

UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE

GÉNIE INFORMATIQUE

Rapport du projet de SY09

Authors :

Crauser JULIEN et Antoine COLLAS

15 avril 2018



1 Cuisine

Question 1

Le jeu de données présent dans le fichier recettes-pays.data contient 51 variables pour 26 individus. Nous pouvons donc noter que nous disposons de plus de variables que d'individus. Parmi ces 51 variables, une seule est qualitative (les origines des recettes) et 50 sont quantitatives. Ces dernières prennent leurs valeurs entre 0 et 1 (0 et 0.82 pour être plus précis). Dans ce jeu de données les recettes ont été agrégées par origine, il y a donc 26 origines (une par ligne de notre tableau individus-variables). De plus, le jeu de données ne présente aucune valeurs manquantes.

Question 2

Nous réalisons une ACP sur le jeu de données. Nous obtenons 26 axes principaux : autant que d'individus. En effet, comme il y a moins d'individus que de variables, il est suffisant de prendre 26 axes pour représenter tous les individus.

Nous obtenons les pourcentages d'inertie expliquée visibles sur la figure 1. Les 3 premiers axes représentent 72% de la variance du nuage de points.

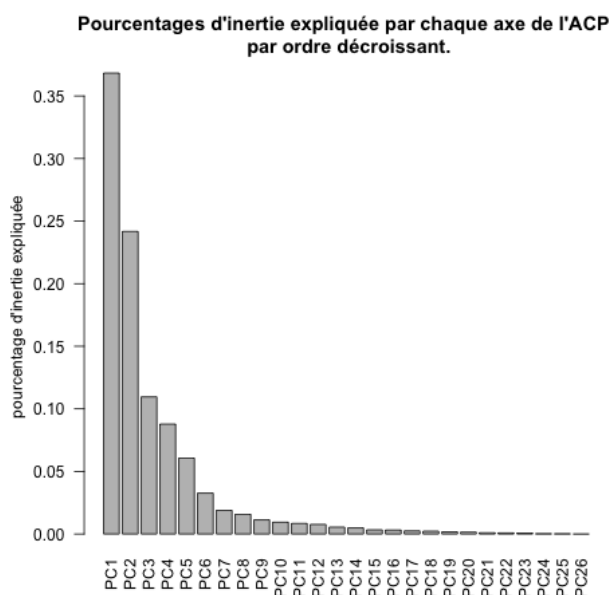


FIGURE 1 – Pourcentages d'inertie expliquée

Comme les 3 premiers axes factoriels représentent 72% de la variance totale, nous analysons le nuage de points seulement avec les 3 premiers plans factoriels. Nous obtenons les plans factoriels des figures 3, ?? et 4. Nous avons associé à chaque plan son cercle des corrélations.

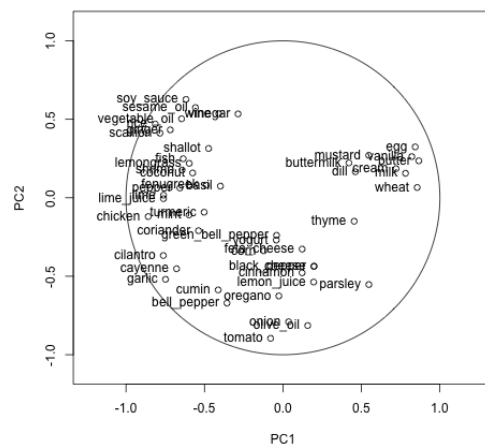
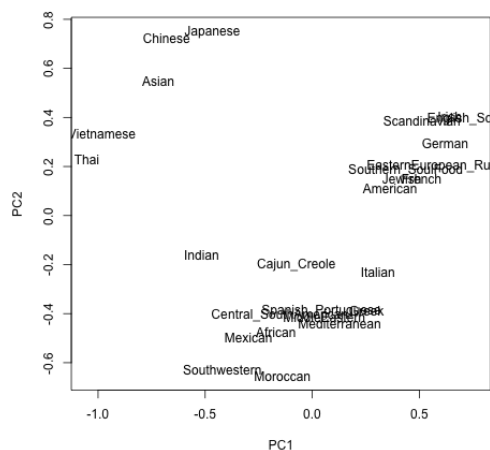


FIGURE 2 – Plan factoriel 1,2 avec son cercle associé

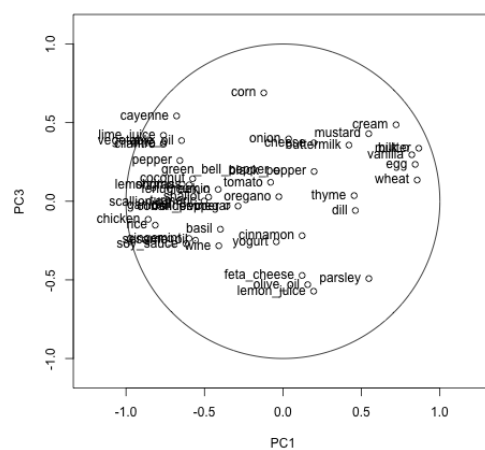
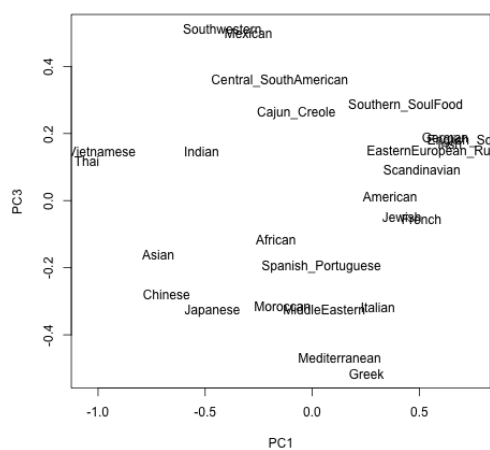
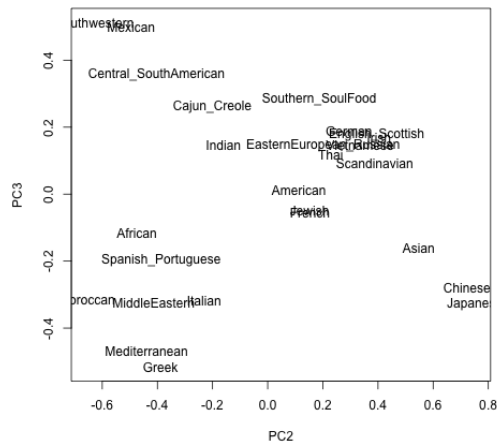
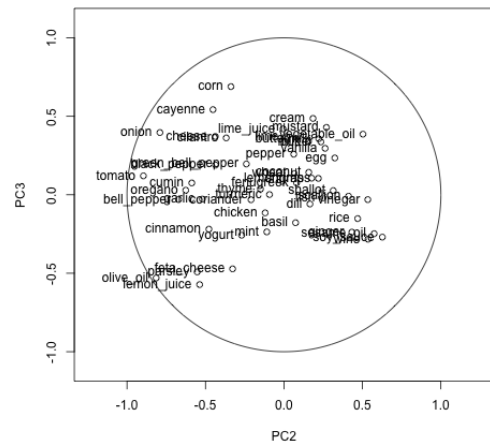


FIGURE 3 – Plan factoriel 1,3 avec son cercle associé



(a) Plan 2,3



(b) Cercle des corrélations

FIGURE 4 – Plan factoriel 2,3 avec son cercle associé

Dans le premier plan factoriel, nous observons que plusieurs groupes se forment. Les pays asiatiques se distinguent avec des ingrédients comme l'huile de sésame, l'huile de soja, le poisson, les échalottes... Un deuxième groupe se distingue : la cuisine occidentale avec des ingrédients comme le blé, les oeufs et le lait. Aussi, les pays consommant de la tomate et de l'huile d'olive sont exposés dans ce premier plan : Maroc, Mexique et Afrique.

Dans le deuxième plan factoriel, la Grèce et la méditerranée sont liés par leur consommation de féta, d'huile d'olive et de jus de citron. Ce plan permet aussi de voir la formation de deux sous groupes parmi les pays asiatiques. Le Vietnam et la Thaïlande consomment du poulet et du piment, ce qui est moins le cas de la Chine et du Japon.

Question 3

Nous obtenons le dendrogramme de la figure 5. Nous observons que trois groupes émergent : ...

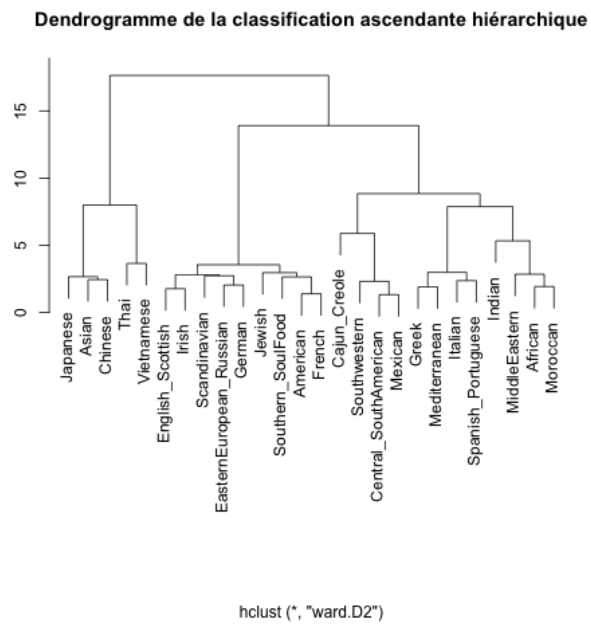


FIGURE 5 – Dendrogramme de la classification ascendante hiérarchique