

09/03/2021

# Mimi Projet de Processus de Data Maning

Classification Non Supervisée

Température  
Histogramme  
Correlation  
décembre Dendogramme  
juillet janvier septembre  
février mai  
Moyenne octobre avril  
**Data Maning**  
ACP juin mars août Biplot  
novembre  
CLASSIFICATION  
Ecart-type  
Kmeans Heatmap

I. INTRODUCTION.....	2
II. PROBLEMATIQUE .....	2
III. EXPLORATION DES VARIABLES .....	2
IV. ETUDE DE LA CORRELATION .....	4
V. ANALYSE DES COMPOSANTES PRINCIPALES.....	4
VI. CLASSIFICATION HIERARCHIQUE ASCENDANTE.....	7
VII. CLASSIFICATION PAR LA METHODE DU KMEANS.....	8
VIII. CONCLUSION .....	9
IX. ANNEXES.....	10

## LISTE DES GRAPHIQUES

<b>Figure 1:</b> HEATMAP des temperatures .....	4
<b>Figure 2:</b> Graphe des individus (ACP) .....	5
<b>Figure 3:</b> Graphe des variables (ACP).....	5
<b>Figure 4:</b> Arbre hiérarchique.....	7
<b>Figure 5:</b> Classification Ascendante Hiérarchique des individus (villes).....	8
<b>Figure 6:</b> Résultat de kmeans .....	9

## LISTE DES ANNEXES

<b>ANNEXE 1:</b> Correlation .....	10
<b>ANNEXE 2:</b> Biplot.....	10
<b>ANNEXE 3:</b> DENDOGRAMME.....	11
<b>ANNEXE 4 :</b> CODE R.....	13

## I. INTRODUCTION

Le jeu de données est constitué de 36 villes de France en ligne et en colonne les températures mensuelles moyennes. Ces températures mensuelles moyennes ont été calculées sur 30 ans. Par exemple à Ajaccio en janvier la température moyenne vaut 7,7 degrés. Cette valeur de 7,7 est la moyenne sur tous les jours de janvier pendant 30 ans. On a ainsi 12 variables qui correspondent aux 12 mois de l'année.

ville	jan	fev	mars	avril	mai	juin	juil	aout	sept	oct	nov	dec
ajac	7.7	8.7	10.5	12.6	15.9	19.8	22.0	22.2	20.3	16.3	11.8	8.7
ange	4.2	4.9	7.9	10.4	13.6	17.0	18.7	18.4	16.1	11.7	7.6	4.9
ango	4.6	5.4	8.9	11.3	14.5	17.2	19.5	19.4	16.9	12.5	8.1	5.3
besa	1.1	2.2	6.4	9.7	13.6	16.9	18.7	18.3	15.5	10.4	5.7	2.0
biar	7.6	8.0	10.8	12.0	14.7	17.8	19.7	19.9	18.5	14.8	10.9	8.2
bord	5.6	6.6	10.3	12.8	15.8	19.3	20.9	21.0	18.6	13.8	9.1	6.2

## II. PROBLEMATIQUE

Les travaux qui seront réalisés durant ce mini-projet auront pour objectif de chercher premièrement les caractéristiques des différents groupes de villes et dans une deuxième étape on cherche à regrouper des villes qui ont des profils météo similaires.

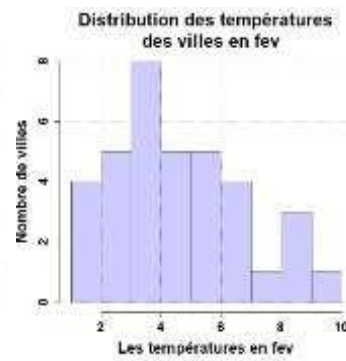
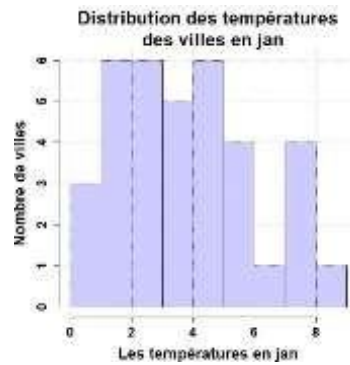
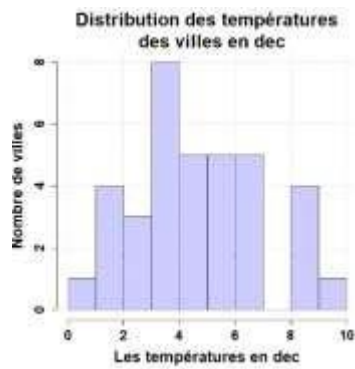
## III. EXPLORATION DES VARIABLES

	mean	sd	median	min	max	Q0.25	Q0.75
jan	3.836111	2.251876	3.45	0.4	8.6	2.075	5.525
fev	4.677778	2.146507	4.25	1.5	9.1	3.125	6.000
mars	8.013889	1.722980	7.70	5.5	11.3	6.875	9.375
avril	10.772222	1.456142	10.40	8.9	13.9	9.675	11.700
mai	14.250000	1.404788	13.90	11.6	17.1	13.300	14.900
juin	17.605556	1.708792	17.20	14.4	21.1	16.575	18.550
juil	19.608333	1.992611	19.10	15.6	23.8	18.400	20.750
aout	19.322222	1.934368	18.75	16.0	23.3	18.125	20.300
sept	16.775000	1.910479	16.15	14.7	20.5	15.300	18.350
oct	12.144444	2.037568	11.45	9.4	16.5	10.650	13.425
nov	7.752778	2.097275	7.15	4.6	12.6	6.400	9.025
dec	4.683333	2.282042	4.30	0.5	9.7	3.100	6.275

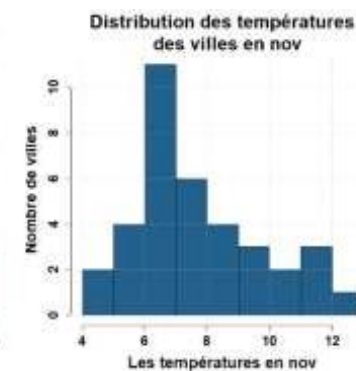
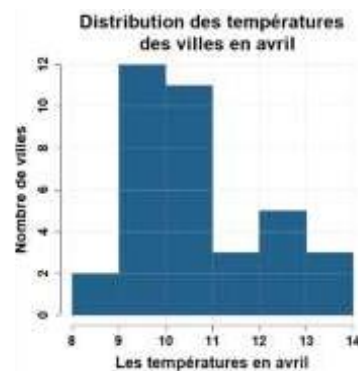
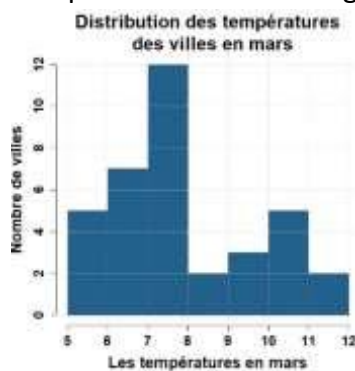
Les mois de janvier, février et décembre sont les mois où les températures sont généralement basses. Elles sont en dessous de 10 degrés.

Mars, avril et novembre sont des mois tempérés avec des températures comprises entre 11 et 13 degrés.

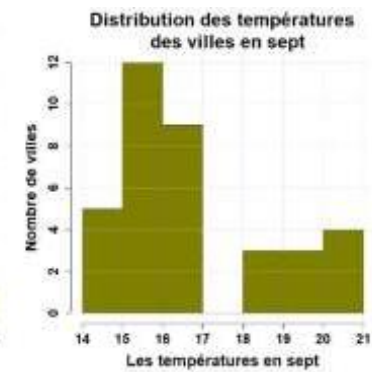
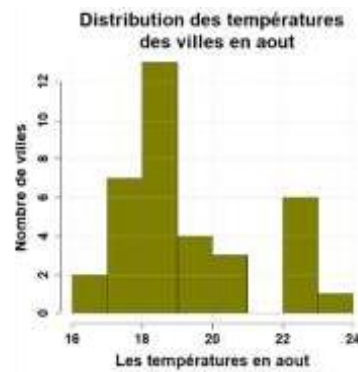
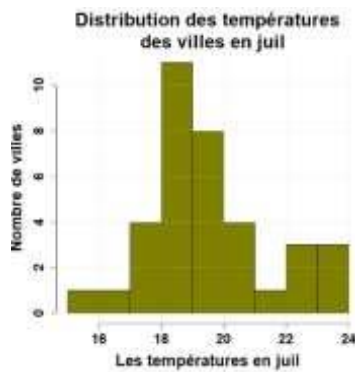
Et enfin les mois de mai, juin, juillet, août et septembre ont des températures élevées allant jusqu'à 24 degrés.



En décembre, janvier et février les villes de France enregistrent des températures moyennes comprises entre 1 et 7 degrés.



En mars, avril et novembre elles enregistrent des températures moyennes comprises entre 6 et 11 degrés.



Enfin les mois de juillet, aout et septembre sont les mois où les villes de France enregistrent leur plus haute température.

#### IV. ETUDE DE LA CORRELATION

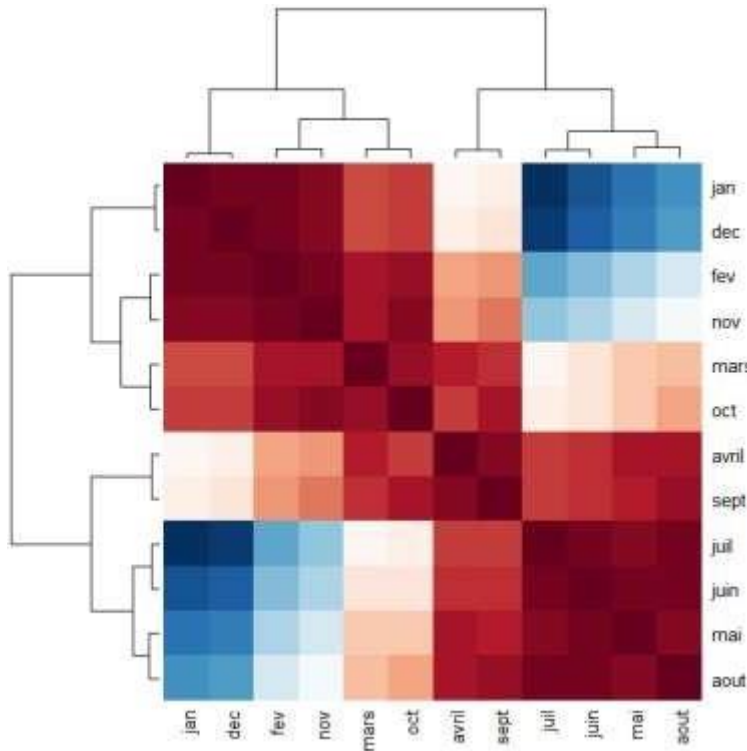
---

Nous pouvons constater deux groupes qui se forment sur l'heatmap :

un groupe composé des mois d'avril, septembre, juin, juillet, mai et août dont les corrélations sont très fortes entre eux ;

un autre groupe composé de janvier, décembre, février, mars, novembre et octobre avec des corrélations très proches de 1

Cette liaison indique que les groupes de variables du jeu de données apportent quasiment la même information



**FIGURE 1:** HEATMAP DES TEMPERATURES

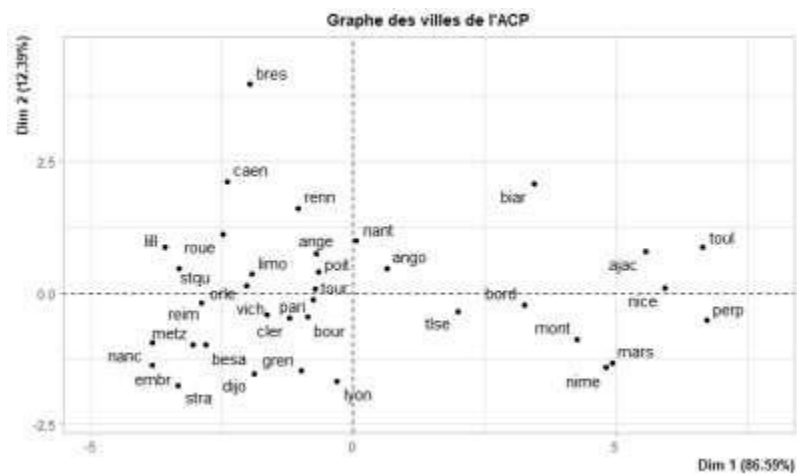
Les mois de janvier, Décembre, Février et Novembre ont des températures froides.

Les mois de Mai, Juin, Juillet et Aout ont des températures chaudes.

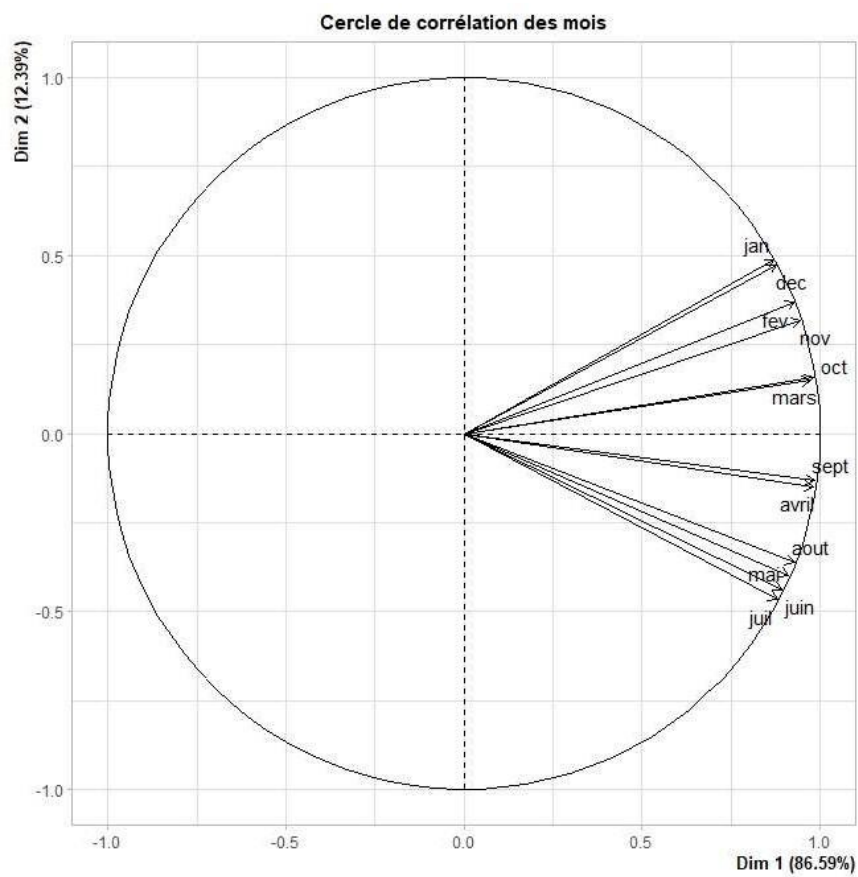
Les mois Mars, Avril, Septembre et Octobre ont des températures tempérées.

#### V. ANALYSE DES COMPOSANTES PRINCIPALES

---



**FIGURE 2: GRAPHE DES INDIVIDUS (ACP)**



**FIGURE 3: GRAPHE DES VARIABLES (ACP)**

La **dimension 1** oppose des individus tels que Perpignan, Montpellier, Nice, Toulouse, Nîmes, Marseille, Biarritz, Bordeaux, Angoulême, Nantes et Ajaccio (à droite du graphique, caractérisés par une coordonnée fortement positive sur l'axe) à des individus comme Brest, Metz, Nancy, Lyon, Lille, Dijon, Strasbourg, Caen, Rennes, Grenoble, Angers, Limoges, Tours, Orléans, Paris, Clermont, Bourgogne, Besançon, Rouen et Evreux (à gauche du graphique, caractérisés par une coordonnée fortement négative sur l'axe).

Le groupe auquel les villes Perpignan, Montpellier, Nice, Toulouse, Nîmes, Marseille, Biarritz, Bordeaux, Angoulême, Nantes et Ajaccio appartiennent (caractérisés par une coordonnée positive sur l'axe 1) partage :

- de fortes températures pour tous les mois de l'année (de la plus extrême à la moins extrême).

Le groupe auquel les villes Brest, Metz, Nancy, Lyon, Lille, Dijon, Strasbourg, Caen, Rennes, Grenoble, Angers, Limoges, Tours, Orléans, Paris, Clermont, Bourgogne, Besançon, Rouen et Evreux appartiennent (caractérisés par une coordonnée négative sur l'axe 1) partage :

- de faibles températures par rapport aux autres villes pour tous les mois (de la plus extrême à la moins extrême).

En conclusion l'axe 1 sépare les villes en deux catégories :

- les villes froides à gauche de l'axe 1 et celles chaudes à droite de l'axe 1.

La **dimension 2** oppose des individus tels que Brest, Caen, Rouen, Rennes, Limoges Poitiers, Tours, Lille, Angers, Toulon, Ajaccio, Nantes, Nice et Biarritz (en haut du graphique, caractérisés par une coordonnée positive sur l'axe) à des individus comme Lyon, Strasbourg, Grenoble, Dijon, Nîmes, Tours, Orléans, Paris, Clermont, Bourgogne, Besançon, Montpellier, Marseille, Perpignan, Toulouse, Bordeaux et Evreux (en bas du graphique, caractérisés par une coordonnée négative sur l'axe).

Parmi les villes froides :

Le groupe auquel les individus Brest, Lille, Caen et Rennes, Rouen, Angers, Poitiers, Limoges et Tours appartiennent (caractérisés par une coordonnée positive sur l'axe) partage :

- De faibles températures pour les mois de Mai, Juin, Juillet et Août et de fortes températures en Janvier et Décembre (de la plus extrême à la moins extrême). Ce sont des villes où il fait froid en été ;

le groupe auquel les individus Metz, Strasbourg, Evreux, Nancy, Reims, Orléans, Clermont, Grenoble, Lyon, Bourgogne et Dijon appartiennent (caractérisés par une coordonnée négative sur l'axe) partage :

- de faibles températures pour les températures en Janvier et Décembre et de fortes températures en Mai, Juin, Juillet et Août (de la plus extrême à la moins extrême). Ce sont des villes où il fait très froid en hiver.

Parmi les villes chaudes :

le groupe auquel les individus Biarritz, Nantes, Ajaccio, Toulon et Nice appartiennent (caractérisés par une coordonnée positive sur l'axe) partage :

- de faibles températures pour les mois de Mai, Juin, Juillet et Août et de fortes températures en Janvier et Décembre (de la plus extrême à la moins extrême). Ce sont des villes où il fait chaud en hiver.

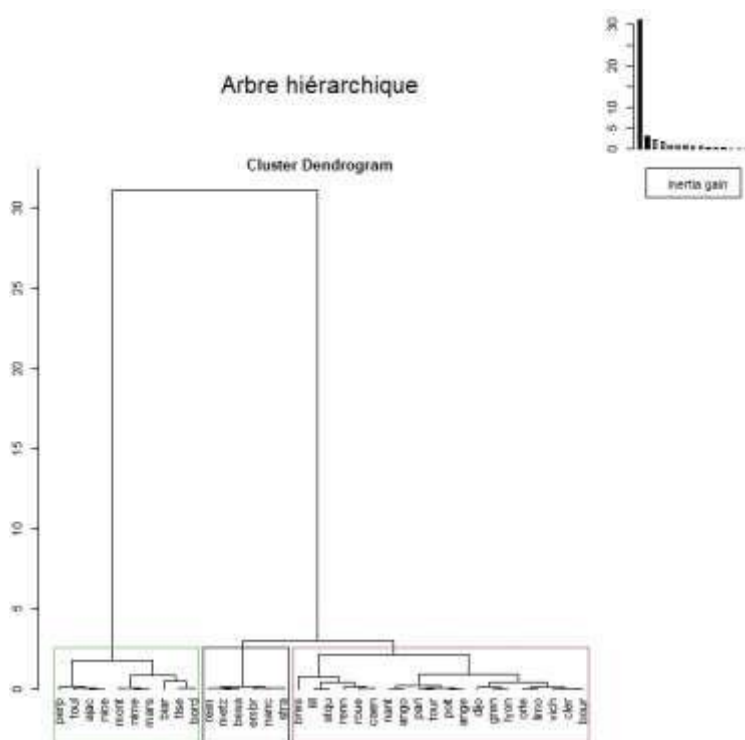


Le groupe auquel les individus Toulouse, Montpellier, Marseille, Nîmes, Perpignan et Bordeaux appartiennent (caractérisés par une coordonnée négative sur l'axe) partage :

- De faibles températures pour les températures en Janvier et Décembre et de fortes températures en Mai, Juin, Juillet et Août (de la plus extrême à la moins extrême). Ce sont des villes où il fait très chaud en été.

## VI. CLASSIFICATION HIERARCHIQUE ASCENDANTE

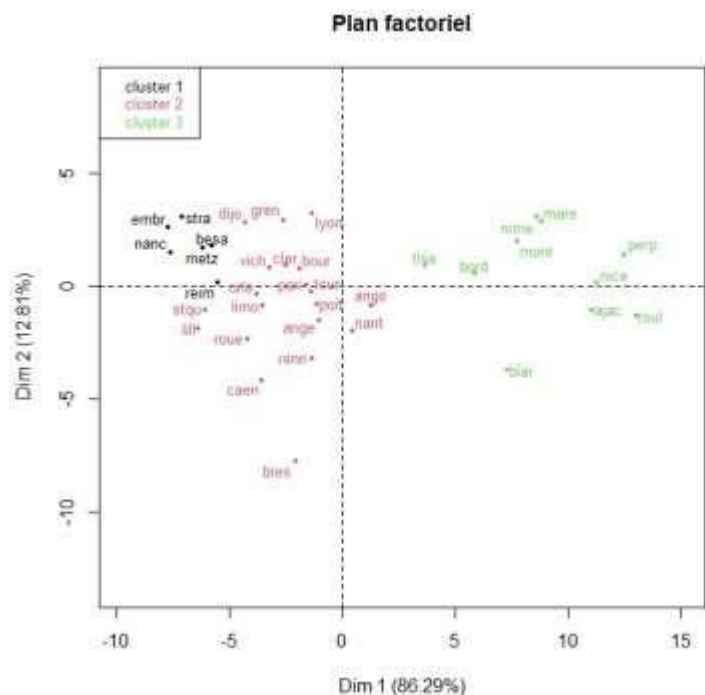
La classification a pour but de regrouper des individus qui ont des caractéristiques similaires (les distances entre individus de même groupe doivent être le plus petites possible et celles des groupes le plus éloignées possible).



**FIGURE 4:** ARBRE HIERARCHIQUE

La classification réalisée avec la distance euclidienne et le critère de Ward sur les individus fait apparaître 3 classes.





**FIGURE 5:** CLASSIFICATION ASCENDANTE HIERARCHIQUE DES INDIVIDUS (VILLES).

La **classe 1** est composée d'individus tels que Embreville, Metz, Nancy, Reims, Besançon et Strasbourg. Ce groupe est caractérisé par :

- de faibles températures pour les mois de décembre, février, janvier, mars, novembre, octobre, avril et septembre (de la plus extrême à la moins extrême).

La **classe 2** est composée d'individus tels que Brest, Caen, Dijon, Grenoble, Lille, Lyon et Rennes. Ce groupe est caractérisé par :

- de faibles températures pour les mois d'Août, Juillet, Mai, Juin, Septembre, Avril, Octobre, Novembre et Mars (de la plus extrême à la moins extrême).

La **classe 3** est composée d'individus tels que Ajaccio, Biart, Marseille, Montpellier, Nice, Nîmes, Perpignan et Toulon. Ce groupe est caractérisé par :

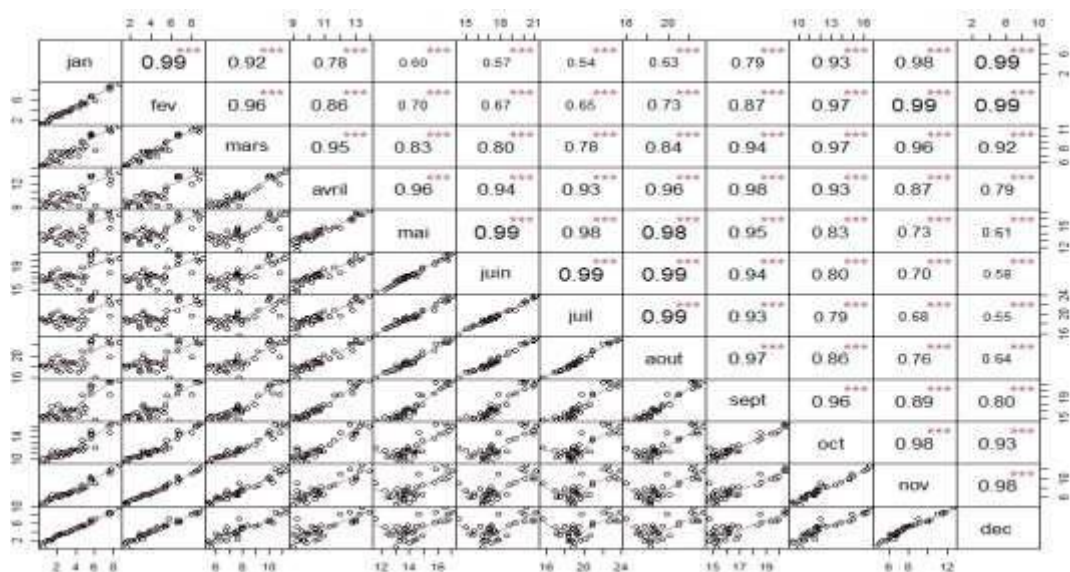
- de fortes valeurs pour des variables telles que Septembre, Octobre, Avril, Août, Mars, Novembre, Mai, Juin, Février et Juillet (de la plus extrême à la moins extrême).

Les classes une et deux regroupent les villes froides sur les douze mois de l'année alors que la classe 3 regroupe les villes chaudes sur les 12 mois de l'année excepté les mois de janvier et décembre.

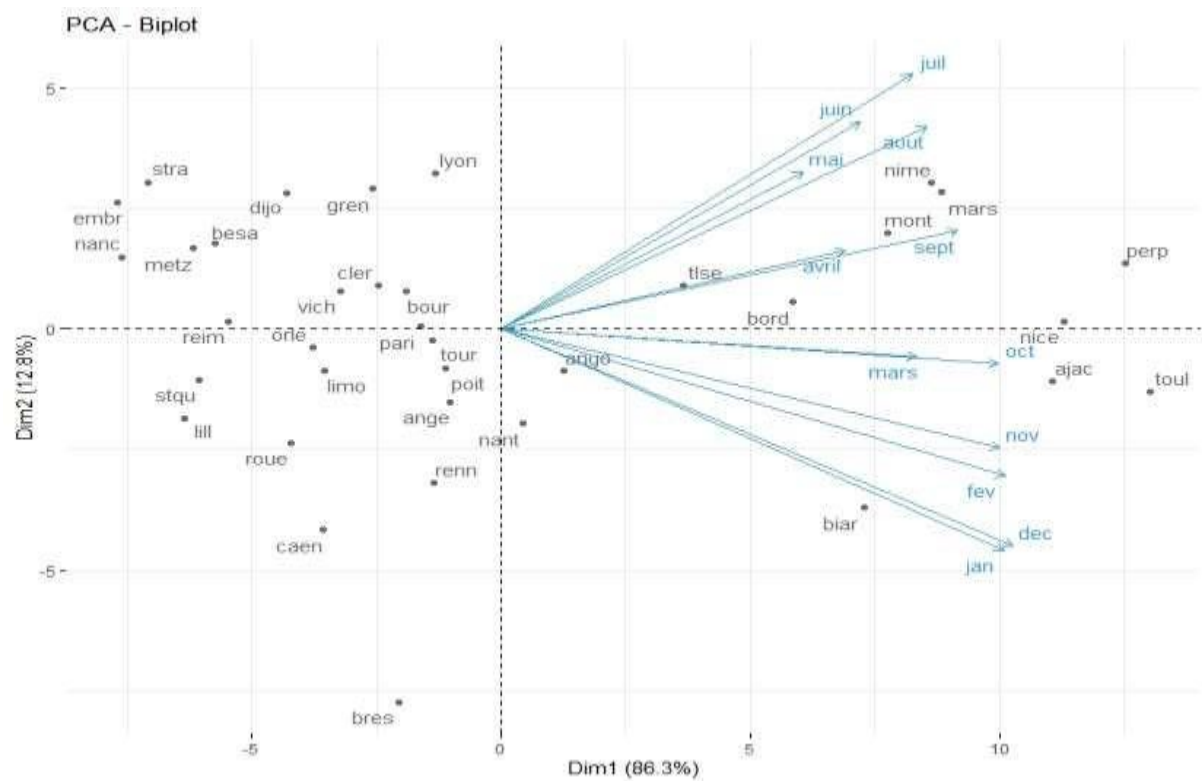
## VII. CLASSIFICATION PAR LA METHODE DU KMEANS



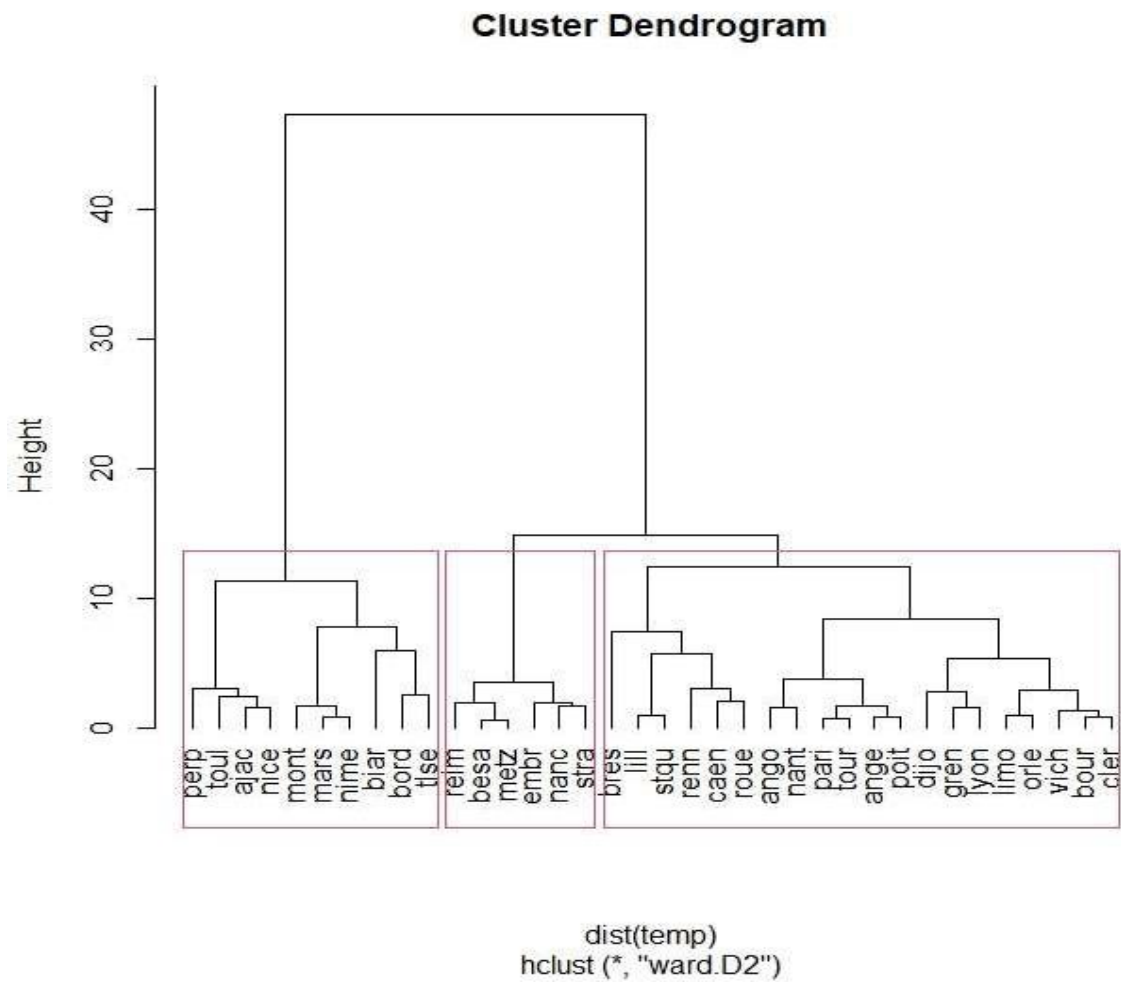
IX. ANNEXES



ANNEXE 1: CORRELATION



ANNEXE 2: BIPLLOT



### ***ANNEXE 3: DENDOGRAMME***