

SY09 - TP1

Classification automatique

Pierre Bathellier et Vançon Virgile

Université de Technologie de Compiègne

Introduction

1 Exercice 1 : Visualisation des données

1.1 Données Iris

La **figure 1** ci-dessous représente les données Iris dans le premier plan factoriel sans tenir compte de l'espèce (après réalisation de l'ACP sur ces données via la fonction `princomp` de R). Avec cette représentation, le jeu de données semble comporter 2 groupes de points différents. La **figure 2** nous montre la même représentation mais avec des couleurs différentes en fonction de l'espèce. On constate que le groupe comportant le plus d'individus rassemble 2 espèces (*versicolor* et *virginia*), même si elles sont séparées au sein de ce groupe. On peut s'attendre à obtenir 2 groupes distincts si l'on recherche une partition des données différentes, alors que trois espèces sont présentes.

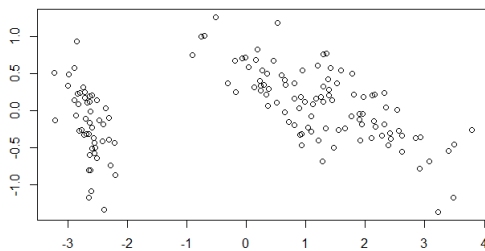


FIGURE 1. Représentation des données Iris dans le premier plan factoriel sans tenir compte de l'espèce

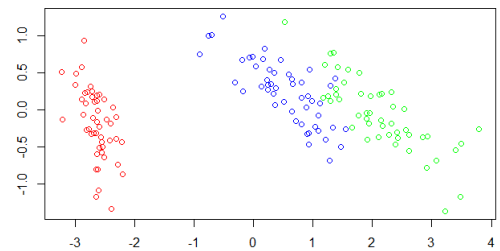


FIGURE 2. Représentation des données Iris dans le premier plan factoriel en tenant compte de l'espèce

1.2 Données Crabs

La **figure 3** ci-dessous représente les données Crabs dans le premier plan factoriel sans tenir compte de l'espèce. Avec cette représentation, 2 groupes de points se détachent clairement (séparés verticalement), mais on peut deviner qu'ils peuvent eux mêmes être séparés horizontalement (les points étant plus concentrés aux extrémités) pour former 4 groupes en tout. La **figure 4** nous montre la même représentation mais avec des couleurs différentes en fonction du sexe, et des symboles différents en fonction de l'espèce. Les 4 groupes qui semblaient se détacher apparaissent nettement désormais.

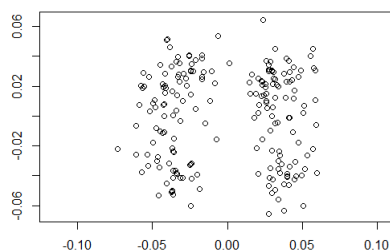


FIGURE 3. Représentation des données Crabs dans le premier plan factoriel sans tenir compte de l'espèce

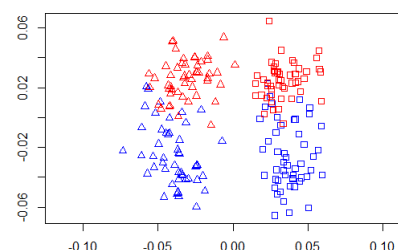


FIGURE 4. Représentation des données Crabs dans le premier plan factoriel en tenant compte de l'espèce et du sexe

1.3 Données Mutations

L'AFTD (Analyse Factorielle d'un Tableau de Distances) est réalisé sous R au moyen de la commande *cmdscale*. Pour des données se présentant sous la forme d'un tableau de dissimilarité, l'AFTD calcul une représentation multidimensionnelle de ces individus.

Pour évaluer la qualité de la représentation, il est intéressant d'utiliser un diagramme de Shepard : ce diagramme représente les distances déterminées par l'AFTD en fonction de la dissimilarité initiale. Sur le diagramme apparaît la bissectrice (droite d'équation $y = x$), et les points se situent sous cette bissectrice puisque les distances déterminées par l'AFTD doivent être inférieures ou égales aux dissimilarités (la représentation est exacte lorsque les dissimilarités sont des distances euclidiennes).

Plus le nombre de variables de représentation est élevé, plus la qualité de la représentation l'est aussi. Avec 2 variables de représentation comme sur la **figure 5**, les points sont beaucoup plus dispersés et la qualité de la représentation est faible (la relation entre distance et dissimilarité est croissante mais peu linéaire). En augmentant le nombre de variables de représentation ($d = 5$) comme sur la **figure 6**, la qualité de la représentation est bien meilleure, et les points forment une droite qui se confond avec la bissectrice.



FIGURE 5. Diagramme de Shepard avec 2 dimensions et la projection de Kruskal **FIGURE 6.** Diagramme de Shepard avec 5 dimensions et la projection de Kruskal

2 Exercice 2 : Classification hiérarchique

2.1 Classification hiérarchique ascendante des données Mutations

Sous R, la fonction `hclust` permet d'effectuer une classification hiérarchique ascendante. Différentes méthodes d'agrégation sont disponibles, comme la méthode de Ward (qui consiste à regrouper les classes de manière à ce que l'augmentation de l'inertie interclasse soit maximum, ou ce qui revient au même d'après le théorème de Huygens, de façon que l'augmentation de l'inertie intraclasse soit minimum), la méthode "single linkage" (correspondant au choix du meilleur cas) ou encore la méthode du "complete linkage" (correspondant au choix du pire cas).

Quelle que soit la méthode utilisée, 3 espèces se distinguent des autres : *Bakers Mould*, *Bread Yeast* et *Skin Fungus*. L'écart entre ces 3 espèces et les autres est immense, bien plus important qu'entre n'importe quelle autre espèce. La majorité des méthodes rassemble ces 3 espèces au sein d'un même groupe, mais la méthode "single linkage" sépare *Bakers Mould* de *Bread Yeast* et *Skin Fungus*, avec cependant un écart minime.

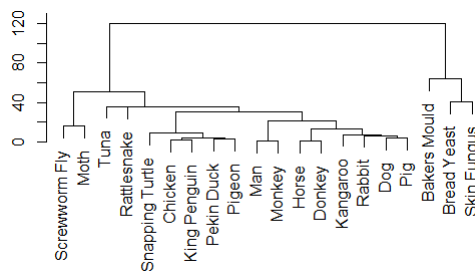


FIGURE 7. Classification hiérarchique ascendante avec la méthode de Ward.d2

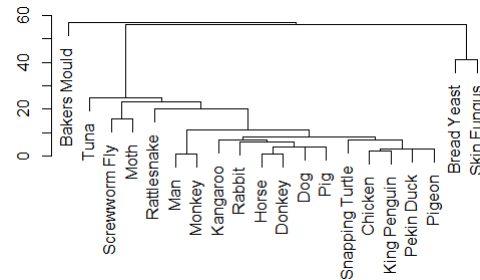


FIGURE 8. Classification hiérarchique ascendante avec la méthode single linkage

2.2 Classification hiérarchique ascendante des données Iris

La classification des données Iris effectuée simplement avec la fonction `hclust` de R est inexploitable : il y a bien trop de branches différentes, comme le montre la **figure 9**. Il nous faut donc réaliser un analyse plus pointu que pour les données *Mutations*.

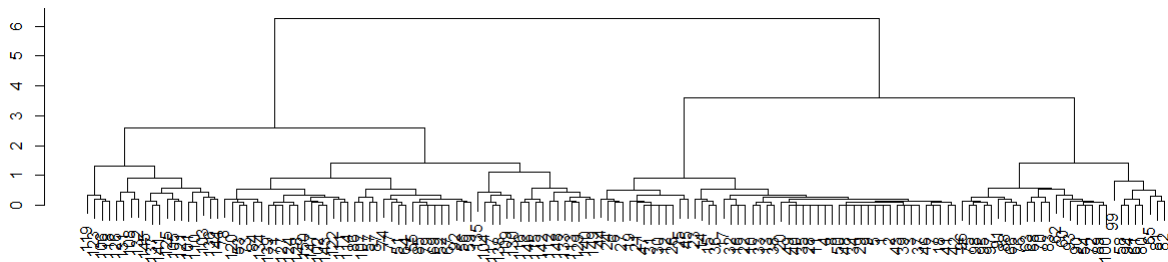


FIGURE 9. Classification hiérarchique ascendante des données Iris

Il est alors intéressant d'utiliser des méthodes plus complexes permettant de colorer chaque individu en fonction de son espèce. Par défaut, la fonction `hclus` utilise l'algorithme de *complete linkage*. Comme expliqué précédemment, cet algorithme choisit comme représentant de chaque groupe l'individu le moins similaire avec les autres individus des autres groupes. La similarité entre cluster est donc le minimum de similarité entre les individus des clusters. La **figure 10** ci-contre montre la classification avec une coloration par espèce (vert pour *versicolor*, rouge pour *virginica* et noir pour *setosa*). Si les *setosas* semblent toutes être dans le même cluster, la séparation entre les *versicolor* et les *virginicas* est moins bonne dans la mesure où l'on a pas une séparation claire entre les deux espèces et que certains *versicolor* sont assimilés au cluster correspondant aux *virginicas*.

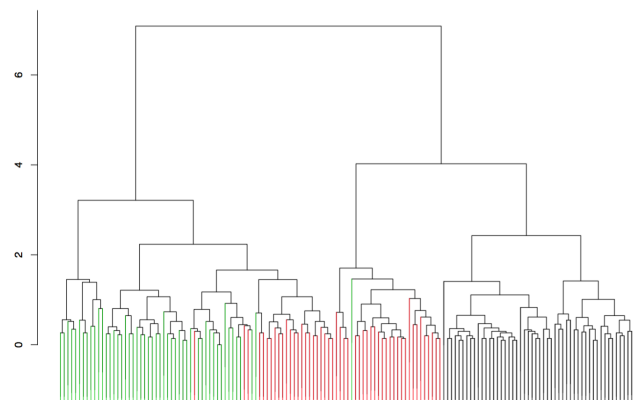


FIGURE 10. Classification hiérarchique ascendante colorée par espèce, méthode *complete linkage*

Si l'on découpe notre arbre en 3 clusters (qui pourraient correspondre à nos 3 espèces) on obtient la répartition représentée par la **figure 11**. Comme nous avons pu l'observer, l'algorithme a bien classifié les *setosas* (les 50 sont dans le cluster numéro 1) et quasiment tous les *virginicas* (49 sur 50 sont dans le cluster numéro 2), mais pas les *versicolors* (23 sont dans le cluster 2 et 27 seulement le sont dans le cluster numéro 3). Ceci est dû au fait que les *versicolors* et les *virginicas* sont proche les uns des autres, si bien que leurs nuages de points s'entremêlent comme le montre la **figure 12**. Sur cette figure, on représente les individus colorés selon leur cluster dans le premier plan factoriel de l'ACP. Les points ne possédant pas la même couleur à l'intérieur (couleur du cluster) et à l'extérieur (couleur de l'espèce) représentent les individus qui ont été mal classés, c'est à dire les 23 *versicolors* du cluster 2.

```
> table(clusterCut, iris$Species)
```

clusterCut	setosa	versicolor	virginica
1	50	0	0
2	0	23	49
3	0	27	1

FIGURE 11. Effectif du clustercut, méthode *complete linkage*

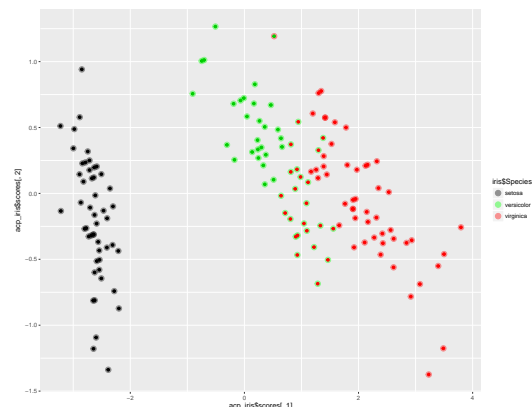


FIGURE 12. Coloration des individus en fonction de leur espèce et de leur classification, représentation dans le premier plan factoriel de l'ACP, méthode *complete linkage*

Pour tenter d'obtenir une meilleur classification des données Iris, on utilise une autre méthode. La fonction *hclust* utilisée avec le critère de Ward fournit en général des clusters de qualité car homogènes. La classification obtenue avec cette méthode est un peu meilleure : on obtient 16 individus mal classés (15 *virginicas* et 1 *versicolor*, contre 28 précédemment).

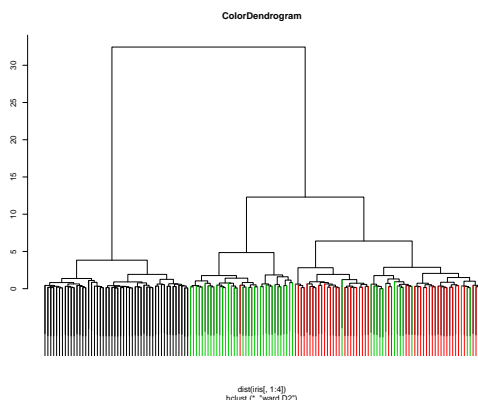


FIGURE 13. Classification hiérarchique ascendante colorée par espèce, critère de Ward

```
> table(clusterCut, iris$Species)
```

clusterCut	setosa	versicolor	virginica
1	50	0	0
2	0	49	15
3	0	1	35

FIGURE 14. Effectif du clustercut selon les espèces, critère de Ward

2.3 Classification hiérarchique descendante des données Iris

Les classifications ascendantes utilisées jusqu'à présent fonctionnent en regroupant les individus similaires au sein d'un même cluster : il faut donc un critère d'agrégation. Les méthodes de classification descendantes divisent les clusters au fur et à mesure : elles nécessitent un critère de séparation.

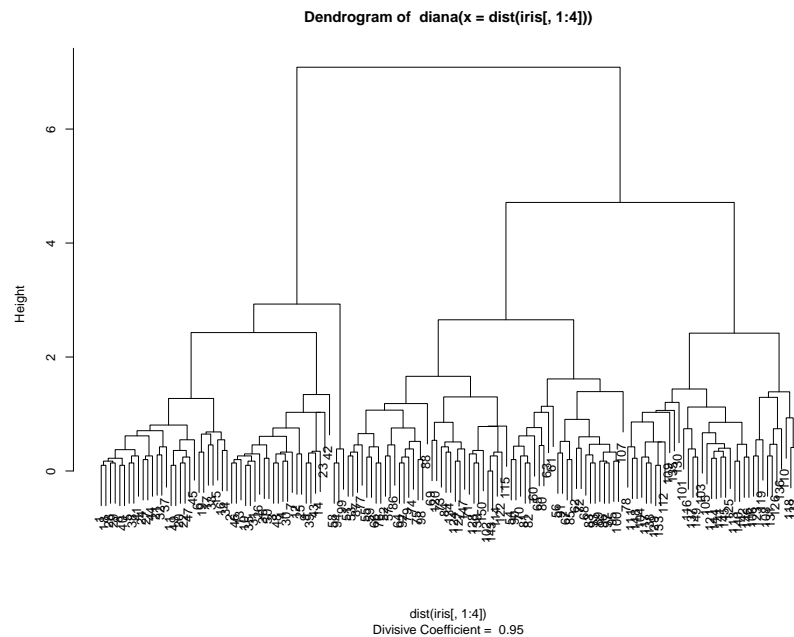


FIGURE 15. Classification hiérarchique descendante des données Iris

L'arbre ci-dessus (**figure 15**) montre la classification descendante des données Iris obtenu avec la fonction *diana*. Si l'on découpe cet arbre en 3 clusters, on obtient la table de contingence ci-contre (**figure 16**). 18 individus ne sont pas dans le bon cluster :

- 4 *versicolors*. 3 se trouvent dans le premier cluster (théoriquement celui des *setosas*) et 1 dans le troisième cluster (théoriquement celui des *virginicas*).
- 14 *virginicas* qui se trouvent dans le deuxième cluster (théoriquement celui des *versicolors*).

Si la mauvaise séparation entre *versicolor* et *virginica* n'est pas nouvelle et est compréhensible, le fait que trois *versicolors* se retrouvent dans le 1^{er} cluster est plus surprenant, étant donné la forte séparation entre ces espèces (cf **figure 12**). Il est alors logique de conclure que la classification hiérarchique descendante est moins efficace dans ce cas que la classification hiérarchique ascendante avec le critère de Ward.

```
> table(dianacut, iris$Species)
```

dianacut	setosa	versicolor	virginica
1	50	3	0
2	0	46	14
3	0	1	36

FIGURE 16. Effectif du clustercut classé selon les espèces, classification hiérarchique descendante

3 Exercice 3 : Méthode des centre mobiles

3.1 Algorithme des *k-means* sur les données Iris

L'algorithme des *k-means* est une méthode de partitionnement de données, basée sur l'algorithme suivant :

- Initiation : choix de k centres au hasard dans l'ensemble des données
- Répéter jusqu'à ce qu'il y ait convergence :
 - affectation de chaque x_i au centre c_k le plus proche
 - Mise à jour des centres : $c_k = \frac{\sum_{i=1}^n z_{ik} x_i}{\sum_{i=1}^n z_{ik}}$

Lorsque l'on applique cet algorithme aux données Iris, avec 2 clusters, on obtient la répartition des effectifs selon les espèces présentée **figure 17**. Alors qu'on pouvait s'attendre à avoir un cluster de 50 individus regroupant toutes les *setosas* (qui sont représentées par le groupe de points très concentrés à gauche du graphique **figure 18**) et un cluster de 100 individus regroupant les *virginicas* et les *versicolores* (représentées par le gros nuage de points), on obtient en réalité un premier cluster de 53 individus avec 3 *versicolores* regroupées avec les *setosas*. Cette "anomalie" avait déjà été observée avec la classification hiérarchique descendante.

Ceci peut s'expliquer par le fait que le gros nuage de points regroupant les *versicolores* et les *virginicas* n'est pas rond mais ovale, et comme l'algorithme des *k-means* cherche à minimiser la distance entre les points des clusters et leurs centre de gravité, il fonctionne mieux avec les classes rondes.

```
> table(ki2$cluster, iris$Species)
```

	setosa	versicolor	virginica
1	50	3	0
2	0	47	50

FIGURE 17. Effectifs des 2 clusters selon les espèces, algorithme des *kmeans* $k=2$

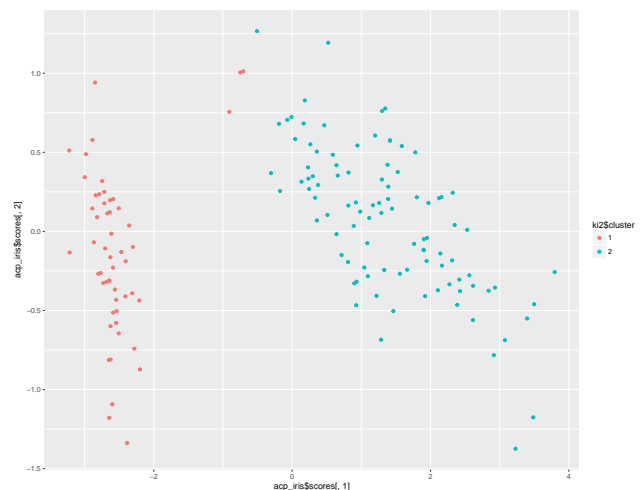


FIGURE 18. Graphique de représentation des des individus Iris dans le premier plan factoriel de l'ACP et coloré selon les clusters obtenues en appliquant algorithme des *kmeans* ($k=2$).

Si l'on cherche à obtenir 3 clusters en appliquant l'algorithme des *kmeans* ($k=3$), toutes les *setosas* se retrouvent ensemble dans un même cluster. Les *versicolores* et les *virginicas* sont réparties dans 2 clusters de tailles inégales avec 14 *virginicas* dans le cluster des *versicolores* et 3 *versicolores* dans le cluster des *virginicas*. La distinction entre ces deux espèces est une nouvelle fois difficile, tellement elles sont proches l'une de l'autre.

	setosa	versicolor	virginica
1	0	3	36
2	50	0	0
3	0	47	14

FIGURE 19. Effectifs des 3 clusters selon les espèces, algorithme des *kmeans* $k=3$.

Avec 4 clusters, il est intéressant de noter que les *versicolors* semblent être séparées en 2 variétés au sein de la même espèce (23 dans le cluster 1, 27 dans le cluster 2). Les *setosas* sont une fois encore seules dans leur cluster.

	setosa	versicolor	virginica
1	0	23	17
2	0	27	1
3	50	0	0
4	0	0	32

FIGURE 21. Effectifs des 4 clusters selon les espèces, algorithme des *kmeans* $k=4$

3.2 Iris - Stabilité du résultat de la partition

L'algorithme des *k-means* semble stable. En effet, dans la majorité des cas, la répartition entre cluster est la même que celle de la **figure 19** avec 3 *versicolors* dans le cluster des *virginicas* et 14 *virginicas* dans le cluster des *versicolors*. Mais comme l'algorithme des *k-means* a un caractère aléatoire (à cause du choix initial des centres) dans d'autre cas on observe la répartition de la **figure 23**, plus rare que la précédente. Avec cette répartition, le cluster des *setosas* est séparé en 2, et la quasi-totalité des *versicolors* et des *virginicas* se retrouve dans le même cluster. Cette répartition se trouve être plus rare que la précédente car beaucoup moins stable ; l'inertie intra-classe du cluster 2 est 3 fois supérieure à celle de la configuration précédente (celle des 2 autres cluster étant sensiblement la même).

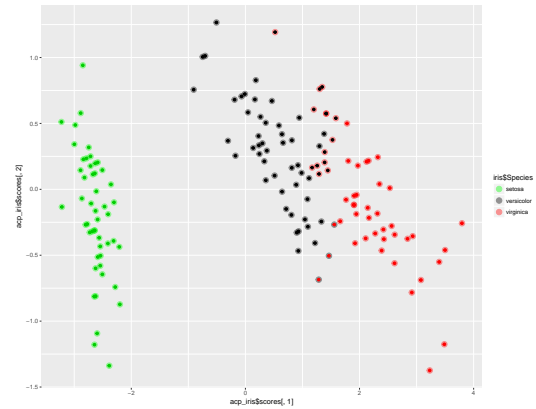


FIGURE 20. Graphique de représentation des des individus Iris dans le premier plan factoriel de l'ACP et coloré selon les clusters obtenues en appliquant algorithme des *kmeans* ($k=3$).

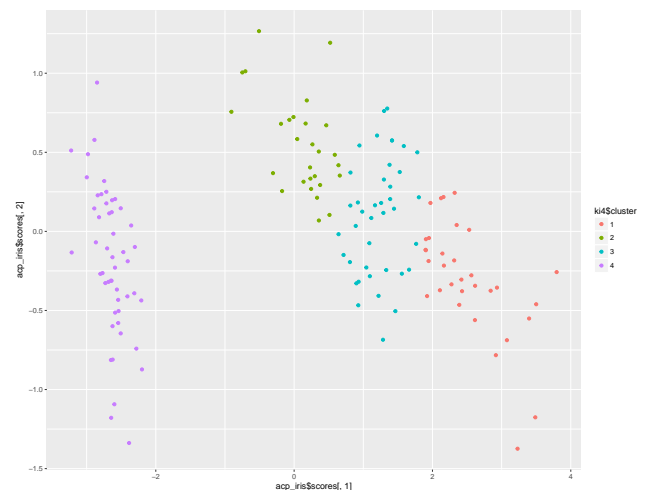


FIGURE 22. Graphique représentant les individus du jeu de données iris colorées selon les clusters dans le premier plan factoriel de l'ACP ,algorithme des *kmeans*.

```

      setosa versicolor virginica
1       18          4         0
2        0         46        50
3       32          0         0
> k13$withinss
[1] 17.221058 105.683170  4.749377

```

FIGURE 23. Effectifs des 3 clusters (bis) triés selon l'espèce, algorithme des *kmeans*

3.3 Iris - Détermination du nombre de classes optimal

La méthode des centres mobiles nécessite de connaître à l'avance le nombre de classes auquel on va aboutir. Un moyen d'estimer le nombre de classes optimal pour découper les données est d'appliquer la méthode du coude. A l'aide de l'évolution de l'inertie intra-classe minimale en fonction du nombre K de classes on peut déterminer un point de rupture. Ce point de rupture estime le nombre optimal de classes pour partitionner l'échantillon.

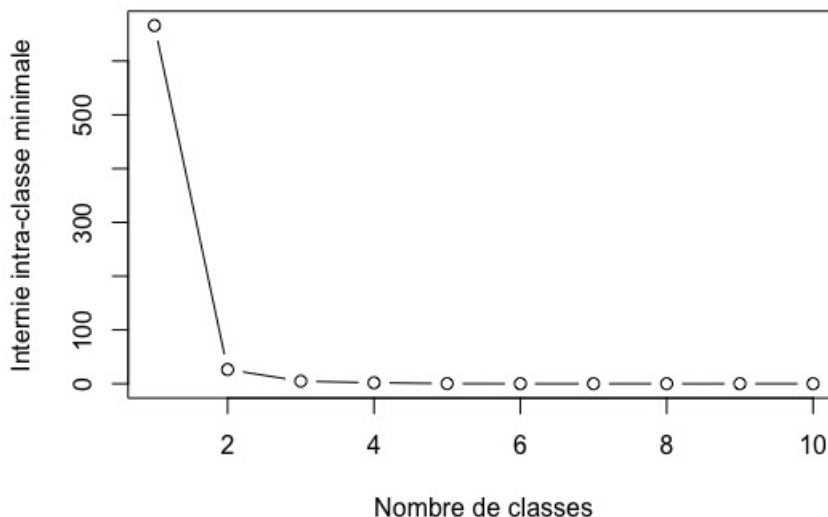


FIGURE 24. Inertie intra-classe minale parmi 100 classifications pour K le nombre de classes allant de 2 à 10

Si l'on observe la **figure 24** on remarque une cassure entre $k=2$ et $k=3$ classes. On observe en effet une grande différence entre $k=1$ (correspond à l'inertie totale) puis entre $k=2$ et $k=3$. On peut donc imaginer que le nombre de classe optimal de classe est $k=2$ ou $k=3$.

3.4 Iris - Comparaison avec la partition réelle en 3 groupes

Le jeu de données Iris peut être décomposé en 3 classes qui correspondent aux espèces : *Setosa*, *Versicolor*, et *Virginica*.

L'indice ajusté de Rand permet de comparer la partition obtenue en utilisant la méthodes des centres mobiles avec la partition réelle. La fonction *adjustedRandIndex* permet de comparer deux classifications en déterminant les paires de points classées identiquement.

```
> adjustedRandIndex(ki3$cluster, iris$Species)
[1] 0.7163421
```

FIGURE 25. Indice ajusté de Rand de la partition obtenue avec les méthodes des kmeans (pour $k = 3$) avec la partition réelle en fonction de l'espèce

On obtient un indice ajusté égal à 72%. Cela signifie que la partion obtenue en appliquant les k-means et la partition réelle sont similaire sur presque les trois-quarts de la population. Les différences sont dues à des erreurs entre les *versicolor* et les *virginicas* qui comme on a pu le constater précédemment, sont deux espèces qui semblent assez proches.

3.5 Crabs - classification en 2 classes

On réalise plusieurs classifications pour $k = 2$ classes. Dans la très grande majorité des cas, la classification semble se faire en fonction de l'espèce. Seul 1 individu est mal classé. On constate graphiquement **figure 26** que cet individu est situé de manière équidistante vis à vis des deux classes.

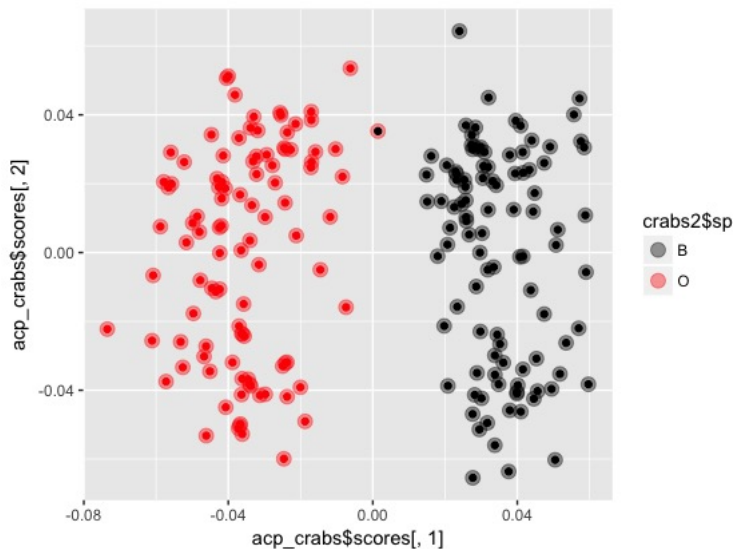


FIGURE 26. Représentation des individus crabs dans le premier plan factoriel (la couleur du centre correspond à la classification et le contour à l'espèce d'appartenance)

```
> table(crabsCluster$cluster, crabs2$sp)
```

	B	O
1	100	1
2	0	99

FIGURE 27. Effectifs obtenus en appliquant les kmeans ($k=2$) et triés selon les espèces.

Cependant, de manière moins fréquente, une autre classification est obtenue pour $k=2$. Cette classification semble se faire en fonction du sexe comme on peut le voir sur les figures. La classification est moins précise puisqu'on obtient cette fois-ci un plus grand nombre d'individu mal classé.

Il est logique que la classification se face davantage et soit plus précise en fonction de l'espèce et plus rarement en fonction du sexe, la distinction entre les espèces est plus prononcée que celle entre les sexes qui sont eux assez proches sur le premier plan factoriel de l'ACP **figure 26** et **figure 28**.

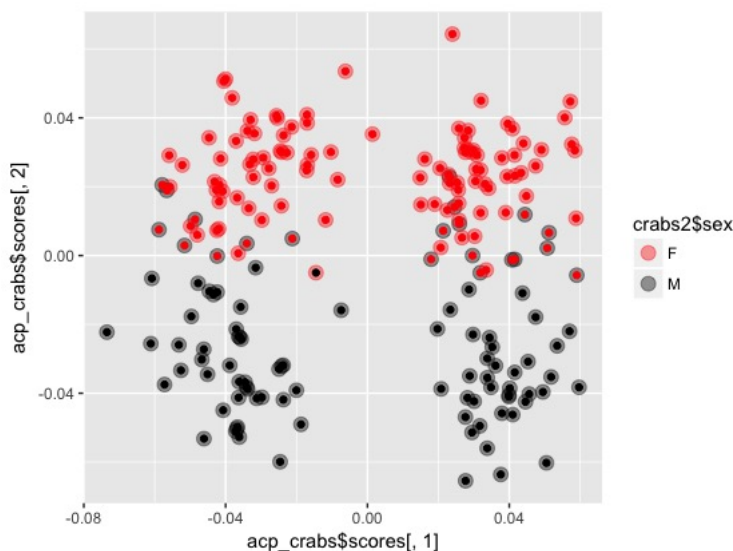


FIGURE 28. Représentation des individus crabs dans le premier plan factoriel (la couleur du centre correspond à la classification et le contour au sexe de l'individu)

```
> table(crabsCluster$cluster, crabs2$sex)
```

	F	M
1	1	79
2	99	21

FIGURE 29. Effectifs obtenus en appliquant les kmeans ($k=2$) et triés selon le sexe.

3.6 Crabs - classification en 4 classes

La classification en 4 classes permet d'obtenir une classification en fonction du sexe et de l'espèce. En effet, chaque cluster correspond à une des quatre combinaisons sexe/espèce possible : Orange/Female, Orange/Male, Blue/Female, Blue/Male.

La classification en 4 clusters est la combinaison des deux classifications que l'on pouvait obtenir pour $k=2$.

La **figure 30** permet de visualiser la classification. On introduit pour cela une colonne dans le jeu de données `species_and_sexe` qui vaut :

- FO pour un individu Orange/Female.
- MO pour un individu Orange/Male.
- FB pour un individu Blue/Female.
- MB pour un individu Blue/Male.

On observe graphiquement que les individus mal répartis correspondent aux individus dont les caractéristiques sont assez éloignées de la moyenne de leur catégorie espèce/sexe.

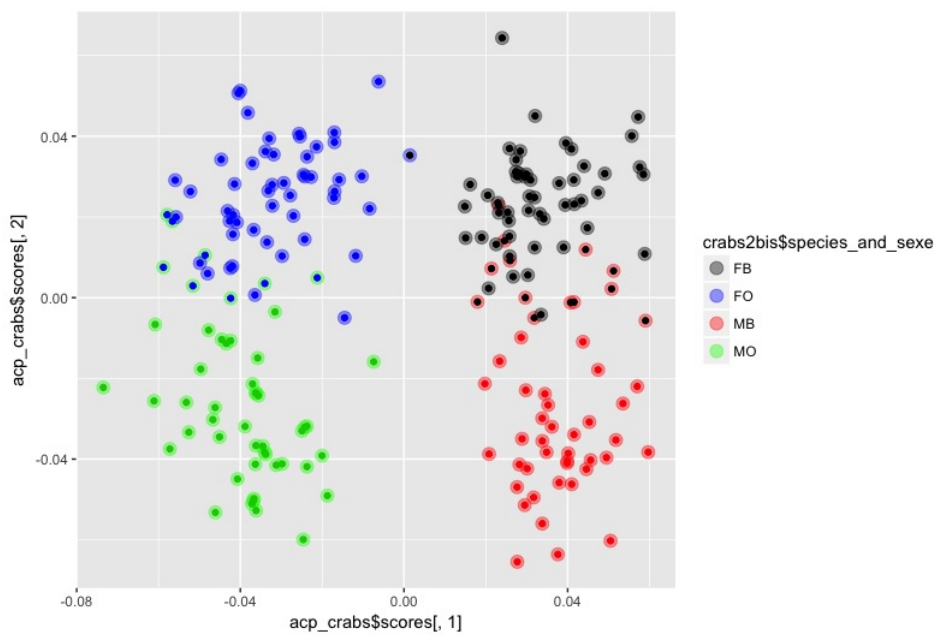


FIGURE 30. Représentation des individus crabs dans le premier plan factoriel de l'ACP (la couleur du centre correspond à la classification et le contour à la combinaison sexe/espèce de l'individu)

La classification semble assez stable. Cependant, nous avons obtenu une fois la classification présentée **figure 31**, alors qu'elle n'est pas fidèle à la réalité et scinde en 3 groupes l'espèce "Blue"

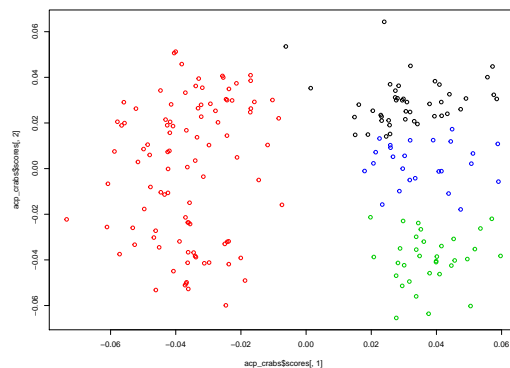


FIGURE 31. Représentation des individus crabs dans le premier plan factoriel de l'ACP (coloré selon les clusters)

3.7 Crabs - classification en 3 classes

Voici **figure 32** la partition la plus fréquente que l'on obtient en faisant tourner l'algorithme des kmeans sur une représentation des données mutations dans un espace de dimension $d=5$:

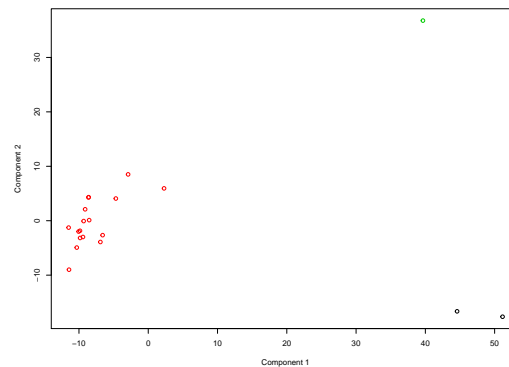


FIGURE 32. Représentation des individus mutations dans le premier plan factoriel de l'AFTD (coloré selon les clusters)

Cette représentation est en accord avec la classification hiérarchique ascendante des données Mutations que l'on avait obtenue **figure 6**.

Le nuage de gauche (points noirs) correspond à une classe, les deux points en bas à droite (Bread Yeast et Skin Fungus en vert) à une autre et finalement Bakers Mould (en rouge) forme une troisième classe. Graphiquement cette classification semble cohérente compte tenue de la forte dissimilarité entre les espèces marginales (Bread Yeast, Skin Fungus et Bakers Mould) du reste des individus.

Cette classification, bien qu'étant celle revenant le plus souvent, n'est pas stable. Les **figures 33 et 34** sont des exemples de classifications que l'on a pu obtenir. Cela illustre l'importance de l'initialisation des centres mobiles dans le résultat final de l'algorithme. Cette instabilité est due au fait que 2 des groupes que l'on cherche à identifier ne présente qu'un ou 2 éléments ce qui est faible par rapport au troisième et dernier groupe. Cela tend à démontrer que l'algorithme est davantage efficace pour classer des groupes de tailles homogènes.

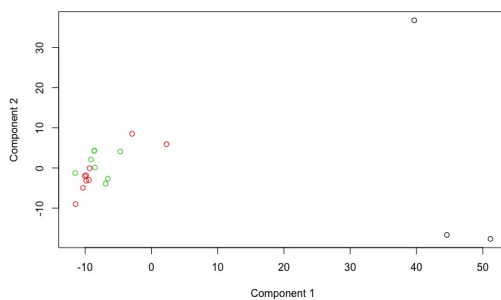


FIGURE 33. Représentation des individus mutations dans le premier plan factoriel de l'AFTD (coloré selon les clusters) - Bis

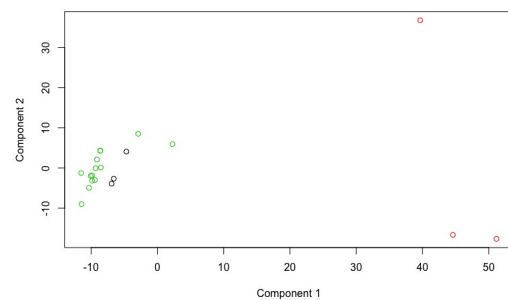


FIGURE 34. Représentation des individus mutations dans le premier plan factoriel de l'AFTD (coloré selon les clusters) - Ter

3.8 Conclusion

Nous avons dans ce nouveau TP expérimenté d'autres méthodes de représentation multidimensionnelle de données.

Celles-ci nous ont permis de partir de sources différentes (comme des tableaux de distances et de dissimilarités) afin de voir qu'il est tout à fait possible de condenser l'information pour obtenir rapidement et efficacement des représentations dans des espaces de faible dimension et donc interprétables par l'œil humain.

Les méthodes de classification fournissent une autre approche de ces méthodes et donnent d'autres représentations comme les dendogrammes qui sont utiles pour catégoriser les individus d'un jeu de données.

Table des figures

1	Représentation des données Iris dans le premier plan factoriel sans tenir compte de l'espèce	1
2	Représentation des données Iris dans le premier plan factoriel en tenant compte de l'espèce	1
3	Représentation des données Crabs dans le premier plan factoriel sans tenir compte de l'espèce	1
4	Représentation des données Crabs dans le premier plan factoriel en tenant compte de l'espèce et du sexe	1
5	Diagramme de Shepard avec 2 dimensions et la projection de Kruskal	2
6	Diagramme de Shepard avec 5 dimensions et la projection de Kruskal	2
7	Classification hiérarchique ascendante avec la méthode de Ward.d2	3
8	Classification hiérarchique ascendante avec la méthode single linkage	3
9	Classification hiérarchique ascendante des données Iris	3
10	Classification hiérarchique ascendante colorée par espèce, méthode <i>complete linkage</i>	3
11	Effectif du clustercut, méthode <i>complete linkage</i>	4
12	Coloration des individus en fonction de leur espèce et de leur classification, représentation dans le premier plan factoriel de l'ACP, méthode <i>complete linkage</i>	4
13	Classification hiérarchique ascendante colorée par espèce, critère de Ward	4
14	Effectif du clustercut selon les espèces, critère de Ward	4
15	Classification hiérarchique descendante des données Iris	5
16	Effectif du clustercut classé selon les espèces, classification hiérarchique descendante	5
17	Effectifs des 2 clusters selon les espèces, algorithme des <i>kmeans</i> k=2	6
18	Graphique de représentation des des individus Iris dans le premier plan factoriel de l'ACP et coloré selon les clusters obtenues en appliquant algorithme des <i>kmeans</i> (k=2).	6
19	Effectifs des 3 clusters selon les espèces, algorithme des <i>kmeans</i> k=3.	7
20	Graphique de représentation des des individus Iris dans le premier plan factoriel de l'ACP et coloré selon les clusters obtenues en appliquant algorithme des <i>kmeans</i> (k=3).	7
21	Effectifs des 4 clusters selon les espèces, algorithme des <i>kmeans</i> k=4	7
22	Graphique représentant les individus du jeu de données iris colorées selon les clusters dans le premier plan factoriel de l'ACP ,algorithme des <i>kmeans</i>	7
23	Effectifs des 3 clusters (bis) triés selon l'espèce, algorithme des <i>kmeans</i>	7
24	Inertie intra-classe minale parmi 100 classifications pour K le nombre de classes allant de 2 à 10	8
25	Indice ajusté de Rand de la partition obtenue avec les méthodes des kmeans (pour k = 3) avec la partition réelle en fonction de l'espèce	8
26	Représentation des individus crabs dans le premier plan factoriel (la couleur du centre correspond à la classification et le contour à l'espèce d'appartenance)	9
27	Effectifs obtenus en appliquant les kmeans (k=2) et triés selon les espèces.	9
28	Représentation des individus crabs dans le premier plan factoriel (la couleur du centre correspond à la classification et le contour au sexe de l'individu)	9
29	Effectifs obtenus en appliquant les kmeans (k=2) et triés selon le sexe.	9
30	Représentation des individus crabs dans le premier plan factoriel de l'ACP (la couleur du centre correspond à la classification et le contour à la combinaison sexe/espèce de l'individu)	10
31	Représentation des individus crabs dans le premier plan factoriel de l'ACP (coloré selon les clusters)	10
32	Représentation des individus mutations dans le premier plan factoriel de l'AFTD (coloré selon les clusters)	11
33	Représentation des individus mutations dans le premier plan factoriel de l'AFTD (coloré selon les clusters) - Bis	11
34	Représentation des individus mutations dans le premier plan factoriel de l'AFTD (coloré selon les clusters) - Ter	11