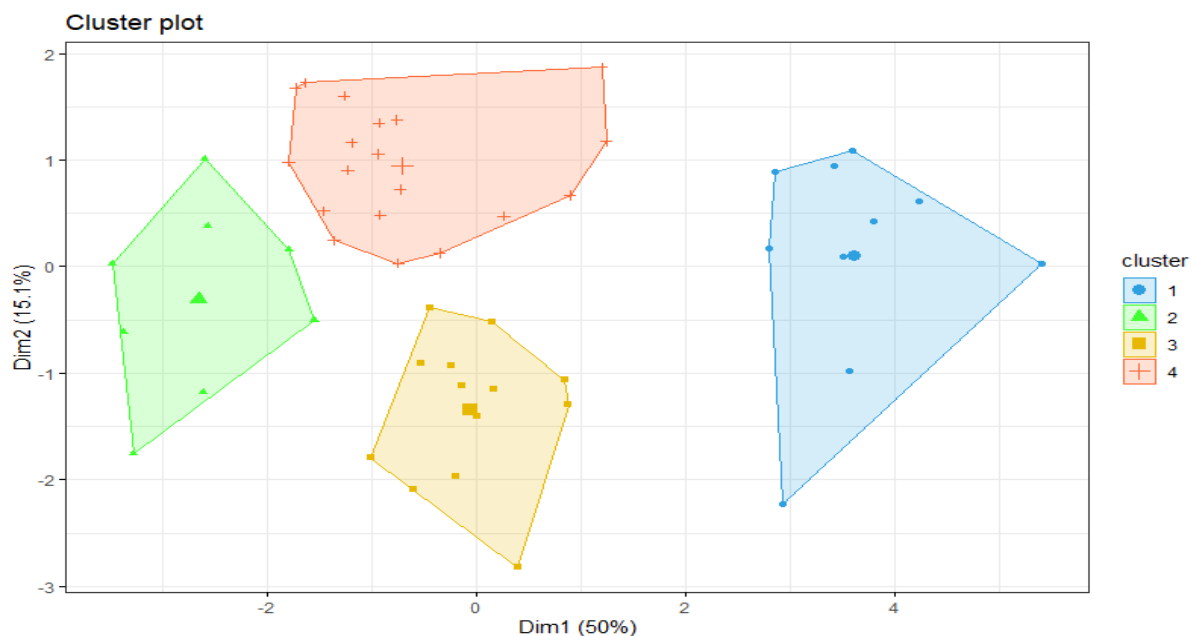


Figure 12: Classes obtenues avec la classification de k-means

#### IV.1.2.1 Description des classes obtenues

Afin de faire une description statistique des variables de chaque classe, nous allons faire une comparaison entre la moyenne de chaque classe et la moyenne générale de la population, en tenant compte de l'effectif par classe et de la dispersion des individus par classe (Ecart type).

Le tableau d'effectif est donné comment suit

Classes	1	2	3	4
Effectif	20	10	13	8

Tableau 4: Effectifs K-means

Le tableau des moyennes et écarts types de chaque classe est le suivant

Classes	T	TM	Tm	PP	V	RA	SN	TS	FG
Moyenne des variables par classe									
1	16.68	21.67	11.2	885.43	12.59	115.30	2.8	28.10	7.70
2	11.57	16.81	6.00	861.69	11.77	163.37	16.25	9.75	80.12
3	13.03	18.26	7.48	666.40	12.97	158.95	10.10	16.60	38.30
4	13.90	19.59	8.32	814.70	9.23	159.53	4.53	7.30	58.92
Moyennes générales									
	13.74	19.04	8.19	777.78	11.75	151.23	8.21	15.41	44.11
Ecart-type par classe									
1	0.63	0.49	1.69	136.59	2.64	23.12	1.47	8.77	7.94
2	0.77	0.76	1.05	154.54	2.55	19.93	5.80	7.99	29.52
3	0.86	1.06	1.47	72.93	2.38	12.05	4.15	8.78	10.61
4	0.52	0.70	0.84	120.40	2.13	14.63	2.47	9.72	24.10
Ecart types générales									
	1.78	1.78	2.10	144.78	2.79	24.12	5.85	11.39	29.30

Tableau 5: Moyennes – Ecart-types (K-means)

La classe 1 contient les villes ayant des valeurs faibles pour le nombre de jours annuel de neige (SN) et le nombre de jours annuel de brouillard (FG) car la moyenne de SN (resp. FG) pour cette classe est 2.8 (resp. 7.7) alors que la moyenne générale est de 8.21 (resp. 44.11) et les écarts types ne sont pas élevés ce qui veut dire que les individus ne sont pas très dispersés. Aussi dans la même classe les températures moyenne, maximale et minimale prennent des valeurs élevées par rapport à la moyenne générale. La valeur moyenne des totaux jours de tempête durant l'année (TS) est 28,10 alors que celle de la population totale est 15,41

Dans la classe 2 les villes ont des valeurs faibles pour le nombre de jours de tempête annuel (TS) dont la moyenne est 9.75 tandis qu'elle est de 15.41 pour la moyenne générale. Par contre les variables FG et SN prennent des valeurs plus élevées dans cette classe

Les classes 3 et 4 ont des températures moyennes, maximales et minimales qui sont quasiment proches aux températures moyennes générales sauf pour les variables PP, RA, FG et particulièrement les variables SN et TS pour la classe 4. Les écarts types de ces deux classes montrent que les individus de la classe 3 sont plus dispersés que les individus de la classe 4.

On constate également que pour les classes 3 et 4 on a la valeur moyenne de la précipitation totale annuelle de pluie et/ou neige fondue (PP) est 666.40 pour la classe 3 alors que celle de la



population totale est 777.78, par contre dans le cas de la classe 4 La valeur moyenne de (PP) est 814.70.

En conclusion, la classe 1 est caractérisée par les variables T, TM, Tm, SN et FG. La classe 2 est caractérisée par les variables SN, TS et FG. La classe 3 et 4 est caractérisée par les variables T, TM et Tm.

## IV.2 Classification Ascendante Hiérarchique (CAH)

Le principe de la CAH est de rassembler des individus selon un critère de ressemblance défini au préalable qui s'exprimera sous la forme d'une matrice de distances, exprimant la distance existante entre chaque individu pris deux à deux. Deux observations identiques auront une distance nulle. Plus les deux observations seront dissemblables, plus la distance sera importante. La CAH va ensuite rassembler les individus de manière itérative afin de produire un dendrogramme ou arbre de classification. La classification est ascendante car elle part des observations individuelles ; elle est hiérarchique car elle produit des classes ou groupes de plus en plus vastes, incluant des sous-groupes en leur sein.

La notion de ressemblance entre observations est évaluée par une distance entre individus. Plusieurs types de distances existent selon les données utilisées. Dans notre cas, on fait le choix de la distance euclidienne vu au cours avec la fonction (dist).

### IV.2.1 Calcul du dendrogramme

Plusieurs méthodes d'agrégation existent pour construire le dendrogramme (saut minimum, distance maximum, moyenne, Ward...). Chacune d'elle produit un dendrogramme différent. Techniquement la plus utilisée et celle que l'on utilise dans le projet est la méthode de Ward qui cherche à minimiser l'inertie intra-classe et à maximiser l'inertie inter-classe afin d'obtenir des classes les plus homogènes possibles. Cette méthode est souvent incorrectement présentée comme une méthode de minimisation de la variance alors qu'au sens strict Ward vise l'augmentation minimum de la somme des carrés ("minimum increase of sum-of-squares (of errors)"). La fonction de base pour le calcul d'un dendrogramme est hclust en précisant le critère d'agrégation avec method. Dans notre cas method = "ward. (Voir graph ci-dessous)

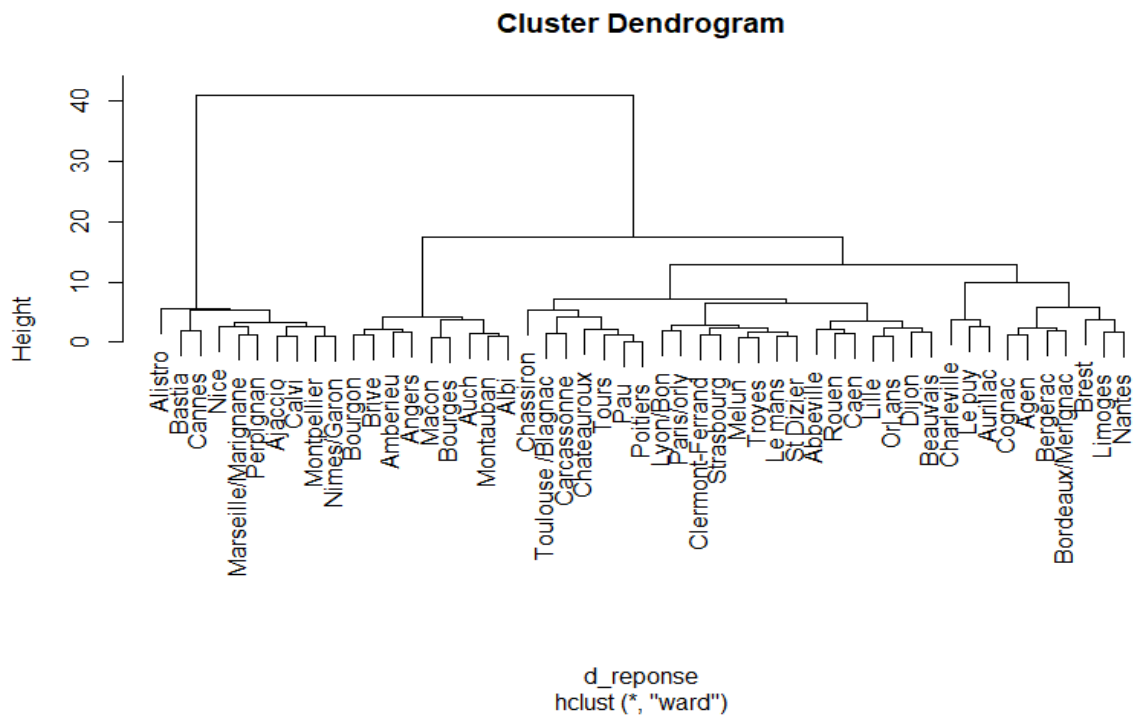


Figure 13: Représentation du dendrogramme CAH (Ward)

#### IV.2.2 Découpage du Dendrogramme

Pour obtenir une partition de la population, il suffit de découper le dendrogramme obtenu à une certaine hauteur. En premier lieu, une analyse de la forme du dendrogramme pourra nous donner une indication sur le nombre de classes à retenir. Dans notre exemple, à première vue, deux branches bien distinctes apparaissent sur l'arbre. Cependant en regardant de plus près le dendrogramme on remarque que le découpage peut bien se faire avec  $k$  équivalent à 3, 4, 5 ou même 6.

Pour nous aider, nous représentons les sauts d'inertie du dendrogramme selon le nombre de classes possibles

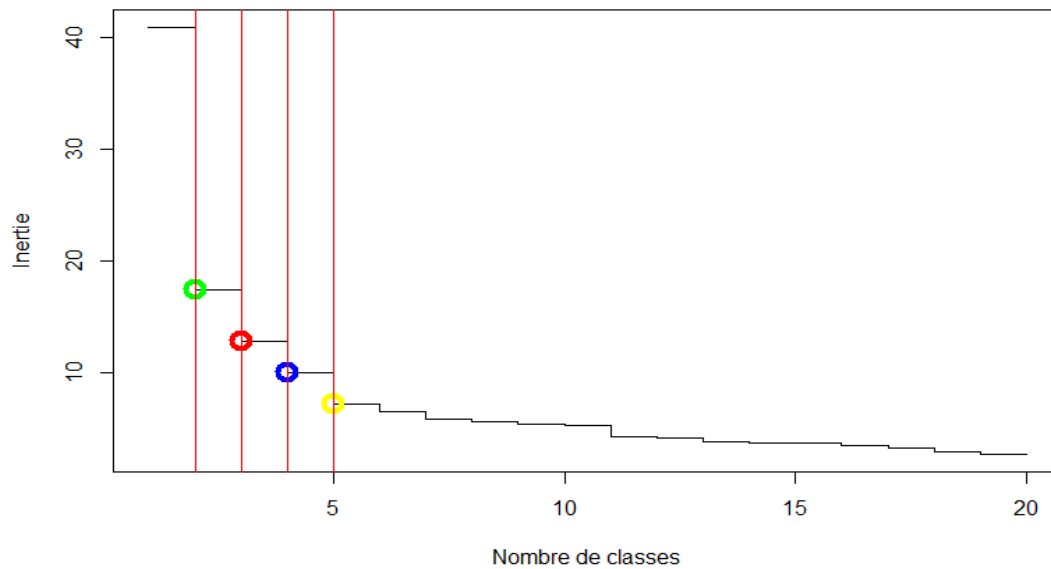


Figure 14: Représentations des sauts d'inertie

Comme dit précédemment on voit avec les sauts d'inertie que le dendrogramme peut être coupé en 2, 3, 4 et 5 groupes.

La fonction `rect.hclust` permet de visualiser les différentes partitions directement sur le dendrogramme.

### Partition en 2, 3, 4 ou 5 classes

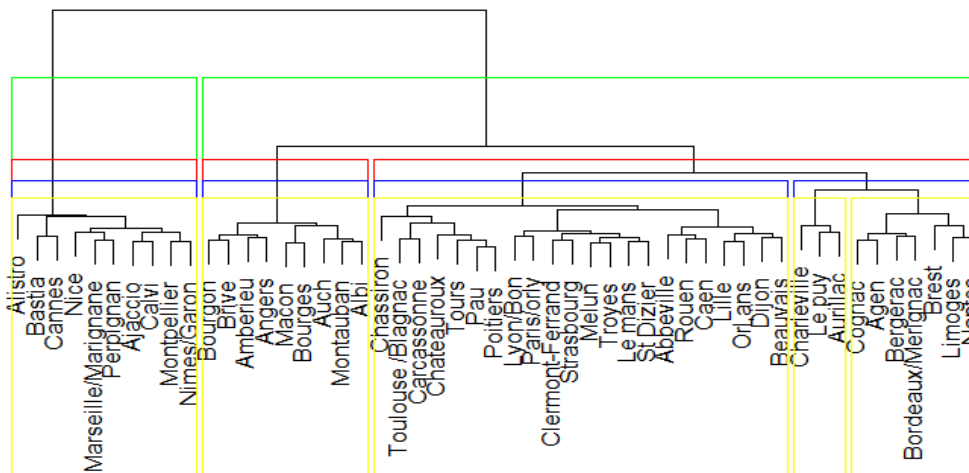


Figure 15: Représentation des différentes partitions possibles pour le dendrogramme

Ces k sont-ils tous optimales ? Comment choisir le k ayant la plus petite perte d’inertie inter classes. Pour répondre à cette question, nous nous référons à une fonction qui permet d’avoir le nombre optimal (qui minimise la perte d’inertie) du découpage de Dendrogramme, la fonction best.cutree. Le graphe ci-dessous illustre bien les différentes pertes d’inertie.

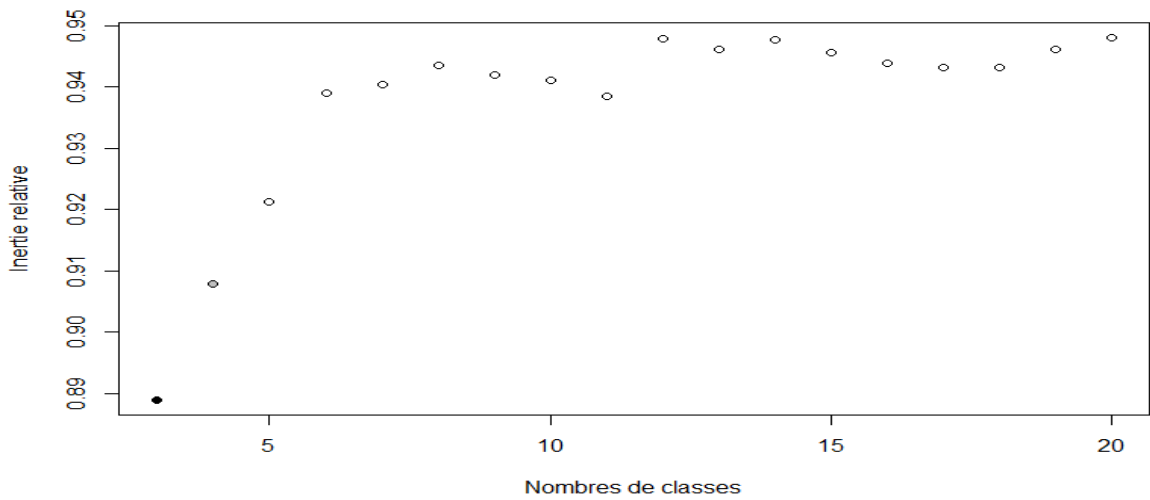


Figure 16: Illustration des différentes pertes d’inertie en fonction du nombre de classe

Ainsi le nombre optimal pour minimiser l’inertie intra-classe est 2, A cet effet le dendrogramme pour k =2 se représente comme suit

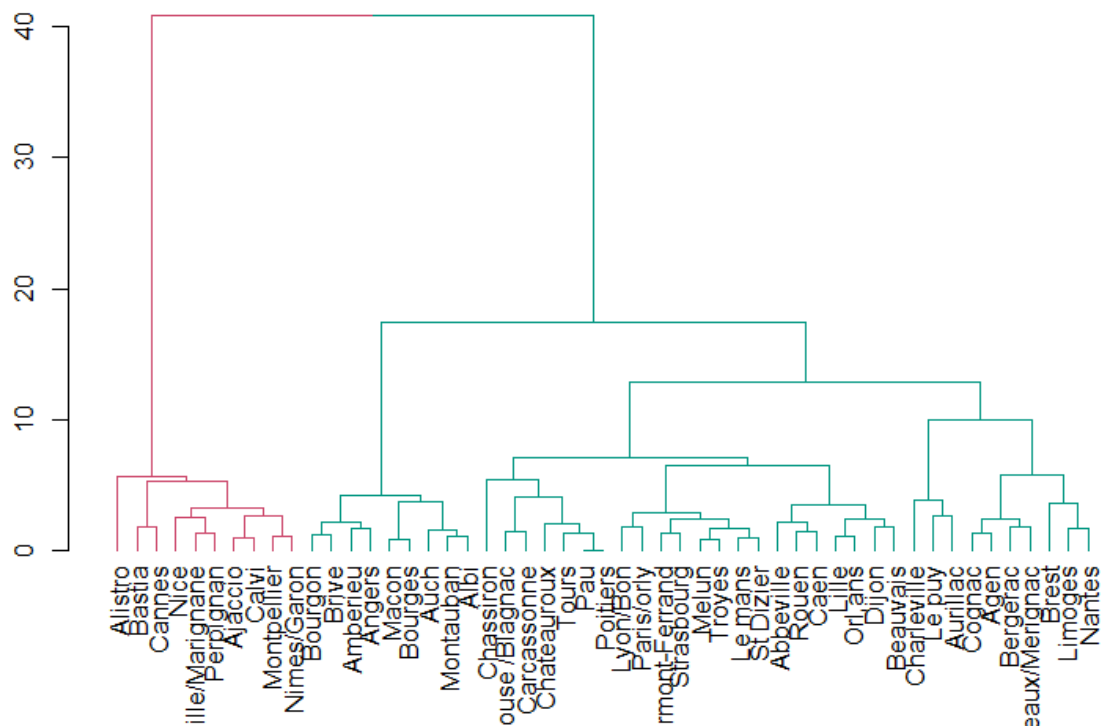


Figure 17: Représentation du dendrogramme avec le nombre optimal de classe pour minimiser l’inertie intra-classe

### IV.2.3 Description des partitions

Etant donné que le jeu de données comporte que des variables numériques, pour une description des individus de chaque classe nous comparerons la moyenne de la classe à la moyenne générale de la population, en tenant compte de l'effectif par classe et de la dispersion des individus par classe (Ecart-type) pour chacune des variables.

Classes	1	2
Effectif	41	10

Tableau 6: Effectifs (CAH)

Le tableau suivant comporte les moyennes et écart-types de chaque classe.

Classes	T	TM	Tm	PP	V	RA	SN	TS	FG
Moyenne par classe									
1	13.02	18.40	7.45	751.53	11.55	160.0	9.53	12.31	53.0
2	16.68	21.67	11.20	885.43	12.59	115.3	2.80	28.10	7.7
Moyennes générales									
	13.74	19.04	8.19	777.78	11.75	151.23	8.21	15.41	44.11
Ecart type par classe									
1	1.10	1.33	1.45	135.65	2.82	14.34	5.76	9.74	25.43
2	0.63	0.49	1.69	136.59	2.64	23.12	1.47	8.77	7.94
Ecart types générales									
	1.78	1.78	2.10	144.78	2.79	24.12	5.85	11.39	29.30

Tableau 7 : Moyennes – Ecart-types (CAH)

Comme vu précédemment 02 groupes distincts ressortent de notre classification.

Le groupe 2 où les températures moyennes (T) et minimales (Tm) excèdent largement les températures générales de la population. Le groupe 1 qui a pratiquement les mêmes températures que les températures générales

Le groupe 2 a pour effectif 10 et néanmoins sa température moyenne et minimale excèdent largement celles de toutes les observations alors que d'après les écart-type des deux variables, ses individus sont moins dispersés. Cela s'explique par le fait qu'ils prennent des valeurs élevées quant à ces variables.

La même remarque est faite pour le groupe 1. Il rassemble les 41 autres individus avec un écart-type faible pour les températures. Ainsi les valeurs des températures de ses individus tournent autour des valeurs moyennes générales de la population.

On note également que pour le groupe 2 que :

La valeur moyenne de jours de neige durant l'année (SN) est 2,80 alors que celle de la population totale est 8.21.

La valeur moyenne de jours de brouillard durant l'année (TS) est 28,10 alors que celle de la population totale est 15,41.

La valeur moyenne de jours de brouillard durant l'année (FG) est 7.7 alors que celle de la population totale est 44.11. On note ici que la dispersion des individus est considérable pour cette variable. Les valeurs varient entre 1 et 15 avec une seule valeur aberrante de 25 pour la ville de Nîmes/Garon

De cette analyse on peut conclure que ce sont les variables Températures moyenne T et minimale Tm, Nombre de jours de pluie RA, de neige SN et de tempête FG qui caractérisent la classe 2.

Rétrospectivement ces mêmes variables caractérisent la classe 1.

### IV.3 CAH avec la fonction HCPC

L'approche HCPC (Hierarchical Clustering on Principal Components ou Classification Hiérarchique sur Composantes Principales) nous permet de combiner trois méthodes standards utilisées dans les analyses de données multivariées (Husson, Josse, and J. 2010) :

Les Méthodes des composantes principales (ACP, AFC, ACM, AFDM, AFM), la classification ascendante hiérarchique et le partitionnement en k-moyennes

#### IV.3.1 Description de la fonction HCPC

L'algorithme de la méthode HCPC, tel qu'il est mis en œuvre dans le paquet FactoMineR, peut être résumé comme suit :

1. **Calcul des méthodes des composantes principales** : PCA, (M)CA ou MFA selon les types de variables de l'ensemble de données et la structure de l'ensemble de données. À cette étape, nous choisirons le nombre de dimensions à retenir dans la sortie en spécifiant l'argument ncp. La valeur par défaut est 5.
2. **Calcul du regroupement hiérarchique** : la classification hiérarchique est effectuée en utilisant le critère de Ward sur les composants principaux sélectionnés. Le critère de Ward est utilisé dans la mise en cluster hiérarchique car il est basé sur la variance multidimensionnelle comme l'analyse des composantes principales.
3. **Choix du nombre de grappes en fonction de l'arbre hiérarchique** : Un premier partitionnement est effectué en coupant l'arbre hiérarchique.
4. Effectuez la classification de K-means pour améliorer la partition initiale obtenue à partir de la classification hiérarchique. La solution finale de partitionnement, obtenue après consolidation avec k-means, peut-être (légèrement) différente de celle obtenue avec le regroupement hiérarchique.

En appliquant la méthode HCPC sous R on obtient le dendrogramme suivant

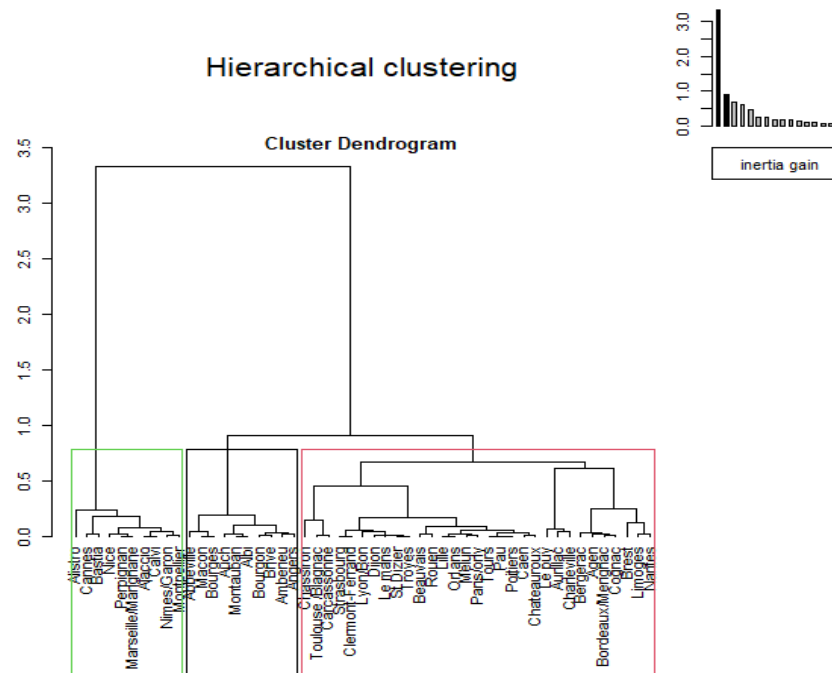


Figure 18: Dendrogramme (HCPC)

### IV.3.2 Visualisation des résultats en 3D

La visualisation en 3D du dendrogramme obtenu avec la méthode HCPC est donnée par le graphe au-dessous

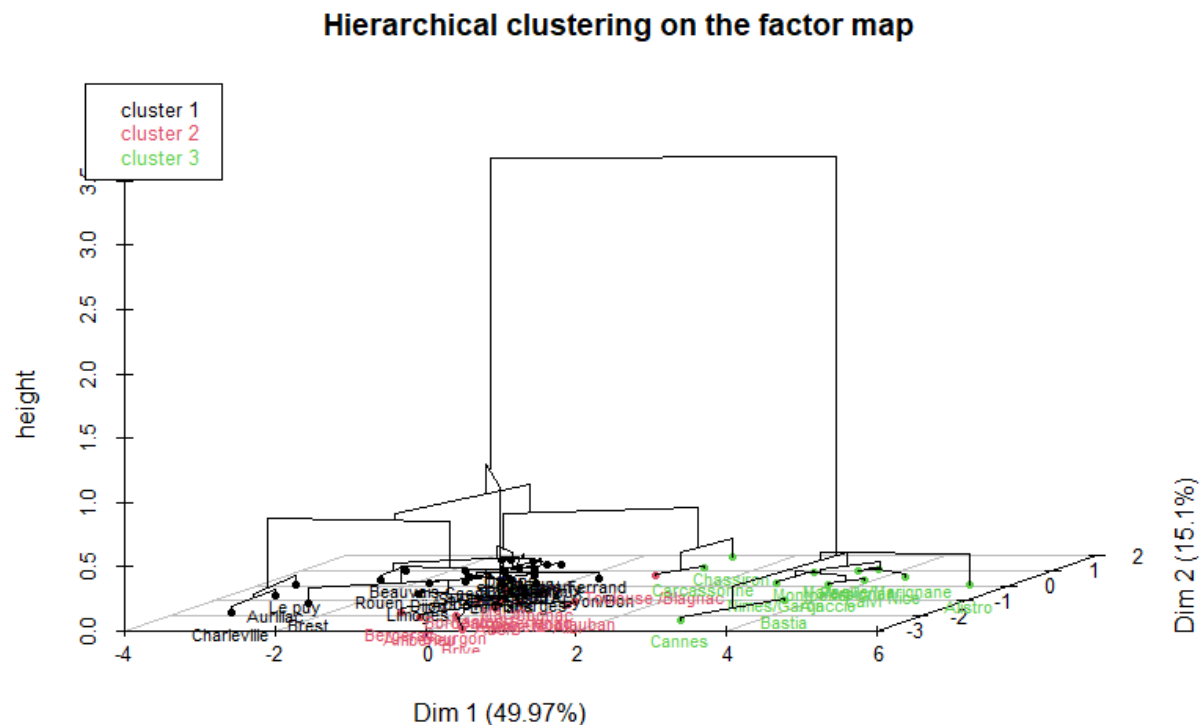


Figure 19: Dendrogramme en 3D (HCPC)

Les individus peuvent être visualisés et coloriés par groupe comme le montre la figure au-dessous

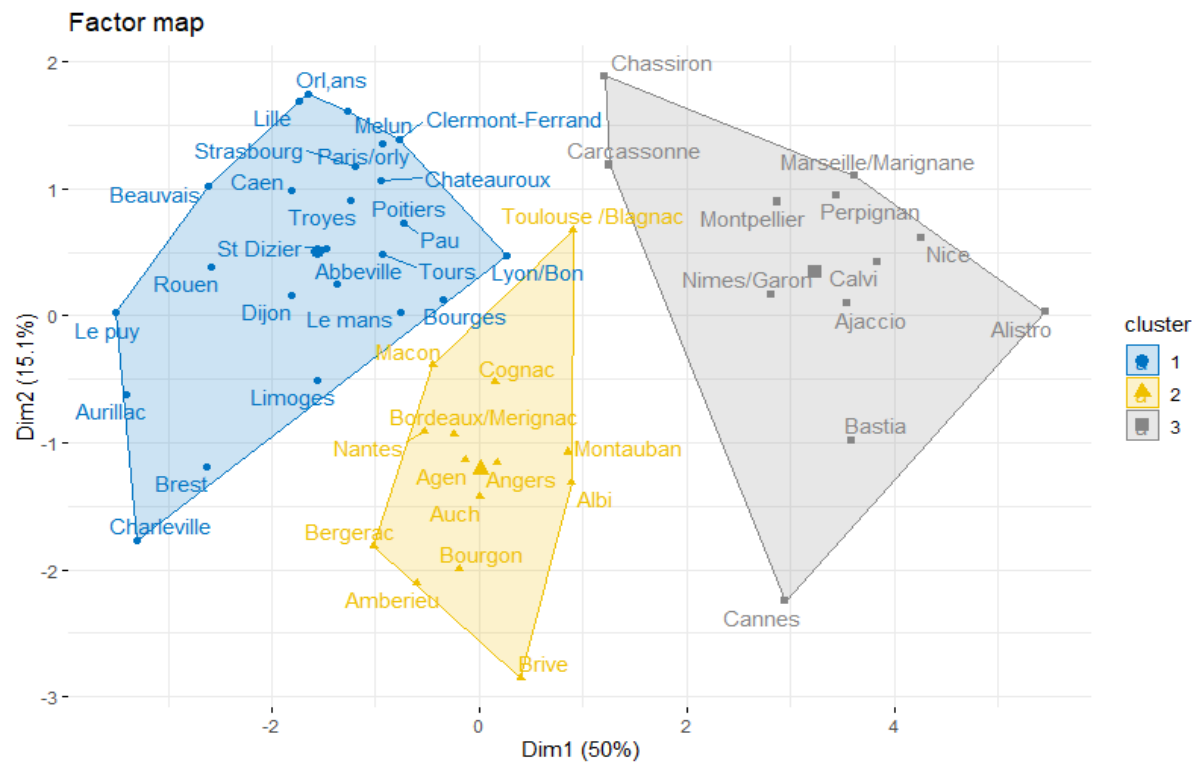


Figure 20: Visualisation des groupes d'individu obtenu à l'aide de la méthode HCPC

IV.3.3 Description des partitions

Contenu du résultat de la fonction HCPC () est donné

**\$data.clust** : Données d'origine avec une colonne supplémentaire appelée clust contenant les groupes.

**\$desc.var** : les variables décrivant les groupes

**\$desc.ind** : les individus les plus typiques de chaque groupe

**\$desc.axes** : les axes décrivant les groupes

Classes	1	2	3
Effectif	25	14	12

Tableau 8: Effectifs (HCPC)



## Variables décrivant le plus chaque classe

```
> res.hcpc$desc.var$quanti
$`1`
      v.test Mean in category Overall mean sd in category Overall sd      p.value
SN  5.580970      12.8800      8.215686      4.6157990      5.794906 2.391803e-08
RA  2.892230     161.2000     151.235294     13.8506318     23.889113 3.825181e-03
FG  2.189287      53.2800     44.117647     25.2039997     29.018349 2.857601e-02
PP -2.687310     722.2252     777.787059     139.2280249    143.359762 7.202999e-03
Tm -4.926208       6.7080      8.192157      1.0480153      2.088985 8.384082e-07
Tm -5.198421     17.7200     19.043137      1.0280078      1.764826 2.009890e-07
T  -5.304458     12.3880     13.741176      0.8496211      1.768812 1.130081e-07

$`2`
      v.test Mean in category Overall mean sd in category Overall sd      p.value
FG  1.984481     57.357143     44.117647     23.017407     29.018349 0.0472022664
TS -2.664944      8.500000     15.411765      9.976759     11.281038 0.0077001219
SN -2.788951      4.500000      8.215686      2.291288      5.794906 0.0052879118
V  -3.525247     9.514286     11.756863      2.222887      2.766980 0.0004230887

$`3`
      v.test Mean in category Overall mean sd in category Overall sd      p.value
T   5.692644     16.308333     13.741176      1.006196      1.768812 1.250873e-08
Tm  5.381593     11.058333      8.192157      1.626068      2.088985 7.382959e-08
TM  4.645458     21.133333     19.043137      1.460213      1.764826 3.393229e-06
TS  3.507595     25.500000     15.411765      9.928914     11.281038 4.521767e-04
V   2.447352     13.483333     11.756863      3.112564      2.766980 1.439101e-02
PP  2.271120     860.795833     777.787059     130.523692    143.359762 2.313972e-02
SN -3.643084      2.833333      8.215686      1.343710      5.794906 2.693913e-04
FG -4.667899      9.583333     44.117647      8.077317     29.018349 3.042954e-06
RA -5.087435    120.250000     151.235294     23.491577     23.889113 3.629378e-07
```

Du résultat ci-dessus, on constate que :

Les variables SN, RA, FG, PP, Tm, TM et T sont les plus significativement associées au cluster 1.

Les variables FG, TS, SN et V sont les plus significativement associées au cluster 2.

Les variables T, Tm, TM, TS, V, PP, SN, FG, RA sont les plus significativement associés au cluster 3

Les classes 1 et 2 ont différentes variables significatives. D'après l'analyse de l'image ci-dessus quant aux moyennes, on peut dire que la classe 1 est caractérisée par toutes les variables qui la décrivent. Les individus dans la classe prennent les valeurs moyennes des variables. Sauf la variable SN dont la moyenne est 12,88. On constate que les individus de la classe 1 prennent les valeurs maximales pour cette variable

Parmi les variables les plus significatives de la classe 2, on voit que SN et TS ont une moyenne largement inférieure à la moyenne générale. Ces deux variables caractérisent bien la classe et les individus dans cette classe prennent les petites valeurs pour ces deux variables

Aussi Les écarts-types en revanche montrent que dans la classe 1 les individus sont plus dispersés que dans la classe 2. Cela est essentiellement dû au nombre d'effectif qu'elle compose.

La 3ème classe s'apparente à la classe 2 de notre classification précédente. Elles comportent les mêmes individus à quelques différences près. Les variables qui caractérisent plus la classe 3 sont ainsi les variables Températures moyenne T maximale TM et minimale Tm, Nombre de jours de pluie RA, de neige SN et de tempête FG.

### Axes principaux associés au cluster

```
> res.hcpc$desc.axes$quant
$`1`
      v.test Mean in category Overall mean sd in category Overall sd      p.value
Dim.2  3.047095      0.5123401  1.864249e-15      0.8565158  1.165845 2.310645e-03
Dim.1 -5.110894     -1.5631131 -3.835712e-15      0.9520498  2.120615 3.206386e-07

$`2`
      v.test Mean in category Overall mean sd in category Overall sd      p.value
Dim.5 -3.465643     -0.6541807 -3.381010e-17      0.6087199  0.8210357 5.289643e-04
Dim.2 -4.514418     -1.2100248  1.864249e-15      0.8195145  1.1658447 6.349096e-06

$`3`
      v.test Mean in category Overall mean sd in category Overall sd      p.value
Dim.1  5.990642      3.238858 -3.835712e-15      1.124616  2.120615 2.090147e-09
```

Les résultats ci-dessus indiquent que les individus dans le groupe 1 ont des coordonnées élevées sur les axes 1 et 2. Les individus du groupe 2 ont des coordonnées élevées sur l'axe 2 et les individus appartenant au 3ème groupe ont des coordonnées élevées sur l'axe 1

### Individus représentatifs de chaque groupe

```
> res.hcpc$desc.ind$para
Cluster: 1
  Troyes      Pau Poitiers St Dizier      Dijon
0.8167394 0.9277131 0.9277131 1.0568530 1.0694338
-----
Cluster: 2
  Agen  Bourgon  Angers  Cognac  Albi
0.9783676 1.0959068 1.3169690 1.5208297 1.5592079
-----
Cluster: 3
Marseille/Marignane      Montpellier      Perpignan      Nice      Nimes/Garon
1.000831      1.030683      1.140735      1.207462      1.338148
> |
```

Pour chaque groupe, les 5 meilleurs individus les plus proches du centre du cluster sont affichés. Ces individus sont appelés paragones. La distance entre chaque individu et le centre du groupe est fournie. Par exemple, les individus représentatifs pour le groupe 1 sont les villes Troyes, Pau, Poitiers, St Dizier et Dijon

## V. Conclusion

Au terme de ces analyses nous voyons que la classification est un outil indispensable quant aux résolutions des problèmes, des équations de la vie. C'est en cela que le machine learning aujourd'hui propose différents types d'algorithmes pouvant aider à la prise de décision dans les entreprises. L'algorithme de K-means et de CAH sont ainsi de bonne méthode de regroupement d'individus.

Plus succinctement, notre analyse nous a permis de regrouper les 51 villes de la France en villes chaudes (Alistro, Bastia, Cannes, Nice, Perpignan, Montpellier, Calvi, Nîmes) et en ville tempérées. Pour la plupart des villes chaudes, le nombre de pluies et de neiges annuels ont des valeurs faibles. Elles ont donc un hiver doux et un été chaud. On peut aussi constater qu'elles sont situées pratiquement sur des régions côtières. Aussi, plus une ville est chaude, plus ses nombres de jours de pluie, de neige de brouillard sont faibles.

Aussi après les résultats obtenus dans l'application de chacune des algorithmes ci-dessus, on pense que la partition la plus pertinente est celle obtenue par l'algorithme de CAH avec la fonction HCPC.

## VI. Table d'illustration (Figures)

Figure 1: Corrélation des variables.....	6
Figure 2: Diagramme d'éboulis des valeurs propres .....	7
Figure 3: Qualité de représentation des variables sur les différents axes .....	8
Figure 4: Contribution de chaque variable sur le 2 <sup>ème</sup> axe factoriel.....	9
Figure 5: Contribution de chaque variable sur le 1 <sup>er</sup> axe factoriel .....	9
Figure 6: Contribution de chaque variable sur le 3 <sup>ème</sup> axe factoriel.....	9
Figure 8: Cercle de corrélation (1,3) .....	10
Figure 7: Cercle de corrélation (1,2) .....	10
Figure 10: Qualité de représentation des individus sur le plan factoriel (1,2) .....	11
Figure 11: Qualité de représentation des individus sur le plan factoriel (1,3).....	11
Figure 12: Résultat de la fonction kmeans sous R .....	15
Figure 13: Classes obtenues avec la classification de k-means .....	15
Figure 14: Représentation du dendrogramme CAH (Ward).....	18
Figure 15: Représentations des sauts d'inertie .....	19
Figure 16: Représentation des différentes partitions possibles pour le dendrogramme.....	19
Figure 17: Illustration des différentes pertes d'inertie en fonction du nombre de classe .....	20
Figure 18: Représentation du dendrogramme avec le nombre optimal de classe pour minimiser l'inertie intra-classe .....	20
Figure 19: Dendrogramme (HCPC).....	23
Figure 20: Dendrogramme en 3D (HCPC) .....	24
Figure 21: Visualisation des groupes d'individu obtenu à l'aide de la méthode HCPC.....	24

## VII. Table d'illustration (Tableau)

Tableau 1: Description des variables.....	4
Tableau 2: Analyses descriptives.....	5
Tableau 3: Matrice de corrélation.....	5
Tableau 4: Effectifs K-means .....	16
Tableau 5: Moyennes – Ecart-types (K-means).....	16
Tableau 6: Effectifs (CAH) .....	21
Tableau 7 : Moyennes – Ecart-types (CAH) .....	21
Tableau 8: Effectifs (HCPC) .....	24

## VIII. Annexe

## Base de données

	T	TM	Tm	PP	V	RA	SN	TS	FG	Kmean	Kc	lust	HCPC
Dijon	12.5	18.1	6.1	745.41	11.1	161	14	23	81	2		1	1
Clermont-Ferrand	12.9	18.9	6.2	513.07	10.0	168	10	31	27	3		1	1
Cognac	14.2	19.7	7.7	858.25	12.5	173	3	18	56	4		1	2
Bourgon	13.6	19.3	8.6	827.73	7.5	155	4	0	58	4		1	2
Le puy	10.1	15.7	4.7	784.99	8.9	139	24	0	39	2		1	1
Lille	12.0	16.8	7.0	596.53	14.9	156	12	16	56	3		1	1
Lyon/Bon	14.2	19.6	8.3	797.55	10.9	152	15	27	28	3		1	1
Macon	13.6	18.7	9.0	704.57	10.3	149	11	0	43	4		1	2
Marseille/Marignane	16.6	21.8	10.4	771.65	15.7	101	2	24	9	1		2	3
Limoges	12.5	17.3	7.8	1011.37	12.8	167	15	16	71	2		1	1
Le mans	13.3	18.6	7.0	749.21	9.8	167	11	20	37	3		1	1
Montpellier	16.1	21.2	9.9	865.58	14.6	120	6	33	15	1		2	3
Nantes	13.3	18.5	7.9	995.06	12.0	174	7	18	60	4		1	2
Nice	17.2	20.7	14.0	896.12	13.9	119	3	33	1	1		2	3
Nîmes/Garon	15.7	21.6	9.7	915.78	13.0	129	3	36	25	1		2	3
Montauban	14.7	20.4	9.5	632.71	8.0	135	4	0	44	4		1	2
Melun	12.6	18.1	6.2	567.94	12.5	156	14	21	34	3		1	1
Orléans	12.5	18.1	6.2	564.88	14.6	155	15	15	53	3		1	1
Pau	13.1	18.6	7.6	704.00	13.2	161	11	14	42	3		1	1
Toulouse /Blagnac	14.7	19.9	9.2	690.83	13.2	166	4	24	37	3		1	2
St Dizier	12.6	18.4	5.6	680.87	10.7	170	12	20	42	3		1	1
Tours	13.1	18.3	7.7	654.99	11.8	175	9	12	38	3		1	1
Troyes	12.6	18.4	6.4	677.76	12.0	157	14	20	44	3		1	1
Strasbourg	12.5	18.3	6.0	567.87	10.5	151	12	30	53	3		1	1
Rouen	11.6	16.6	6.1	792.30	13.9	177	13	9	68	2		1	1
Poitiers	13.1	18.6	7.6	704.00	13.2	161	11	14	42	3		1	1
Perpignan	16.8	21.8	11.3	758.67	15.6	130	1	24	12	1		2	3
Paris/orly	13.1	18.0	8.5	681.92	12.9	176	16	24	33	3		1	1
Abbeville	11.7	16.1	7.8	723.68	13.2	146	5	0	46	3		1	1
Agen	14.2	20.1	7.2	795.96	9.5	168	4	16	68	4		1	2
Ajaccio	16.7	22.0	10.4	802.57	9.7	123	3	38	4	1		2	3
Albi	14.5	20.2	9.3	747.17	8.3	147	3	0	28	4		1	2
Amberieu	13.1	19.0	7.8	992.61	9.1	155	7	0	60	4		1	2
Alistro	18.0	22.4	14.4	835.54	13.8	56	2	9	0	1		2	3
Angers	13.3	18.5	8.7	859.30	10.0	146	3	0	29	4		1	2
Auch	14.1	20.0	8.9	610.00	6.9	158	3	0	57	4		1	2
Aurillac	11.3	17.0	5.3	931.67	10.2	172	24	11	81	2		1	1
Bastia	16.6	21.5	11.6	1043.31	10.1	121	4	29	8	1		2	3
Beauvais	11.6	16.9	5.7	659.76	12.1	158	19	14	59	2		1	1
Bergerac	13.5	20.0	6.6	886.99	8.8	185	5	21	101	4		1	2
Brest	11.8	15.9	7.2	1133.96	16.3	197	7	5	107	2		1	1
Brive	14.2	20.4	8.5	900.05	5.2	152	3	0	51	4		1	2
Bourges	13.4	18.7	8.5	661.37	11.0	134	12	0	34	3		1	1
Bordeaux/Mérignac	14.4	19.9	8.5	780.73	11.9	177	2	22	111	4		1	2
Caen	11.8	16.5	6.8	681.51	14.5	166	8	10	51	3		1	1
Cannes	16.2	21.5	10.0	1187.88	8.0	139	3	21	2	1		2	3
Chassiron	14.1	16.7	11.9	730.12	19.6	132	2	6	18	3		1	3
Calvi	16.9	22.2	10.3	777.20	11.5	115	1	34	1	1		2	3
Charleville	11.2	17.0	5.1	834.08	8.9	136	14	0	135	2		1	1
Carcassonne	14.8	20.2	8.8	745.13	16.3	158	4	19	20	3		1	3
Chateauroux	12.6	18.5	6.3	634.94	14.7	172	5	9	31	3		1	1

```
> res_pca$ind$contrib
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dijon	1.426675e+00	3.810273e-02	7.187055e-01	1.50004057	1.829468e-01
Clermont-Ferrand	2.561517e-01	2.784456e+00	2.840439e-01	8.79131140	1.547784e+00
Cognac	1.008780e-02	3.970870e-01	2.492549e+00	0.00631197	2.394293e+00
Bourgon	1.662485e-02	5.713973e+00	1.709390e+00	0.06001598	7.784557e-01
Le puy	5.369680e+00	9.385605e-04	5.351637e+00	0.01249225	1.736436e+01
Lille	1.325302e+00	4.121412e+00	3.679614e-02	0.48573437	1.381347e-01
Lyon/Bon	3.096254e-02	3.297546e-01	2.199127e-01	2.72179599	2.942197e+00
Macon	8.725847e-02	2.188134e-01	3.655380e+00	0.39509562	1.547618e-02
Marseille/Marignane	5.722513e+00	1.751437e+00	2.524335e-02	0.53106015	8.555531e-02
Limoges	1.063407e+00	3.757474e-01	4.384598e+00	0.83934222	4.533996e+00
Le mans	2.484127e-01	1.357087e-03	1.515130e-03	1.97971717	2.484941e-02
Montpellier	3.609546e+00	1.163143e+00	1.526595e+00	0.08398884	1.440579e+00
Nantes	1.227905e-01	1.198373e+00	4.790244e+00	0.23636594	4.863102e-02
Nice	7.924432e+00	5.476263e-01	1.088543e+00	0.41393895	1.253178e+00
Nîmes/Garon	3.460636e+00	4.108299e-02	3.232837e+00	0.74667092	3.069535e-01
Montauban	3.194900e-01	1.658031e+00	8.521587e+00	0.02676125	2.050461e+00
Melun	7.088553e-01	3.776757e+00	5.337796e-01	1.45957631	4.617419e-03
Orléans	1.182169e+00	4.374672e+00	3.090071e-01	0.01660766	1.819794e-02
Pau	2.358075e-01	7.694168e-01	5.718736e-03	0.01923946	3.147348e-02
Toulouse /Blagnac	3.558425e-01	6.539726e-01	5.811396e-01	0.16486042	3.428333e+00
St Dizier	9.474485e-01	4.012358e-01	1.142303e-03	2.47683951	4.430967e-02
Tours	3.756630e-01	3.459099e-01	1.931073e-01	0.03036577	1.430952e+00
Troyes	6.689710e-01	1.201325e+00	2.083786e-02	1.04398143	2.775745e-01
Strasbourg	6.304025e-01	1.982275e+00	1.066838e-01	5.47899873	5.449090e-02
Rouen	2.919891e+00	2.160995e-01	7.308035e-01	1.22280201	1.571925e-02
Poitiers	2.358075e-01	7.694168e-01	5.718736e-03	0.01923946	3.147348e-02
Perpignan	5.180156e+00	1.299477e+00	2.461161e-01	0.44457727	9.504759e-01
Paris/orly	3.854130e-01	2.661447e+00	3.378053e-01	0.66716004	2.885248e-01
Abbeville	8.189869e-01	8.713137e-02	1.546956e+00	6.25753400	2.727645e-01
Agen	7.746848e-03	1.848236e+00	3.005048e-01	1.05465393	2.842971e+00
Ajaccio	5.463005e+00	1.355815e-02	8.034069e-02	5.43379000	1.679873e-01
Albi	3.435873e-01	2.470976e+00	4.574349e+00	0.02233596	1.215848e+00
Amberieu	1.582293e-01	6.408261e+00	1.385222e-02	0.86514524	4.326530e-01
Alistro	1.298142e+01	1.291358e-03	6.182082e+00	5.52059545	4.256464e+00
Angers	1.283091e-02	1.936764e+00	1.565475e+00	1.88949202	2.997450e-02
Auch	1.878965e-05	2.908493e+00	6.371682e+00	0.39082603	6.595316e+00
Aurillac	5.046659e+00	5.566854e-01	8.379326e-01	0.34911974	1.043391e+01
Bastia	5.634407e+00	1.395276e+00	9.956557e-01	0.32464710	4.088021e+00
Beauvais	2.990623e+00	1.491786e+00	2.918351e-01	0.26419345	1.813121e+00
Bergerac	4.513875e-01	4.740982e+00	4.305043e+00	2.30108943	3.257567e+00
Brest	3.022451e+00	2.031496e+00	1.592614e+01	13.71311387	5.390278e-02
Brive	7.040518e-02	1.171644e+01	2.161473e+00	0.31710235	3.078022e-01
Bourges	5.491259e-02	2.404671e-02	6.290310e+00	0.45513044	4.271341e-01
Bordeaux/Mérignac	2.535546e-02	1.264590e+00	4.581705e+00	0.08385433	9.118777e+00
Caen	1.425926e+00	1.422259e+00	2.434324e-04	1.98860561	8.486309e-01
Cannes	3.799114e+00	7.283703e+00	1.499028e+00	0.45191694	4.631015e+00
Chassiron	6.399048e-01	5.143699e+00	3.989408e-02	22.57030028	3.445377e-01
Calvi	6.406280e+00	2.596014e-01	7.096642e-04	2.19790325	1.435077e-06
Charleville	4.752127e+00	4.516162e+00	4.138768e-01	0.32617277	1.163769e+00
Carcassonne	6.833211e-01	2.040131e+00	8.325413e-01	0.75202583	1.656899e+00
Chateauroux	3.909006e-01	1.645089e+00	7.693101e-02	0.59556031	4.356943e+00

```
> |
```



```
> res_pca$ind$cos2
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dijon	6.391073e-01	5.158963e-03	6.891819e-02	0.1396985406	1.228496e-02
Clermont-Ferrand	7.532117e-02	2.474673e-01	1.787884e-02	0.5374197801	6.822293e-02
Cognac	9.221539e-03	1.097110e-01	4.877359e-01	0.0011995324	3.280836e-01
Bourgon	7.327923e-03	7.612351e-01	1.612866e-01	0.0054995873	5.143482e-02
Le puy	5.790436e-01	3.059023e-05	1.235332e-01	0.0002800553	2.806868e-01
Lille	4.722892e-01	4.439120e-01	2.806917e-03	0.0359858858	7.378982e-03
Lyon/Bon	2.351561e-02	7.569516e-02	3.575231e-02	0.4297494877	3.349590e-01
Macon	7.048434e-02	5.342155e-02	6.320505e-01	0.0663479271	1.873912e-03
Marseille/Marignane	8.296768e-01	7.674928e-02	7.834357e-04	0.0160068526	1.859385e-03
Limoges	3.500579e-01	3.738469e-02	3.089613e-01	0.0574406580	2.237286e-01
Le mans	3.026022e-01	4.996475e-04	3.950777e-04	0.5013506765	4.537480e-03
Montpellier	7.762470e-01	7.560282e-02	7.027574e-02	0.0037549922	4.643921e-02
Nantes	7.439413e-02	2.194439e-01	6.212490e-01	0.0297713768	4.416594e-03
Nice	8.660476e-01	1.808904e-02	2.546558e-02	0.0094048101	2.052989e-02
Nimes/Garon	7.619929e-01	2.734102e-03	1.523748e-01	0.0341793988	1.013136e-02
Montauban	1.046473e-01	1.641425e-01	5.974828e-01	0.0018222895	1.006752e-01
Melun	3.109005e-01	5.006572e-01	5.011410e-02	0.1330855036	3.035732e-04
Orl,ans	4.210506e-01	4.709317e-01	2.355903e-02	0.0012297119	9.715771e-04
Pau	4.394548e-01	4.333870e-01	2.281344e-03	0.0074540038	8.792281e-03
Toulouse /Blagnac	2.763463e-01	1.535016e-01	9.660743e-02	0.0266165874	3.990976e-01
St Dizier	5.563213e-01	7.120776e-02	1.435773e-04	0.3023486676	3.900038e-03
Tours	3.842873e-01	1.069494e-01	4.228540e-02	0.0064577620	2.194234e-01
Troyes	5.114878e-01	2.776171e-01	3.410479e-03	0.1659437574	3.181324e-02
Strasbourg	2.289907e-01	2.176313e-01	8.295308e-03	0.4137530096	2.967044e-03
Rouen	8.541515e-01	1.910643e-02	4.576179e-02	0.0743642904	6.892875e-04
Poitiers	4.394548e-01	4.333870e-01	2.281344e-03	0.0074540038	8.792281e-03
Perpignan	8.649099e-01	6.557738e-02	8.796344e-03	0.0154317566	2.378863e-02
Paris/only	1.870440e-01	3.903847e-01	3.509280e-02	0.0673112302	2.098943e-02
Abbeville	2.766694e-01	8.896442e-03	1.118655e-01	0.4394682687	1.381250e-02
Agen	5.854884e-03	4.221904e-01	4.861595e-02	0.1657079428	3.220816e-01
Ajaccio	8.118809e-01	6.090022e-04	2.555821e-03	0.1678818257	3.742291e-03
Albi	1.428112e-01	3.104210e-01	4.069944e-01	0.0019300553	7.575381e-02
Amberieu	6.515744e-02	7.975812e-01	1.221042e-03	0.0740638485	2.670651e-02
Alistro	7.706860e-01	2.317178e-05	7.856409e-02	0.0681367131	3.787951e-02
Angers	7.471921e-03	3.408860e-01	1.951439e-01	0.2287494354	2.616538e-03
Auch	5.486438e-06	2.566832e-01	3.982539e-01	0.0237244080	2.886740e-01
Aurillac	6.983582e-01	2.328315e-02	2.482090e-02	0.0100435854	2.164320e-01
Bastia	8.010578e-01	5.995611e-02	3.030114e-02	0.0095954924	8.712224e-02
Beauvais	7.725181e-01	1.164692e-01	1.613685e-02	0.0141876065	7.020596e-02
Bergerac	1.149108e-01	3.647853e-01	2.345973e-01	0.1217825364	1.243097e-01
Brest	3.049412e-01	6.194845e-02	3.439550e-01	0.2876292263	8.152076e-04
Brive	1.655006e-02	8.324314e-01	1.087625e-01	0.0154965361	1.084593e-02
Bourges	3.344195e-02	4.426220e-03	8.200227e-01	0.0576229734	3.899277e-02
Bordeaux/Merignac	6.950780e-03	1.047778e-01	2.688582e-01	0.0047788970	3.747131e-01
Caen	5.610783e-01	1.691464e-01	2.050401e-05	0.1626728491	5.005476e-02
Cannes	4.821855e-01	2.794098e-01	4.072637e-02	0.0119242357	8.810657e-02
Chassiron	8.946347e-02	2.173519e-01	1.193914e-03	0.6560059447	7.220502e-03
Calvi	9.075267e-01	1.111522e-02	2.151992e-05	0.0647294870	3.047397e-08
Charleville	5.478548e-01	1.573637e-01	1.021369e-02	0.0078174521	2.011151e-02
Carcassonne	3.120752e-01	2.816113e-01	8.139062e-02	0.0714014905	1.134307e-01
Chateauroux	1.776528e-01	2.259710e-01	7.484131e-03	0.0562693152	2.968164e-01

```
> |
```

```
> res_pca$var
```

```
$coord
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T	0.9834393	-0.06389934	0.059229455	0.06806077	-0.05519658
TM	0.8891562	-0.18999914	0.004236464	0.32733085	-0.12137582
Tm	0.9121588	-0.03746051	-0.041132669	-0.29182126	0.01822712
PP	0.2416396	-0.63417877	0.540830398	-0.24507000	0.41274089
V	0.1177526	0.72156262	0.369977135	-0.55485632	-0.05541513
RA	-0.7426251	-0.08378377	0.399601861	0.13279701	-0.38146677
SN	-0.7340925	0.26352802	-0.066450348	0.21770390	0.54528001
TS	0.4463909	0.42053724	0.543469671	0.54870914	0.07201966
FG	-0.7384579	-0.37610383	0.261861928	-0.06263632	-0.18581898

```
$cor
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T	0.9834393	-0.06389934	0.059229455	0.06806077	-0.05519658
TM	0.8891562	-0.18999914	0.004236464	0.32733085	-0.12137582
Tm	0.9121588	-0.03746051	-0.041132669	-0.29182126	0.01822712
PP	0.2416396	-0.63417877	0.540830398	-0.24507000	0.41274089
V	0.1177526	0.72156262	0.369977135	-0.55485632	-0.05541513
RA	-0.7426251	-0.08378377	0.399601861	0.13279701	-0.38146677
SN	-0.7340925	0.26352802	-0.066450348	0.21770390	0.54528001
TS	0.4463909	0.42053724	0.543469671	0.54870914	0.07201966
FG	-0.7384579	-0.37610383	0.261861928	-0.06263632	-0.18581898

```
$cos2
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T	0.96715285	0.004083126	3.508128e-03	0.004632269	0.0030466623
TM	0.79059881	0.036099672	1.794763e-05	0.107145488	0.0147320887
Tm	0.83203375	0.001403290	1.691896e-03	0.085159647	0.0003322279
PP	0.05838971	0.402182716	2.924975e-01	0.060059305	0.1703550383
V	0.01386568	0.520652615	1.368831e-01	0.307865537	0.0030708366
RA	0.55149204	0.007019720	1.596816e-01	0.017635046	0.1455168949
SN	0.53889177	0.069447016	4.415649e-03	0.047394990	0.2973302941
TS	0.19926482	0.176851574	2.953593e-01	0.301081723	0.0051868321
FG	0.54532000	0.141454095	6.857167e-02	0.003923308	0.0345286917

```
$contrib
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T	21.5065782	0.3004079	0.364432853	0.4954842	0.45196029
TM	17.5805460	2.6559621	0.001864443	11.4606691	2.18544700
Tm	18.5019347	0.1032443	0.175758295	9.1089841	0.04928469
PP	1.2984121	29.5897987	30.385348997	6.4241606	25.27149501
V	0.3083311	38.3059874	14.219745160	32.9304120	0.45554644
RA	12.2635285	0.5164621	16.588115305	1.8863083	21.58685483
SN	11.9833365	5.1094270	0.458708257	5.0695396	44.10777114
TS	4.4310518	13.0115051	30.682635999	32.2047907	0.76944599
FG	12.1262810	10.4072055	7.123390691	0.4196513	5.12219462