Comparaison d'individus sur de nombreux critères

La méthode Statistique ACP

Comparaison des villes sur leurs températures : CORRIGE

Corrigé du TP sur villes européennes

Le partage Web fourni par le Professeur :

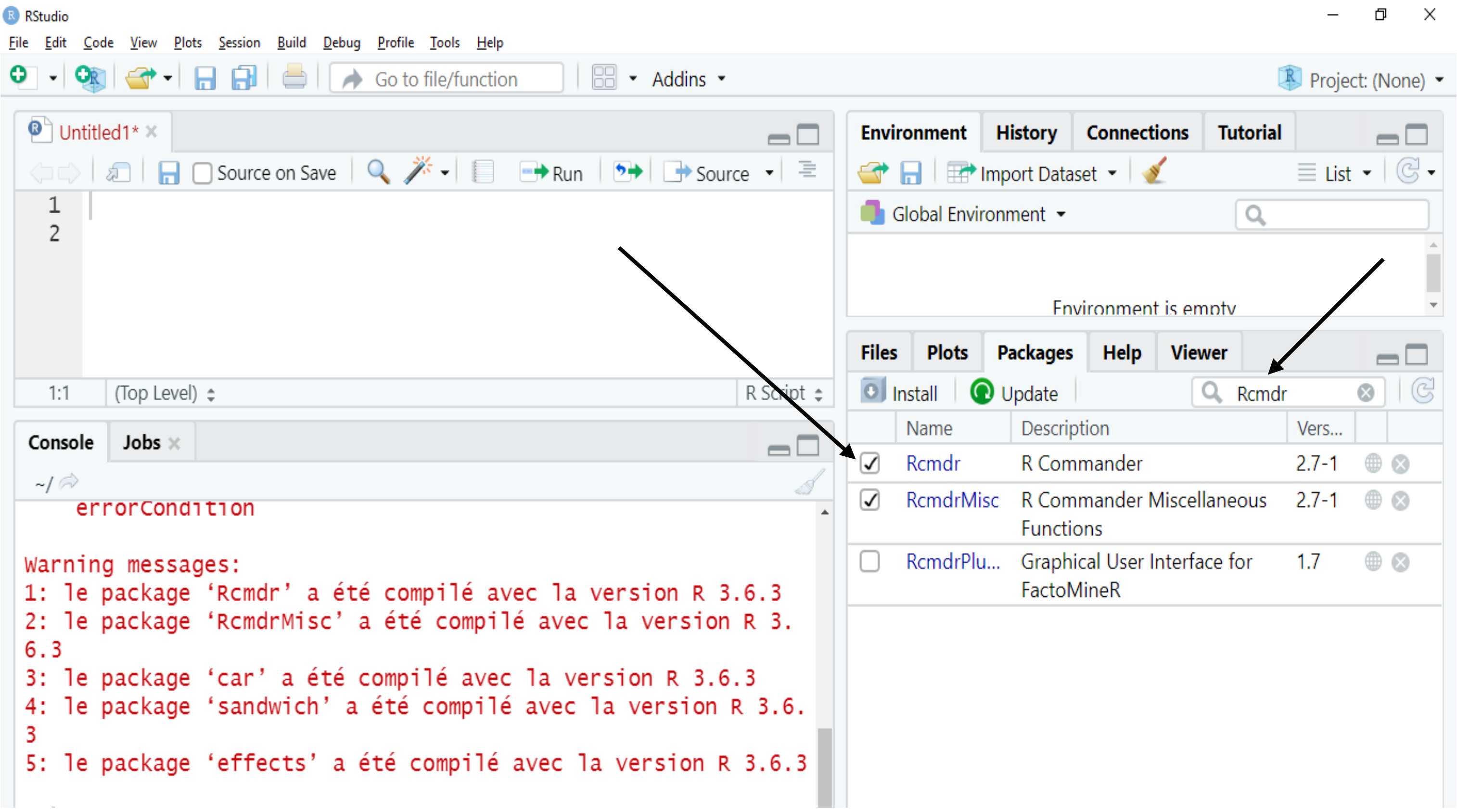
- ➪ Lire la théorie sur la méthode ACS dans le PDF « LogicielR_50_ACP_1_LaTheorie.pdf »

Savoirs à Acquérir :

- ➪ Savoir lancer le logiciel R puis afficher la fenêtre R Commander et charger le composant FactoMiner,
- ➪ Savoir décider si l'analyse demandée nécessite ou pas l'utilisation de la méthode ACP.
- ➪ Savoir manipuler le logiciel R jusqu'à afficher les résultats attendus par l'ACP,
- ➪ Savoir analyser ces résultats.

Corrigé : lancer le logiciel

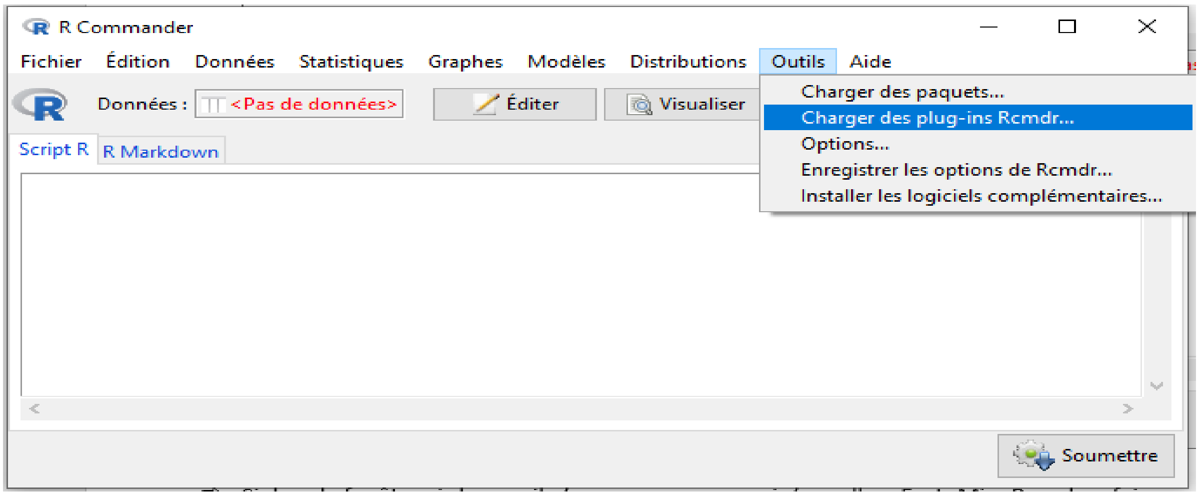
- ➪ **Lancer R Studio**, chercher **Rcmdr** et cocher à gauche de **Rcmdr** :



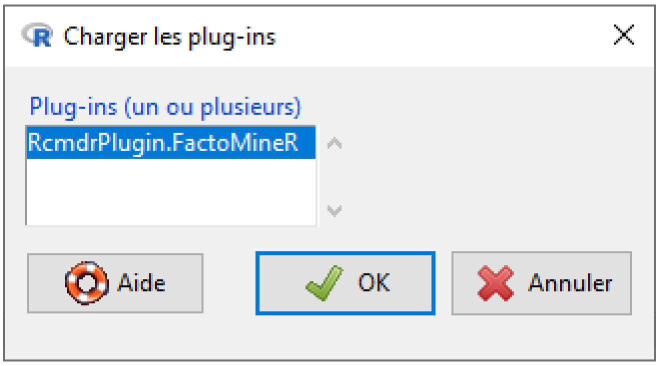
- ➪ Et attendre que ma fenêtre de R Commander s'affiche : C'est une fenêtre flottante et qui ne s'affiche pas

Forcément au-dessus de R Studio : il faut faire Alt/Tab sur un PC pour la chercher et l'afficher :

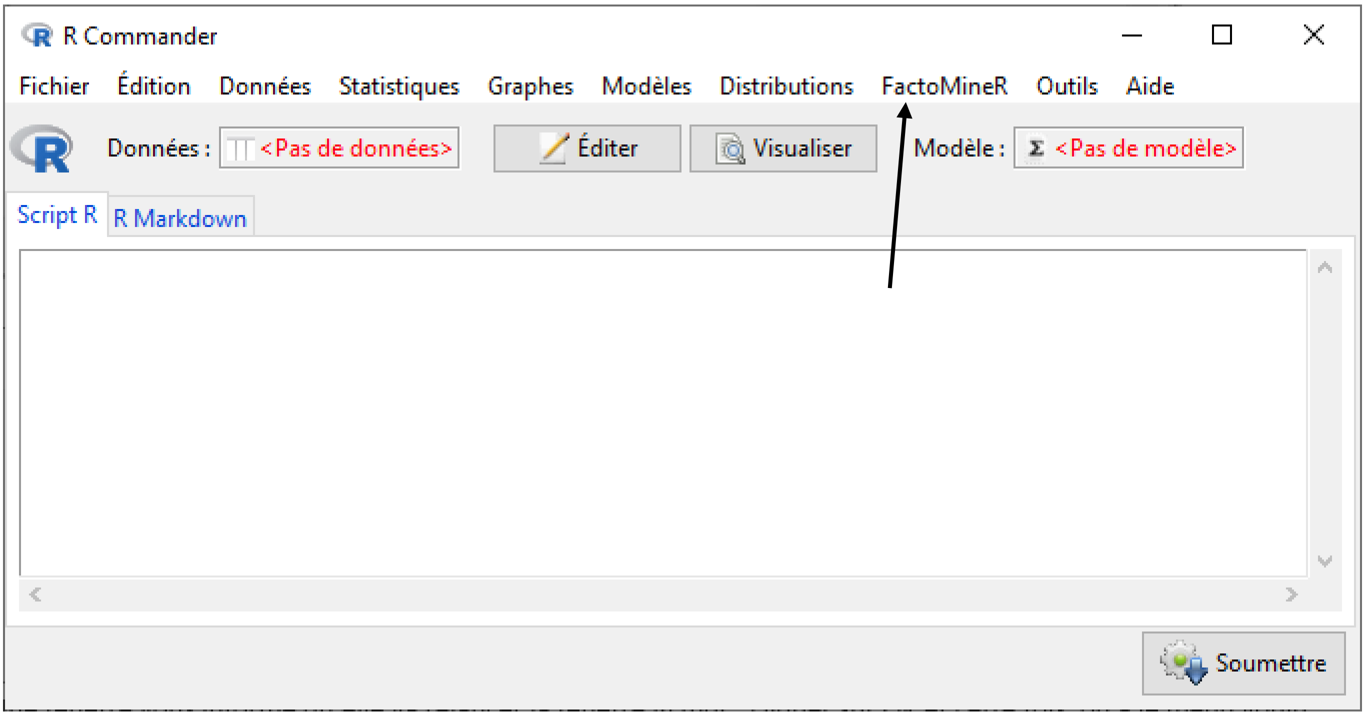
⇨ Si dans la fenêtre ci-dessus, il n'y a pas un menu qui s'appelle « **FactoMineR** », alors faire ceci :



Puis cliquer sur OK dans cette fenêtre :



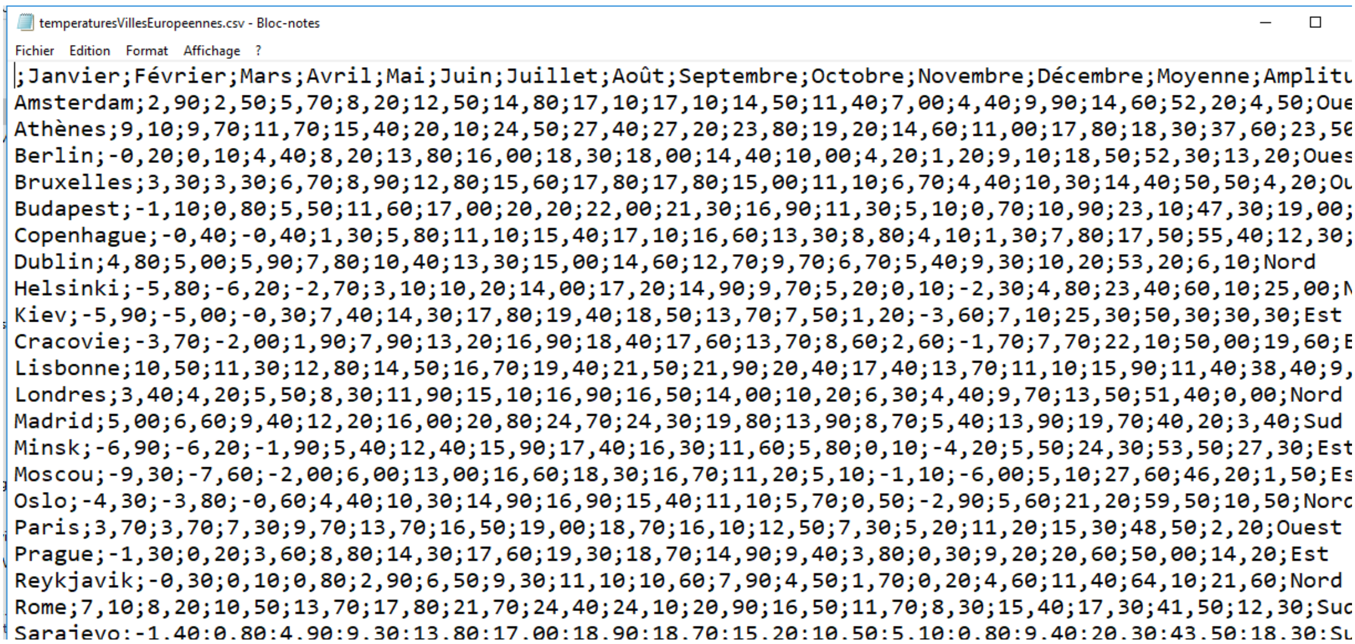
Une fenêtre vous informe qu'elle va relancer la fenêtre Rcmdr : cliquer sur OK et cette fois, on a le menu voulu :



Attention : Vous n'arrivez pas à accéder aux plugins de Rcmdr : cela veut qu'il manque un package nommé **« RcmdrPlugin.FactoMineR »** : il faut l'installer comme l'installation de Rcmdr er relancer R Studio

Corrigé : Examinez, en dehors de R, le fichier data et comprendre sa structure :

- ⇨ Le fichier fourni est « temperaturesVillesEuropeennes.csv »,
- ⇨ Ne pas double-cliquer dessus, car il sera ouvert par Excel et on n'aura pas toutes les infos dessus,
- ⇨ Ouvrez ce fichier par le « Bloc-notes » :


temperaturesVillesEuropeennes.csv - Bloc-notes

Fichier Edition Format Affichage ?

```
;Janvier;Février;Mars;Avril;Mai;Juin;Juillet;Août;Septembre;Octobre;Novembre;Décembre;Moyenne;Amplitu
Amsterdam;2,90;2,50;5,70;8,20;12,50;14,80;17,10;17,10;14,50;11,40;7,00;4,40;9,90;14,60;52,20;4,50;Oue
Athènes;9,10;9,70;11,70;15,40;20,10;24,50;27,40;27,20;23,80;19,20;14,60;11,00;17,80;18,30;37,60;23,5e
Berlin;-0,20;0,10;4,40;8,20;13,80;16,00;18,30;18,00;14,40;10,00;4,20;1,20;9,10;18,50;52,30;13,20;Oues
Bruxelles;3,30;3,30;6,70;8,90;12,80;15,60;17,80;17,80;15,00;11,10;6,70;4,40;10,30;14,40;50,50;4,20;Ou
Budapest;-1,10;0,80;5,50;11,60;17,00;20,20;22,00;21,30;16,90;11,30;5,10;0,70;10,90;23,10;47,30;19,00;
Copenhague;-0,40;-0,40;1,30;5,80;11,10;15,40;17,10;16,60;13,30;8,80;4,10;1,30;7,80;17,50;55,40;12,30;
Dublin;4,80;5,00;5,90;7,80;10,40;13,30;15,00;14,60;12,70;9,70;6,70;5,40;9,30;10,20;53,20;6,10;Nord
Helsinki;-5,80;-6,20;-2,70;3,10;10,20;14,00;17,20;14,90;9,70;5,20;0,10;-2,30;4,80;23,40;60,10;25,00;N
Kiev;-5,90;-5,00;-0,30;7,40;14,30;17,80;19,40;18,50;13,70;7,50;1,20;-3,60;7,10;25,30;50,30;30,30;Est
Cracovie;-3,70;-2,00;1,90;7,90;13,20;16,90;18,40;17,60;13,70;8,60;2,60;-1,70;7,70;22,10;50,00;19,60;E
Lisbonne;10,50;11,30;12,80;14,50;16,70;19,40;21,50;21,90;20,40;17,40;13,70;11,10;15,90;11,40;38,40;9,
Londres;3,40;4,20;5,50;8,30;11,90;15,10;16,90;16,50;14,00;10,20;6,30;4,40;9,70;13,50;51,40;0,00;Nord
Madrid;5,00;6,60;9,40;12,20;16,00;20,80;24,70;24,30;19,80;13,90;8,70;5,40;13,90;19,70;40,20;3,40;Sud
Minsk;-6,90;-6,20;-1,90;5,40;12,40;15,90;17,40;16,30;11,60;5,80;0,10;-4,20;5,50;24,30;53,50;27,30;Est
Moscou;-9,30;-7,60;-2,00;6,00;13,00;16,60;18,30;16,70;11,20;5,10;-1,10;-6,00;5,10;27,60;46,20;1,50;Es
Oslo;-4,30;-3,80;-0,60;4,40;10,30;14,90;16,90;15,40;11,10;5,70;0,50;-2,90;5,60;21,20;59,50;10,50;Norc
Paris;3,70;3,70;7,30;9,70;13,70;16,50;19,00;18,70;16,10;12,50;7,30;5,20;11,20;15,30;48,50;2,20;Ouest
Prague;-1,30;0,20;3,60;8,80;14,30;17,60;19,30;18,70;14,90;9,40;3,80;0,30;9,20;20,60;50,00;14,20;Est
Reykjavik;-0,30;0,10;0,80;2,90;6,50;9,30;11,10;10,60;7,90;4,50;1,70;0,20;4,60;11,40;64,10;21,60;Nord
Rome;7,10;8,20;10,50;13,70;17,80;21,70;24,40;24,10;20,90;16,50;11,70;8,30;15,40;17,30;41,50;12,30;Suc
Sarajevo;-1,40;0,80;4,90;9,30;13,80;17,00;18,90;18,70;15,20;10,50;5,10;0,80;9,40;20,30;43,50;18,30;Su
```

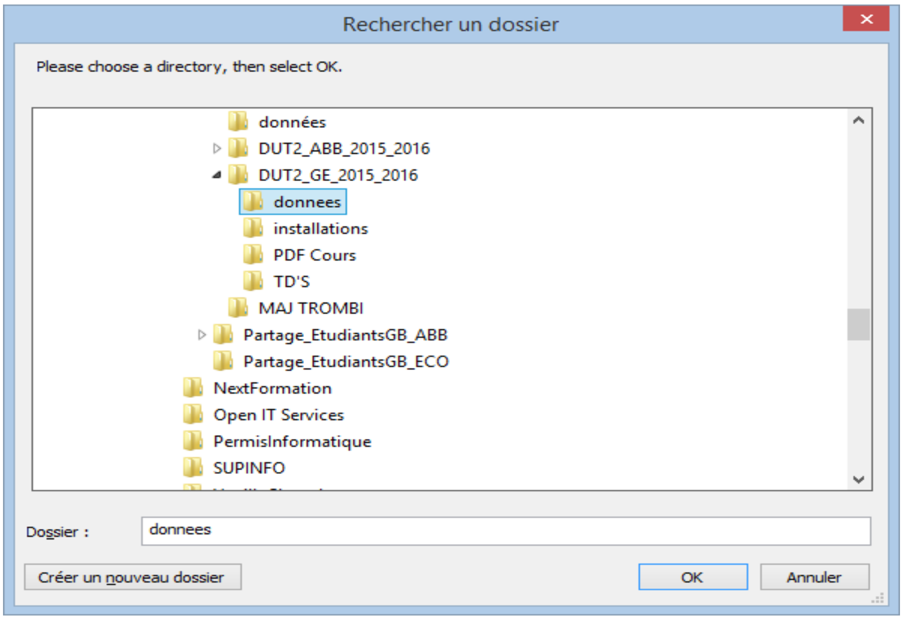
⇨ Constatations :

- La 1ère ligne contient les noms des variables et pas les données : « Janvier », « Février », ….
- Les colonnes sont séparées par le symbole « ; »
- Sur la 1ère ligne, la 1ère colonne ne contient pas de valeur,
- Dans les autres lignes (sauf la 1ère), la 1ère colonne contient des noms de villes (Amsterdam, Athènes, …)

Corrigé : Charger et visualiser les données dans R Commander :

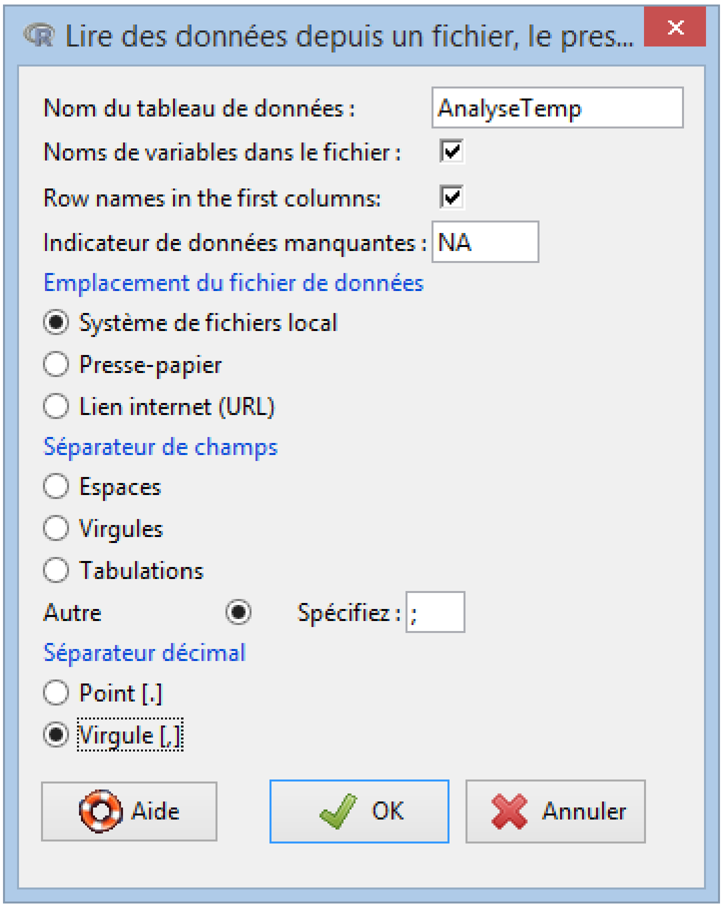
Se positionner sur le dossier contenant le fichier de données à analyser

Menu « Fichier » → option « changer le répertoire de travail »

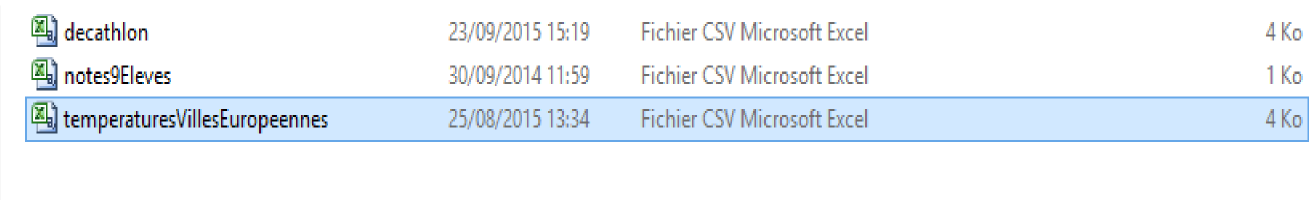


Charger le fichier de données

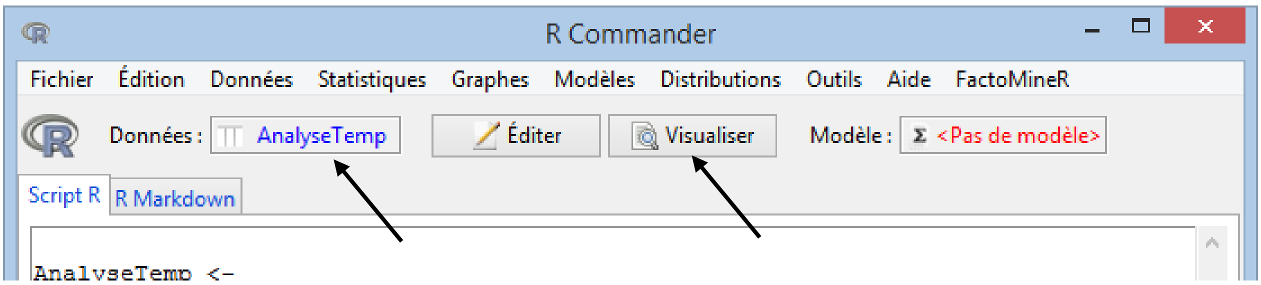
Menu « FactoMineR »: option Lire les données depuis …. » :



et montrer le fichier sur le disque :



Si le fichier de données a été bien chargé, son nom apparaît en haut. Données : « AnalyseTemp » :



Ce nom « AnalyseTemp » a été choisi aléatoirement pour nommer la matrice de données dans le logiciel.

Cette matrice contient les données issues du fichier sur le disque,

Visualiser le fichier de données : cliquer sur le bouton « visualiser »

AnalyseTemp

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Moyenne	Amplitude	Latitude	Longitude	Région
Amsterdam	2.9	2.5	5.7	8.2	12.5	14.8	17.1	17.1	14.5	11.4	7.0	4.4	9.9	14.6	52.2	4.5	Ouest
Athènes	9.1	9.7	11.7	15.4	20.1	24.5	27.4	27.2	23.8	19.2	14.6	11.0	17.8	18.3	37.6	23.5	Sud
Berlin	-0.2	0.1	4.4	8.2	13.8	16.0	18.3	18.0	14.4	10.0	4.2	1.2	9.1	18.5	52.3	13.2	Ouest
Bruxelles	3.3	3.3	6.7	8.9	12.8	15.6	17.8	17.8	15.0	11.1	6.7	4.4	10.3	14.4	50.5	4.2	Ouest
Budapest	-1.1	0.8	5.5	11.6	17.0	20.2	22.0	21.3	16.9	11.3	5.1	0.7	10.9	23.1	47.3	19.0	Est
Copenhague	-0.4	-0.4	1.3	5.8	11.1	15.4	17.1	16.6	13.3	8.8	4.1	1.3	7.8	17.5	55.4	12.3	Nord
Dublin	4.8	5.0	5.9	7.8	10.4	13.3	15.0	14.6	12.7	9.7	6.7	5.4	9.3	10.2	53.2	6.1	Nord
Helsinki	-5.8	-6.2	-2.7	3.1	10.2	14.0	17.2	14.9	9.7	5.2	0.1	-2.3	4.8	23.4	60.1	25.0	Nord
Kiev	-5.9	-5.0	-0.3	7.4	14.3	17.8	19.4	18.5	13.7	7.5	1.2	-3.6	7.1	25.3	50.3	30.3	Est
Cracovie	-3.7	-2.0	1.9	7.9	13.2	16.9	18.4	17.6	13.7	8.6	2.6	-1.7	7.7	22.1	50.0	19.6	Est
Lisbonne	10.5	11.3	12.8	14.5	16.7	19.4	21.5	21.9	20.4	17.4	13.7	11.1	15.9	11.4	38.4	9.1	Sud
Londres	3.4	4.2	5.5	8.3	11.9	15.1	16.9	16.5	14.0	10.2	6.3	4.4	9.7	13.5	51.4	0.0	Nord
Madrid	5.0	6.6	9.4	12.2	16.0	20.8	24.7	24.3	19.8	13.9	8.7	5.4	13.9	19.7	40.2	3.4	Sud
Minsk	-6.9	-6.2	-1.9	5.4	12.4	15.9	17.4	16.3	11.6	5.8	0.1	-4.2	5.5	24.3	53.5	27.3	Est
Moscou	-9.3	-7.6	-2.0	6.0	13.0	16.6	18.3	16.7	11.2	5.1	-1.1	-6.0	5.1	27.6	46.2	1.5	Est
Oslo	-4.3	-3.8	-0.6	4.4	10.3	14.9	16.9	15.4	11.1	5.7	0.5	-2.9	5.6	21.2	59.5	10.5	Nord
Paris	3.7	3.7	7.3	9.7	13.7	16.5	19.0	18.7	16.1	12.5	7.3	5.2	11.2	15.3	48.5	2.2	Ouest
Prague	-1.3	0.2	3.6	8.8	14.3	17.6	19.3	18.7	14.9	9.4	3.8	0.3	9.2	20.6	50.0	14.2	Est
Reykjavik	-0.3	0.1	0.8	2.9	6.5	9.3	11.1	10.6	7.9	4.5	1.7	0.2	4.6	11.4	64.1	21.6	Nord
Rome	7.1	8.2	10.5	13.7	17.8	21.7	24.4	24.1	20.9	16.5	11.7	8.3	15.4	17.3	41.5	12.3	Sud
Sarajevo	-1.4	0.8	4.9	9.3	13.8	17.0	18.9	18.7	15.2	10.5	5.1	0.8	9.4	20.3	43.5	18.3	Sud
Sofia	-1.7	0.2	4.3	9.7	14.3	17.7	20.0	19.5	15.8	10.7	5.0	0.6	9.6	21.7	42.4	23.2	Est
Stockholm	-3.5	-3.5	-1.3	3.5	9.2	14.6	17.2	16.0	11.7	6.5	1.7	-1.6	5.8	20.7	59.2	18.0	Nord
Anvers	3.1	2.9	6.2	8.9	12.9	15.5	17.9	17.6	14.7	11.5	6.8	4.7	10.3	15.0	51.1	4.2	Ouest
Barcelone	9.1	10.3	11.8	14.1	17.4	21.2	24.2	24.1	21.7	17.5	13.1	10.0	16.2	15.1	41.2	2.2	Sud
Bordeaux	5.6	6.7	9.0	11.9	15.0	18.3	20.4	20.0	17.6	13.5	8.5	6.1	12.7	14.8	44.5	0.3	Ouest
Edimbourg	2.9	3.6	4.7	7.1	9.9	13.0	14.7	14.3	12.1	8.7	5.3	3.7	8.3	11.8	55.0	3.0	Nord
Francfort	0.2	1.8	5.4	9.7	14.3	17.5	19.0	18.3	14.8	9.8	4.9	1.7	9.8	18.8	50.1	8.4	Ouest
Genève	0.1	1.9	5.1	9.4	13.8	17.3	19.4	18.5	15.0	9.8	4.9	1.4	9.7	19.3	46.1	6.1	Ouest
Gènes	8.7	8.7	11.4	13.8	17.5	21.0	24.5	24.6	21.8	17.8	12.2	10.0	16.1	15.9	44.3	9.4	Sud

Vérifier que les données sont bien séparées dans les colonnes, quelles sont cohérentes.

Termes statistiques

➔ Dans notre exemple des températures des villes :

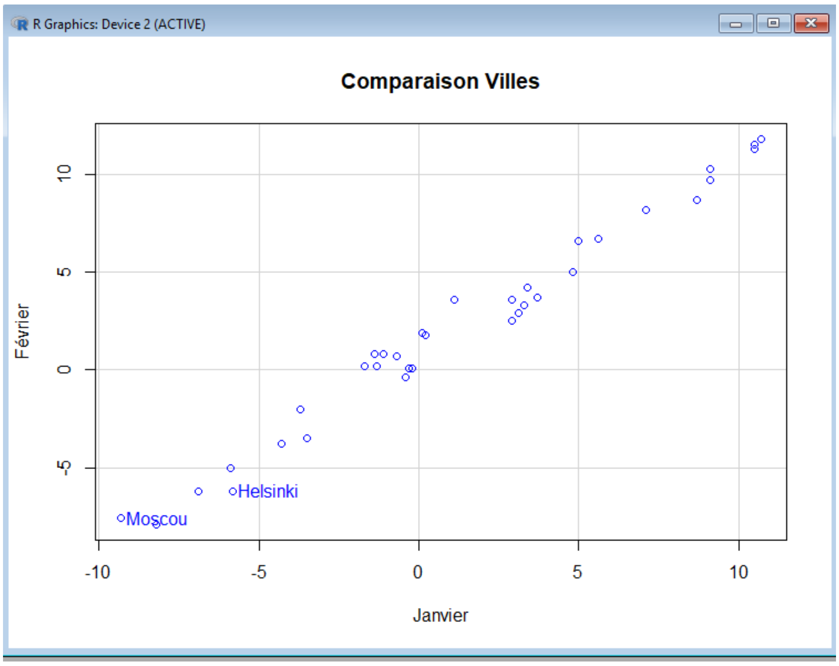
- Les **lignes** sont des **individus** = ici des villes,
- Les **colonnes** sont des variables (les caractères) :
  - Variables numériques qui vont servir à l'analyse (**variables actives**) :
    - L'analyse multidimensionnelle se base essentiellement sur ces variables,
    - Ici : les douze mois de l'année,
    - Le contenu d'une cellule de ces colonnes :
      - une température moyenne du mois, calculée sur 30 ans,
  - **Variables qualitatives supplémentaires** (ou facteurs illustratifs) :
    - Aide à mieux affiner l'analyse,
    - Ici : région ou se trouve la ville,
  - **Variables numériques supplémentaires** (ou variables illustratives) :
    - Aide à mieux affiner l'analyse,
    - Ici : latitude et longitude de la ville.

Corrigé : Peut-t-on utiliser des Statistiques descriptives pour répondre aux demandes posées ?

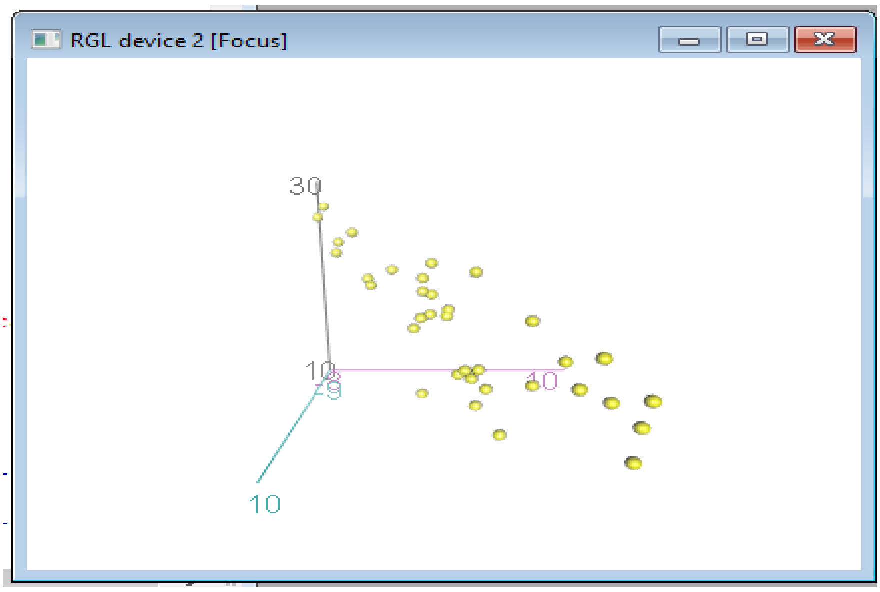
Ici il s'agit de comparer plusieurs villes sur 12 critères (variables) : les 12 mois de températures,

⇨ On peut toujours calculer des valeurs statistiques comme : moyennes, médianes, écart-type, ….
Sur les 12 variables et comparer,

⇨ On peut essayer de comparer par un nuage de points mais :
Exemple : comparer les villes sur Janvier et Février :



Exemple : comparer les villes sur Janvier, Février et Mars :



Mais on ne sait pas afficher un nuage de points de 4,5, ….. dimensions (axes)

Corrigé : Doit-on utiliser la méthode ACP, pour répondre aux demandes Sur ce fichier ?

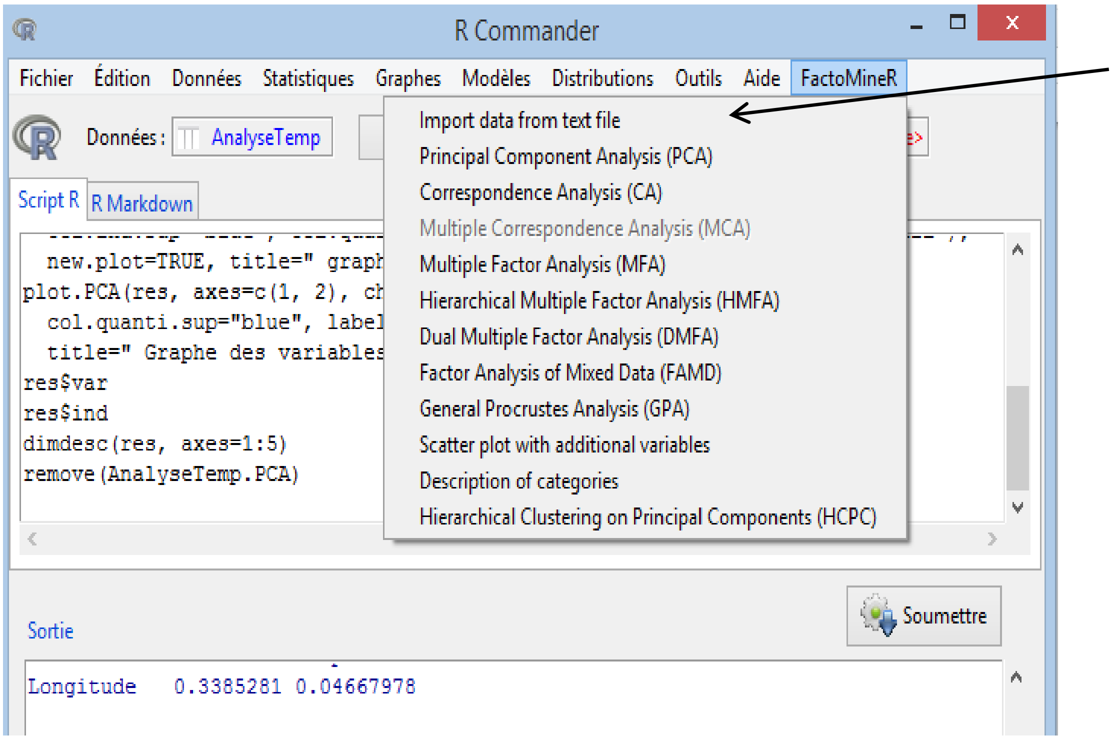
- ⇨ Est-ce que dans le sujet : il s'agit de comparer des individus (villes, voitures, médicaments, bactéries, étudiants, … entre eux ?
- ⇨ 2) Est ce que chaque ligne (à part la 1ere) représente un individu (1ere colonne qui porte le nom de la ville Ou de la voiture ou de ….) ?

- ⇨ Est-ce que les lignes sont décrites, par au moins 4 colonnes (variables) quantitatives ?
- ⇨ Si OUI à la question précédente : ces colonnes représentent elles toutes la même nature de données (Température, score, note, puissance, ...) ?
- ⇨ Si Oui aux deux questions précédentes : Est-ce que les nombres dans chacune de ces colonnes quantitatives représentent une valeur et pas un comptage (une température et pas un nombre de températures, une note et pas un nombre de notes, ...)?

Si on répond oui aux questions ci-dessus : Il faut utiliser la méthode ACP

Réalisation de l'ACP (Analyse de Composantes Principales)

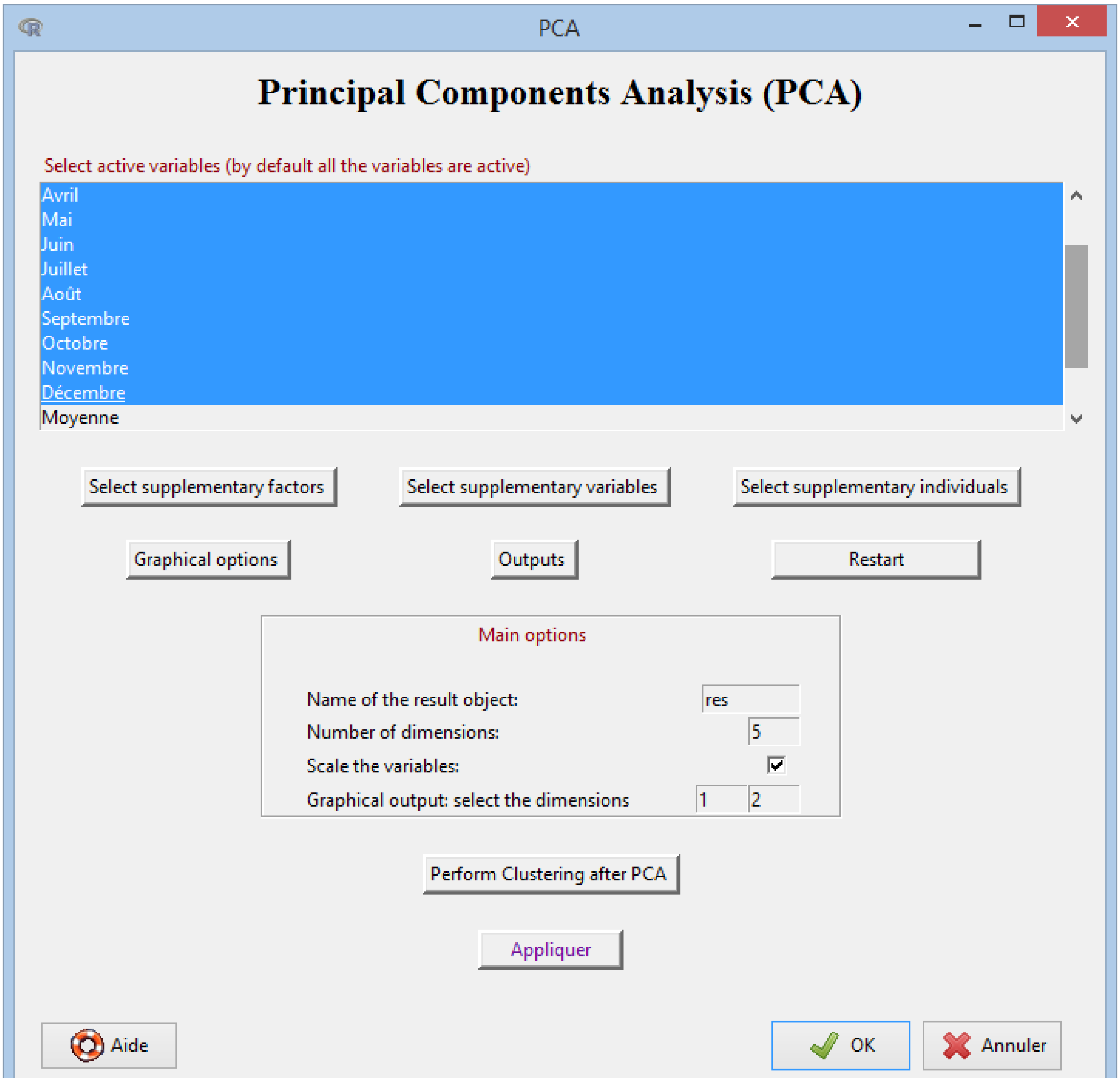
Lancer l'ACP



import data text from file = Lire les données depuis un ficher, ...

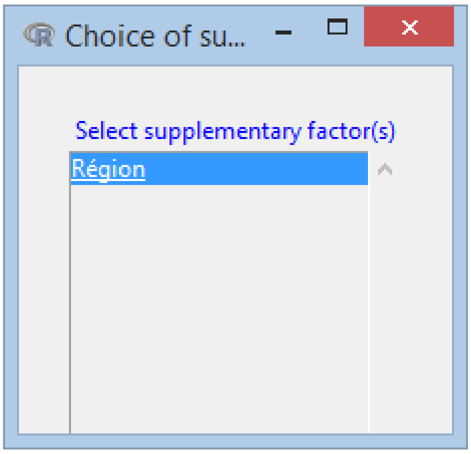
Choisir les variables actives

➔ **Fenêtre « ACP (PCA en Anglais) » :** Dans la liste en haut, choisir les variables actives (ici les douze mois de l'année) :



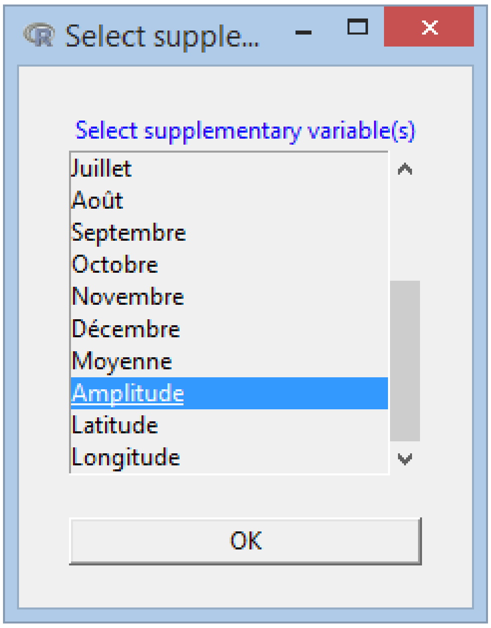
<u>Choisir les variables qualitatives supplémentaires (les facteurs) si nécessaire :</u>

- ➔ **Toujours la Fenêtre « ACP (ou PCA) » :** Cliquer sur le bouton **«Choisir facteurs illustratifs (ou Select supplementary factors)** » et choisir dans la petite fenêtre un ou plusieurs facteurs,
- ➔ Ici choisir « Région » :



Choisir les variables numériques supplémentaires si nécessaire :

- ➔ **Toujours dans la Fenêtre « ACP (ou PCA) » :** Cliquer sur le bouton **« Select supplementary variables** » et choisir dans la petite fenêtre une ou plusieurs variables,
- ➔ Ici choisir « Amplitude » :



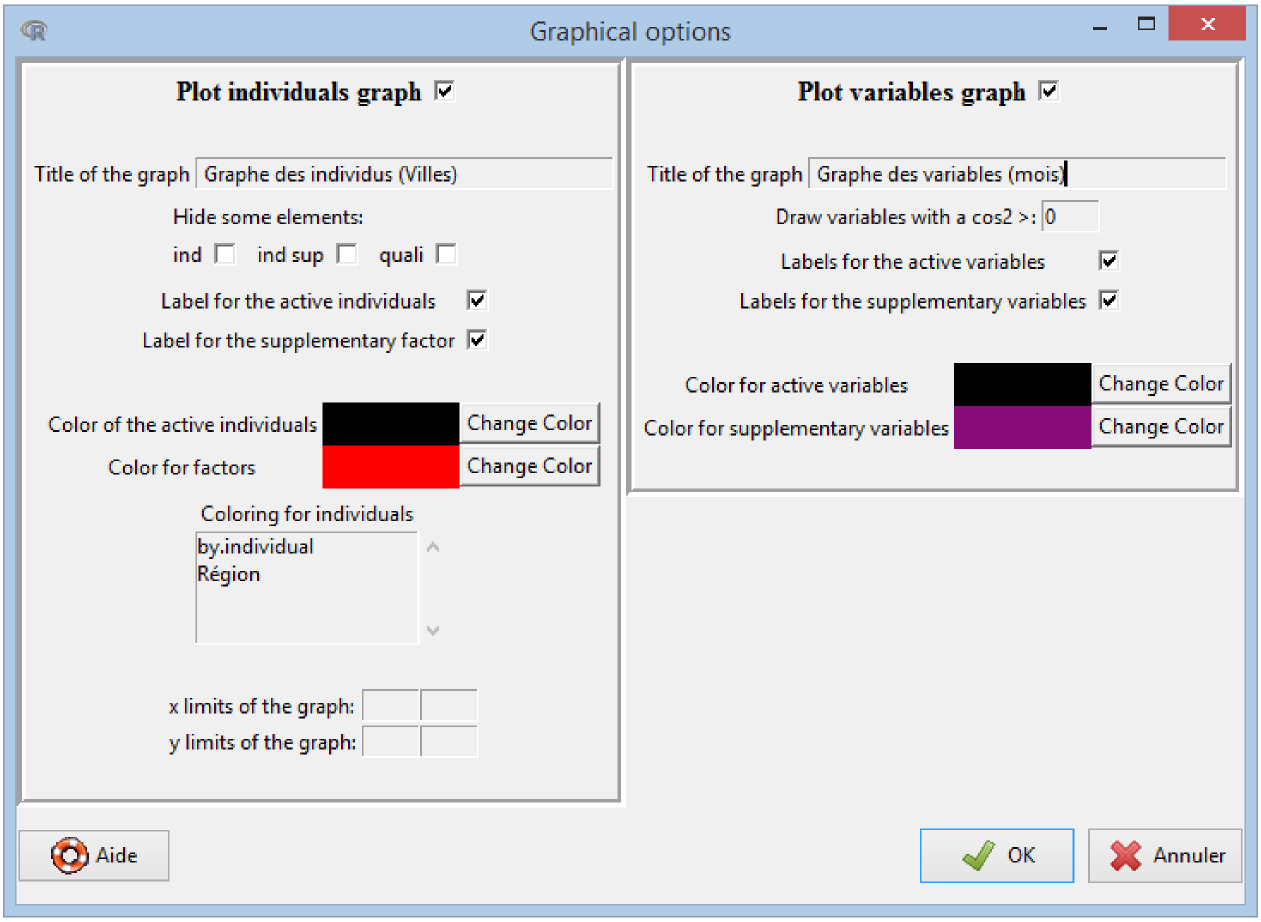
Donner des titres parlants aux graphes d'interprétation

➔ Toujours dans **Fenêtre « ACP (ou PCA) » :** Deux graphes principaux vont bientôt s'afficher et permettront d'analyser les données,

- Le graphe des individus et,
- Le graphe des variables

➔ Avant de les afficher, il faut changer leurs propriétés : donner un titre parlant, changer les couleurs des individus, Des variables, ….

➔ Cliquer sur le bouton « **Graphical Options** » et changer ainsi les titres :

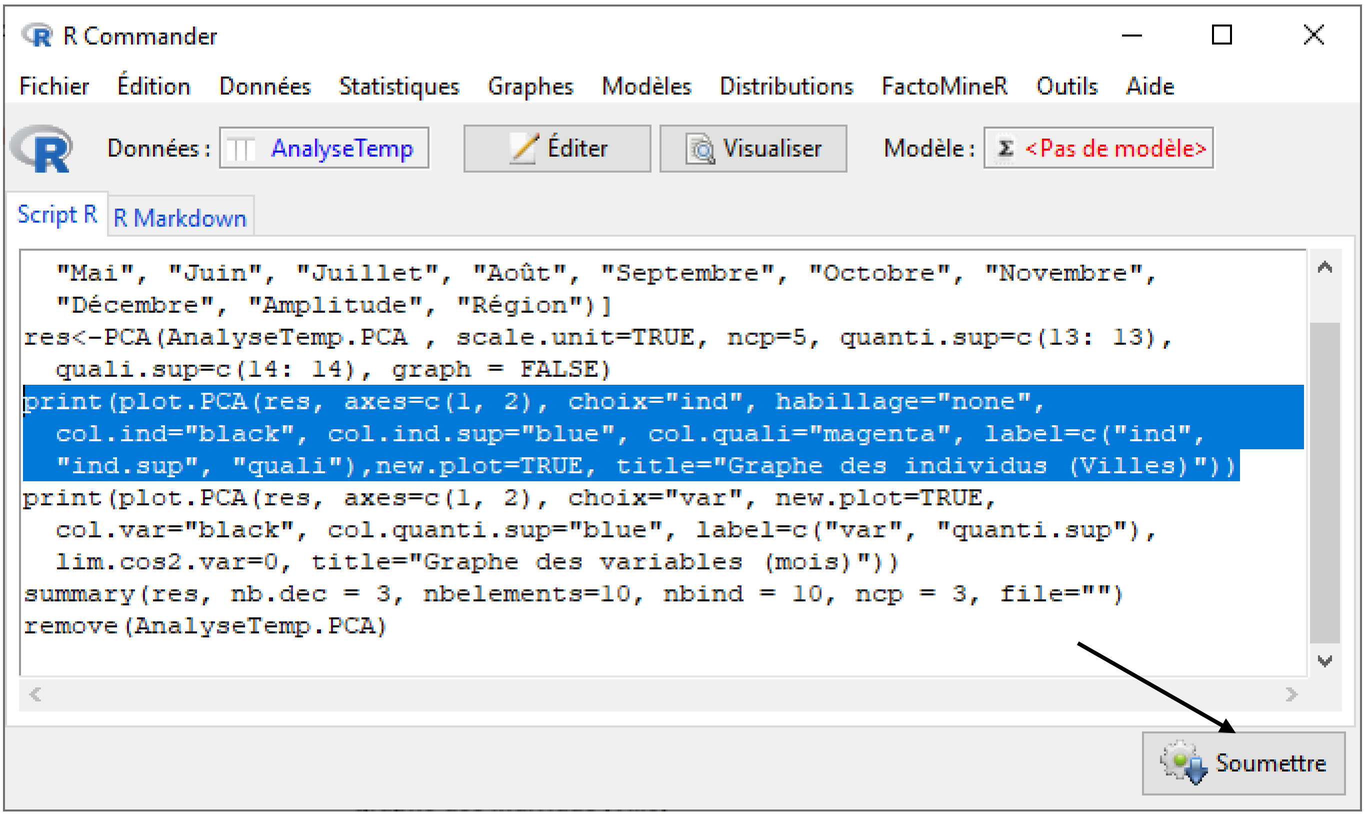


Produire les résulatats sur l'ACP

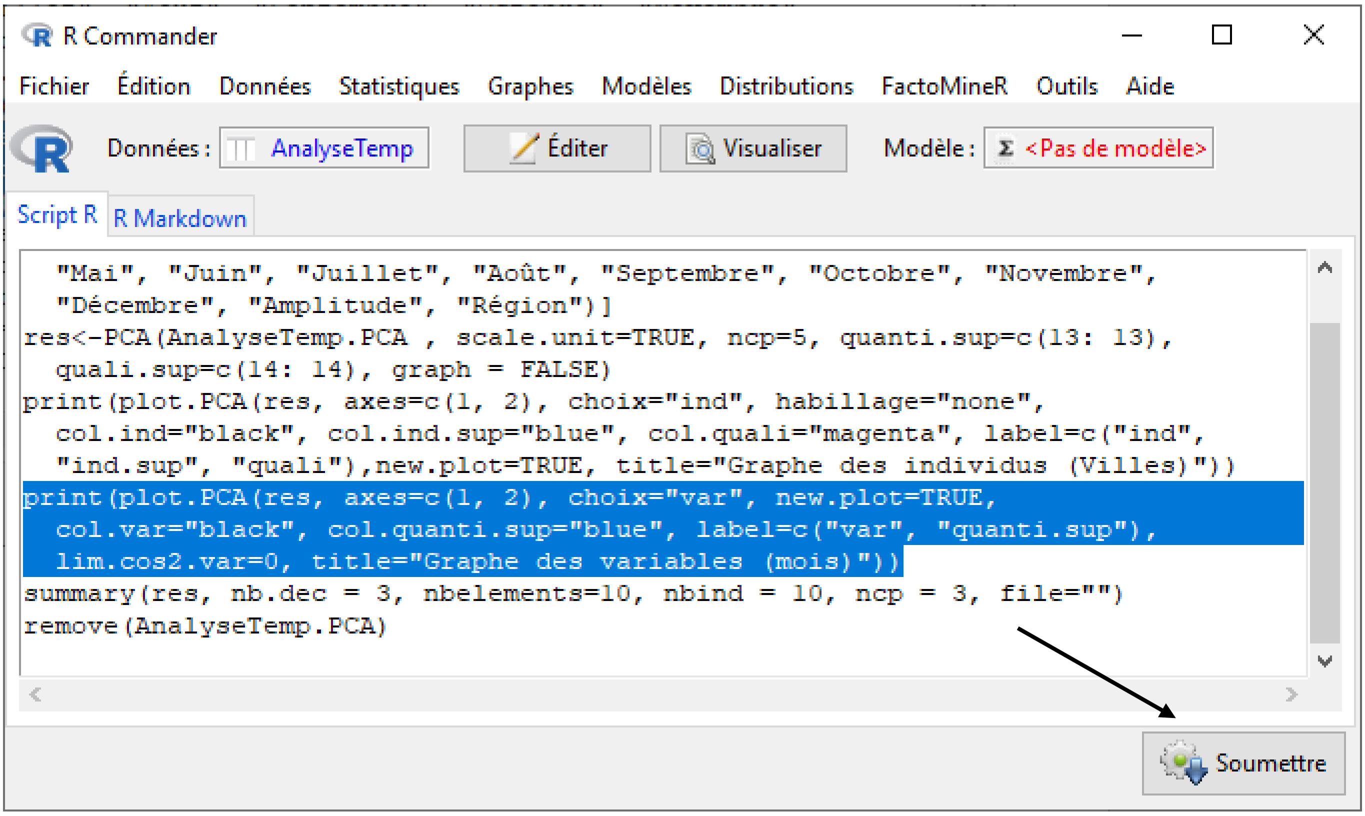
- ➔ Actuellement, il y a **Fenêtre «ACP (ou PCA) » d'ouverte** et la fenêtre **« R Commander »,**
- ➔ On peut passer de l'une à l'autre sans les fermer,
- ➔ Retourner dans R Commander et effacer le contenu des deux volets « **Script R** » et « **Sortie** »,
- ➔ Retourner dans la **Fenêtre « ACP (ou PCA) » :** Cliquer sur le bout « Appliquer » :
    - La fenêtre de R Commander contient maintenant dans le **volet « Sortie », plein d'infos écrites : Nous reviendrons dessus**

ATTENTION :

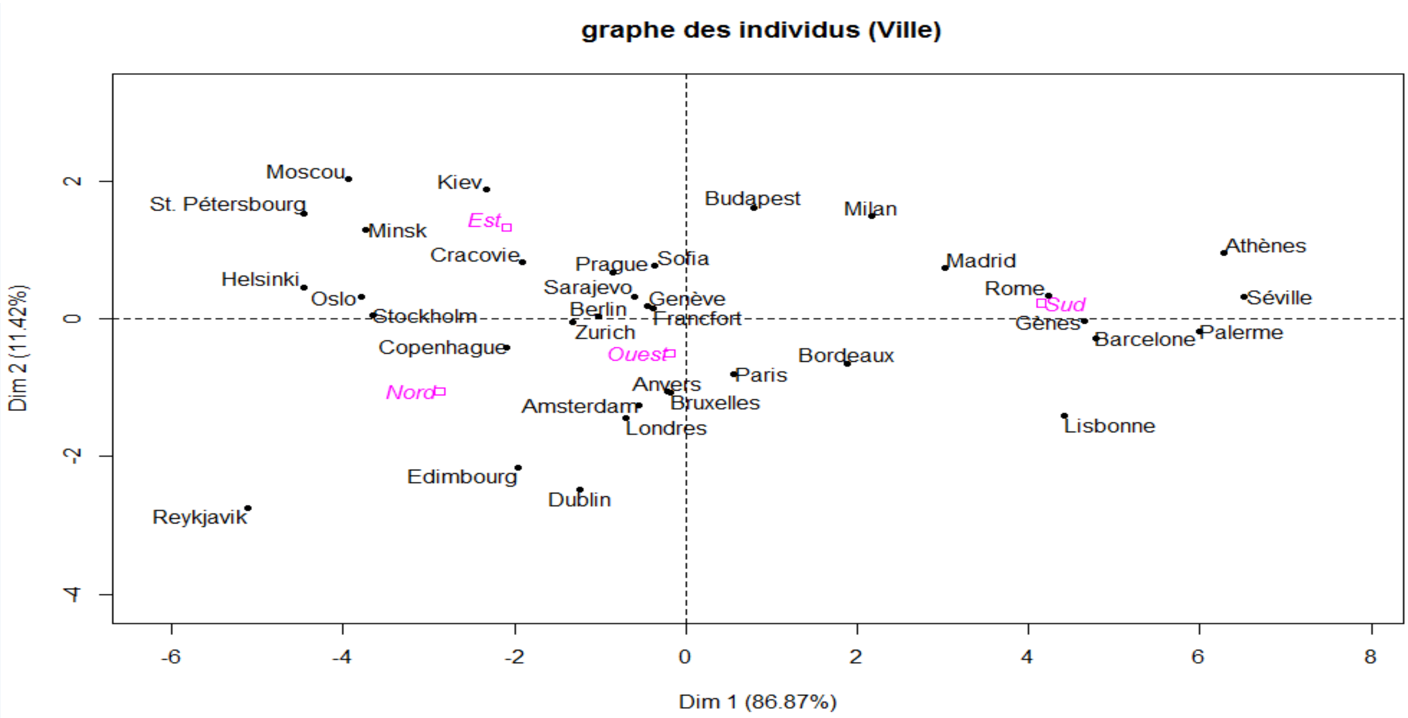
- **Si on utilise le logiciel R de Base : deux graphes s'affichent en même temps,**
- **Si on utilise R Commander via R Studio : Les 2 graphes sont superposés :**
  Pour afficher l'un ou l'autre, dans la fenêtre « R Graphics Device », repérez dans le script de R Commander les 2 commandes « print » et exécuter l'une ou l'autre pour afficher le graphe souhaité

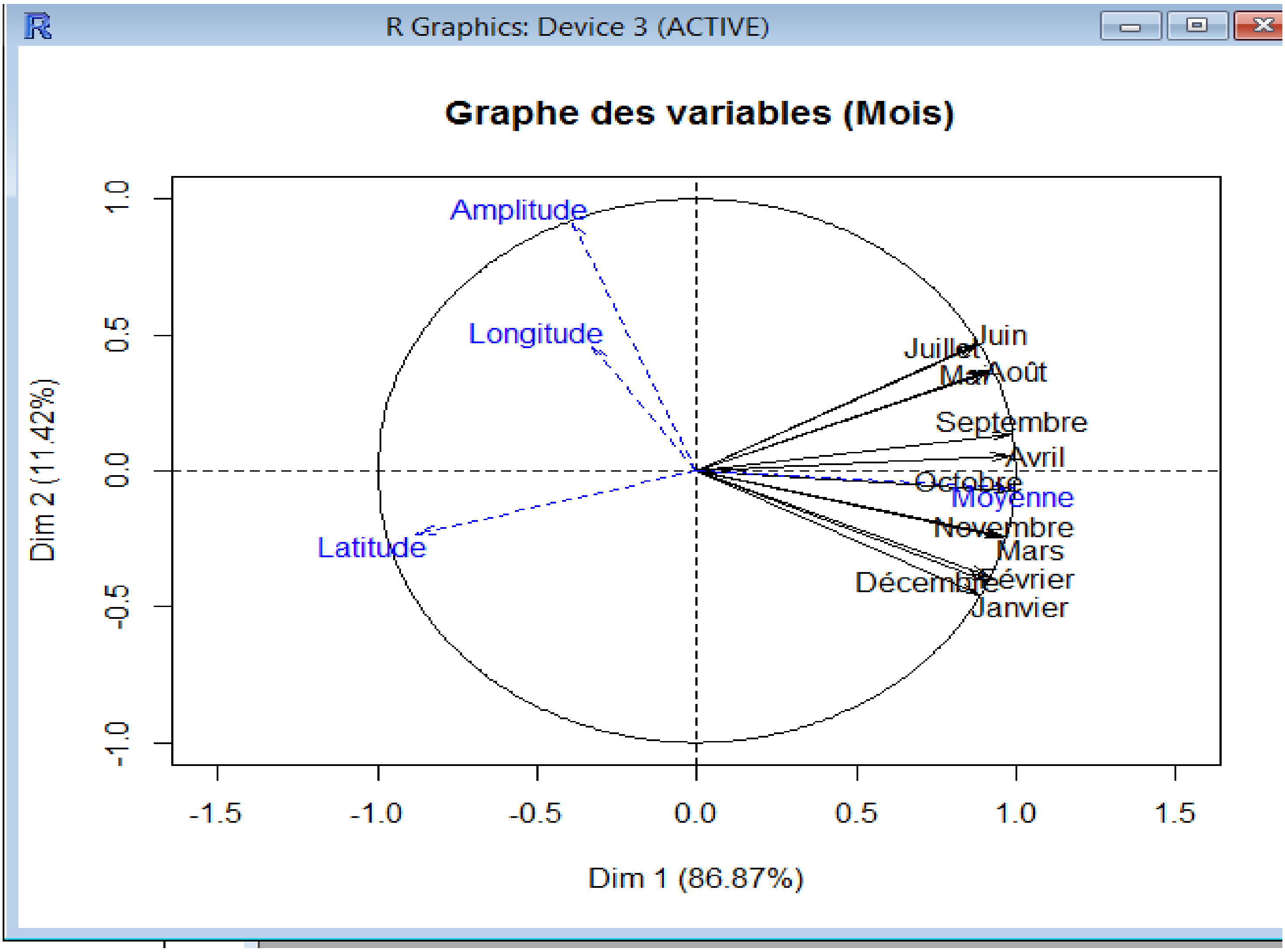


Ou



➔ Deux graphes :



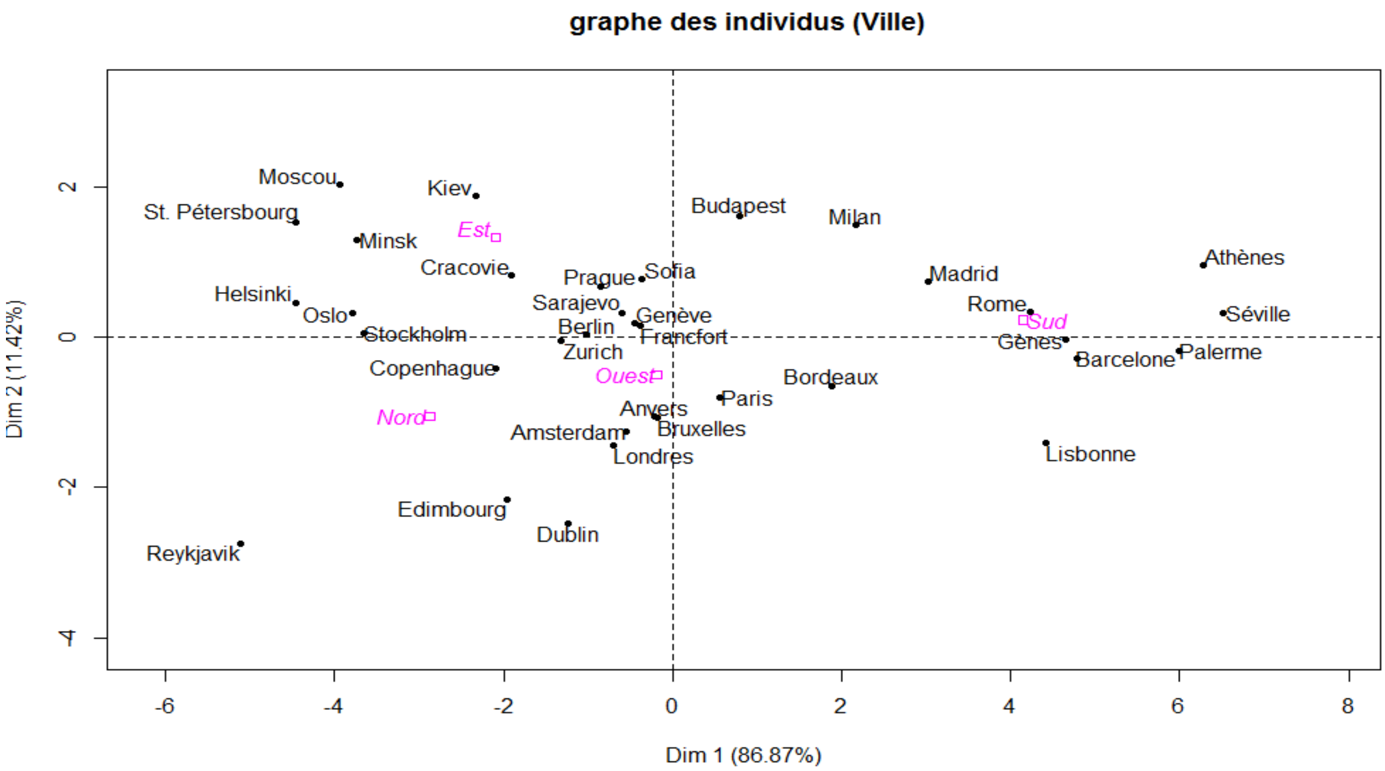


S'arranger pour les afficher côte à côte.

Analyse des résultats

L'analyse des données peut se faire de plus façons :

Analyser les individus



,

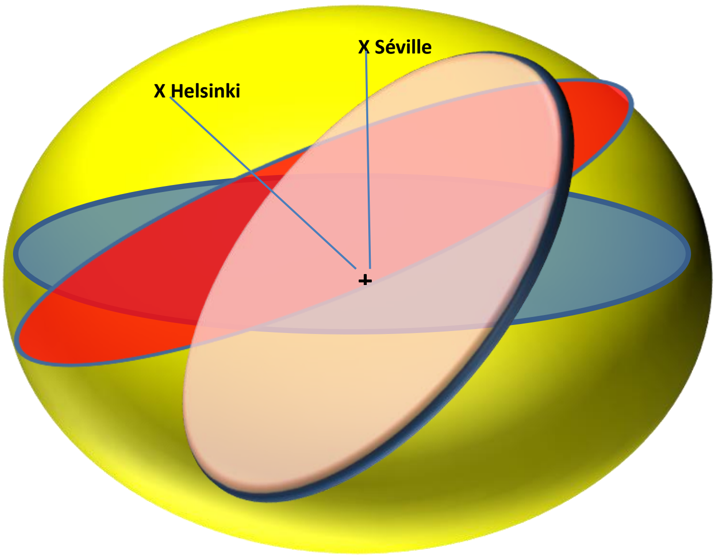
D’où vient le graphique des individus ?

➔ Les deux axes :

- L’axe horizontal est nommé simplement « **Dim 1** »,
- L’axe vertical est nommé simplement **« Dim 2** »,
- Si vous examinez la fenêtre « ACP (ou PCA) » vue ci-dessus, on parle de « **Number of dimensions** » = 5,
- **Que représentent ces deux dimensions** ?
    - D’abord on aime mieux analyser des données graphiquement,
    - dans notre étude, un individu est une ville,
    - chaque individu (chaque ville) est caractérisée par 12 valeurs (12 variables actives = ici 12 températures de l’année),
    - On aimerait représenter les individus (les villes) géométriquement dans un nuage de points,
    - Or nous ne savons représenter des points qu’avec 3 coordonnées au maximum (3 dimensions) dans l’espace. Or ici chaque individu (chaque ville) est représenté par 12 coordonnées.
    - Comment faire ?
    - L’ACP, par des calculs mathématiques (voir PDF du cours) va tenter de projetter les points de l’espace de n dimensions (ici 12 dimensions) sur un plan à deux dimensions. Or il y a une infinité de plans sur lesquels on peut projeter ces points.
    - L’ACP, par des calculs de distances et d’angles entre les points dans l’espace d’origine (n dimensions) , va choisir le meilleur plan qui respecterai ces distances et ses angles.
    - Dans Dim 1 et Dim 2 sont deux axes sans unité d’un plan choisi par l’ACP pour nous représenter les individus,
    - Les pourcentages entre parenthèses pour Dim 1 et Dim 2 :

86,87 % **+** **11,42 % =** **98,29 %** de fidélité de projection des points dans l'espace de 12 dimensions en ce plan de deux dimensions

Imaginons que l'espace est représenté par une Sphère (Jaune) et que dans notre étude les individus sont dans cette sphère (ici on représente Séville et Helsinki) :



On a représenté 3 plans au hasard (rose, rouge et bleu), parmi une infinité de plans possibles, sur lesquels, on pourrait projeter orthogonalement ces deux individus :

Quel est le meilleur plan qui donnerait :

- ➔ Presque la **même distance** entre Séville et Helsinki dans l'espace (Sphère Jaune) et dans le plan choisi (rose ou rouge ou bleu ou autre) ?
- ➔ Presque le **même angle** entre les deux vecteurs partant du centre de la sphère et rejoignant Séville et Helsinki et l'angle les deux vecteurs partant du centre du plan choisi (rose ou rouge ou bleu ou autre),et rejoignant Séville et Helsinki ?

L'ACP se charge, pour nous de :

- ➔ Trouver le meilleur plan pour la projection des individus,
- ➔ Et la fidélité de la projection se mesure en sommant les pourcentages des deux axes (dimensions) choisis sur ce plan.

Que représentent les deux dimensions dans le graphe des individus?

- ➔ **La dimension 1** :
  - Représente l'unité des variables actives,
  - Ici les individus représentent les villes et les villes ont comme coordonnées 12 variables mesurées en degrés Celsius.
  - On peut donc affirmer que la dimension 1 représente à gauche un temps **froid**, au milieu un temps **doux** et à droite un temps **chaud**,

- On dit que la Dim 1 est celle qui représente le mieux les points,

➔ **La dimension 2** :

- Représente l'amplitude (la variabilité) des valeurs actives,
- Plus on monte vers le haut, plus il y a des différences entre les températures les plus froides et les températures les plus chaudes,

Comment interpréter le graphe des individus ?

Observez ce graphe dans le logiciel R :

- Plus les individus ont une grande valeur sur la Dim 1, plus elles sont bien représentées :
  - Par exemple : les villes « Reykjavik », « Séville », « Palerme » et « Athènes » ont une coordonnée grande sur Dim 1 : elles sont bien représentées par projection sur ce plan,
- Il y a des individus très proches :
  - Par exemple « **Gènes** » et « **Barcelone** » sont très proches : elles ont un peu près en moyenne les mêmes températures, et ce quel que soit le mois de l'année,
  - Idem pour « **Cracovie** » et « **Prague** »,
  - Par contre « **Helsinki** » et « **Séville** » ont des comportements très différents. Ces deux villes sont complétement opposées sur la Dim 1. Mais maintenant qu'est ce qui oppose **Helsinki** à **Séville** ?
    - Si je connais bien le sujet : je peux affirmer que c'est évident qu'à Helsinki il fait plus froid qu'à Séville. (c'est un raisonnement général dans la vie courante)
    - Mais je veux me baser uniquement sur les données analysées (35 villes et 12 températures moyennes par ville) : il faudra compléter l'analyse l'étude du graphe des variables (les mois),

Analyser les variables

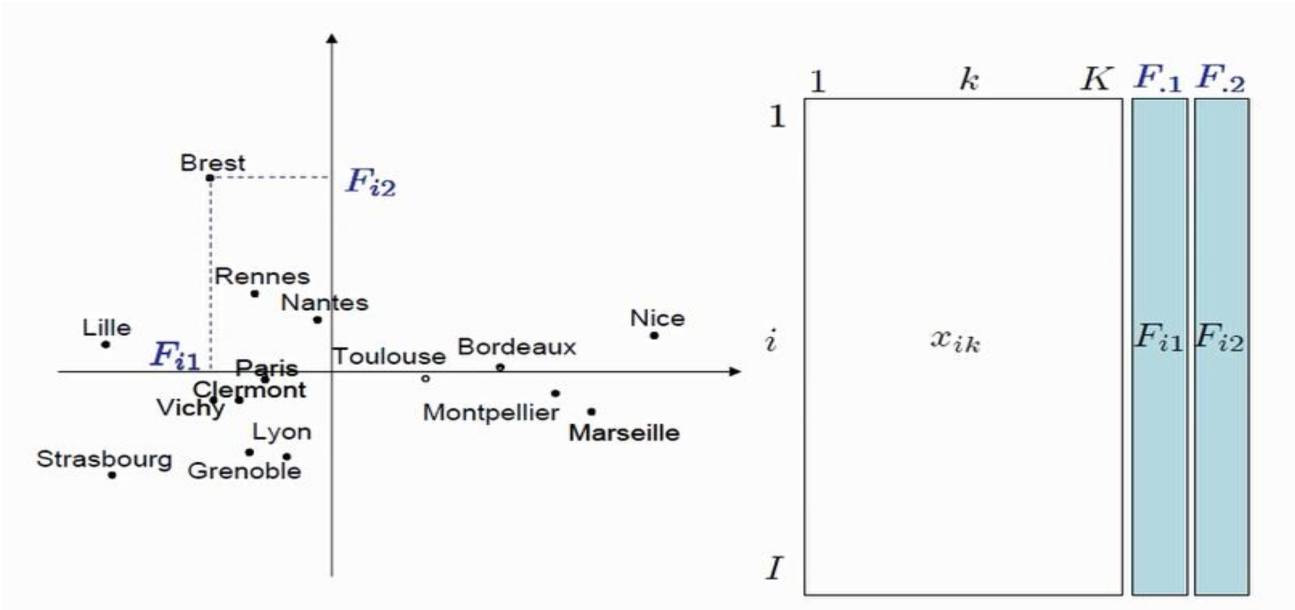
Comment interpréter le graphe des variables ?

Observez ce graphe qu'on appelle aussi « **Le Cercle des corrélations** » :

- Dans le graphe des individus :

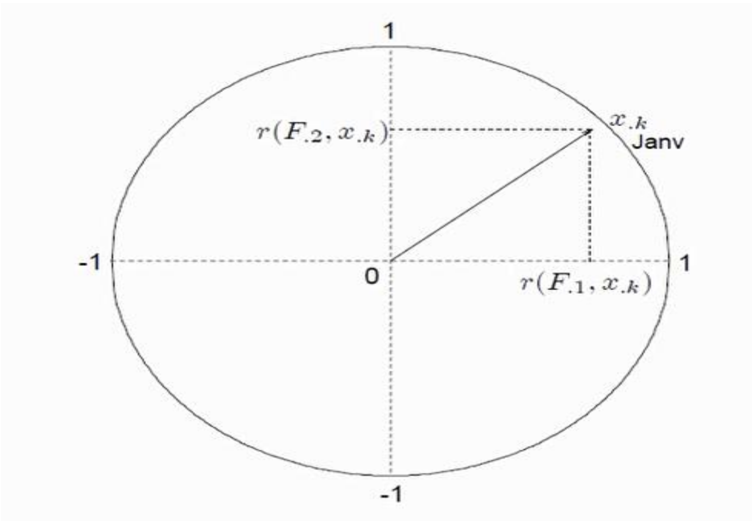
  Considérons les coordonnées des individus comme des variables

  Nous allons identifier chaque individu (par exemple Brest est l'individu n° 1) par sa coordonnée sur Dim1 : F_{11} et sa coordonnée sur Dim2 : F_{12} :



- Donc j'aurai pour chaque variable (par exemple Janvier) un ensemble de valeur représentent les coordonnées
- Sur Dim 1 des 35 valeurs : c'est le vecteur **F_{i1}** et un ensemble de valeur représentent les coordonnées

  Sur Dim 2 des 35 valeurs : c'est le vecteur **F_{i2}**
- On représente sur un cercle les variables par des vecteurs : corrélation entre une variable et les deux vecteurs :



- Sur Le cercle complet des corrélations :
- On a ici toutes les variables qui sont à droite de la Dim 1 (ce n'est pas toujours le cas).

- Quand **une flèche d'un vecteur touche le cercle** : on dit qu'elle est corrélée à Dim 1. C'est-à-dire bien représentée,

- Ici nous avons toutes les variables qui sont bien corrélées à Dim 1, de corrélation positive (à droite) et de valeur supérieure ou égale à environ 0,7. (**coefficient de corrélation =1 : variable très bien corrélée**),

- **<u>Deux vecteurs qui sont presque très proches</u>** (presque colinéaires) et de même sens :
  Les deux variables ont le même comportement. Ici par exemple « **Juin** » et « **Juillet** » ont à peu près le même comportement de variabilité de température,
- Ici deux variables qui sont très bien corrélées (car valeurs très grandes sur Dim 1) : « **Avril** » et « **Octobre** »,

<u>Comment interpréter les individus à partir du graphe des variables ?</u>

- Si nous reprenons **Avril** qui est bien corrélé à Dim 1 :
  - **<u>Les villes qui sont à gauche de Dim 1</u>** et qui ont une faible coordonnée sur Dim 1 (valeurs très négatives), prennent des faibles valeurs en Avril :
    - Ici : « **Raykjavik** », « **Helsinki** », « **Saint Petersburg** », … prennent de faibles valeurs en Avril (froid),
  - **<u>Les villes qui sont au milieu de Dim 1</u>** prennent des faibles valeurs moyennes en Avril :
    - Ici : « **Sofia** », « **Francfort** », « **Genève**», … prennent de moyennes valeurs en Avril (doux),
  - **<u>Les villes qui sont à droite de Dim 1</u>** prennent de fortes valeurs en Avril :
    - Ici : « **Séville** », « **Athènes**», « **Palerme**», … prennent de for tes valeurs en Avril (doux),
- On résume dans cet exemple :
  - **<u>Qu'est ce qui séparent les villes par rapport à la Dim 1 ?</u>**
    - Comme toutes les variables (les mois) sont à droite de Dim1 :
      - A droite, on a ou les individus (les villes) qui ont de fortes valeurs sur Dim 1 tous les mois de l'année : **il y est fait plutôt chaud,**
      - A gauche, tous les individus (les villes) qui ont de faibles valeurs tous les mois de L'année : **il y est fait plutôt froid,**

  - **<u>Qu'est ce qui séparent les villes par rapport à la Dim 2 ?</u>**
    - Dimension (Axe) de variabilité des variables (amplitude thermique),
    - En haut de Dim 2: villes avec petite amplitude thermique, en bas de Dim2 : villes avec forte amplitude thermique,
    - En haut du graphe des individus (forte variabilité), on va trouver les villes où il fait plutôt chaud en **Juillet** et **Juin**, et plutôt froid en Décembre, Janvier : **Moscou**, **Budapest**, **Kiev**

<u>Autre indicateurs d'analyse</u>

On a des variables supplémentaires :

➔ Ici Région et Amplitude : elles ne servent pas à construire les dimensions,
➔ Variable supplémentaire numérique : projetée sur le graphe des variables,
➔ Variable supplémentaire qualitative (facteur) : projetée sur le graphe des individus,
➔ Ici :
- L'amplitude est fortement corrélée positivement à Dim 2 : villes en haut : forte amplitude : par exemple Moscou (on passe de très froides températures à de très grandes températures),
- La région qui prend 4 valeurs (4 modalités) : Nord, Sud, Est, Ouest,
  - On a représenté le barycentre de chaque modalité ;
  - Cela m'indique en plus, par exemple, les villes qui sont au Sud : Rome, Gènes, …..

Qualité de représentation d'un individu et d'une variable

(Voir Volet « Sorties » sur la fenêtre de R Commander)

$cos2

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Janvier	**0.7851298**	**0.209319354**	0.0027104597	1.342339e-04	9.695892e-04
Février	**0.8383139**	**0.152348962**	0.0003745193	2.842813e-03	5.180285e-03
Mars	0.9185534	0.060900080	0.0089040613		

….

cos2

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Amsterdam	**0.15193054**	7.785927e-01	2.838786e-04	6.199515e-03	6.031806e-02
Athènes	**0.96806614**	2.251021e-02	6.743766e-03	2.245916e-03	7.169556e-07
Berlin	**0.87544892**	8.663055e-04	3.555474e-02	2.496223e-03	4.785784e-02
Bruxelles	0.02524338	9.373941e-01 1.3			

…

Contribution des variables et des individus à la construction de Dim1 et Dim 2

$contrib

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Janvier	7.531617	**15.2732245**	2.2492018	0.3170905	4.22980196
Février	8.041803	11.1163151	0.3107848	6.7153618	22.59882892
Mars	8.811526	4.4436435	7.3887947	23.4055358	0.50133217
Avril	9.214820	0.2422683	27.7538220	3.3103562	0.02336811
Mai	7.983460	9.4536838	25.6073120		

```
….
```

Fortes coordonnées : forte contribution

```
contrib
               Dim.1          Dim.2          Dim.3          Dim.4
Amsterdam      0.083620931  3.259529087    1.351579e-02   0.840237468
Athènes        10.822798022 1.914206418    6.521924e+00   6.183047023
Berlin         0.284208341    0.002139201  9.984865e-01   0.199555524
Bruxelles      0.008387938    2.369215441  3.883717e-02    0.543104625
Budapest       0.174547245    5.420831540  5.733259e+00   0.648203056
…..
```