

Projet d'Analyse de Données

BOCQUET Cécile, HOUNTON Jean-Phillipes, DJEMA Alissa

04/06/2020

Contents

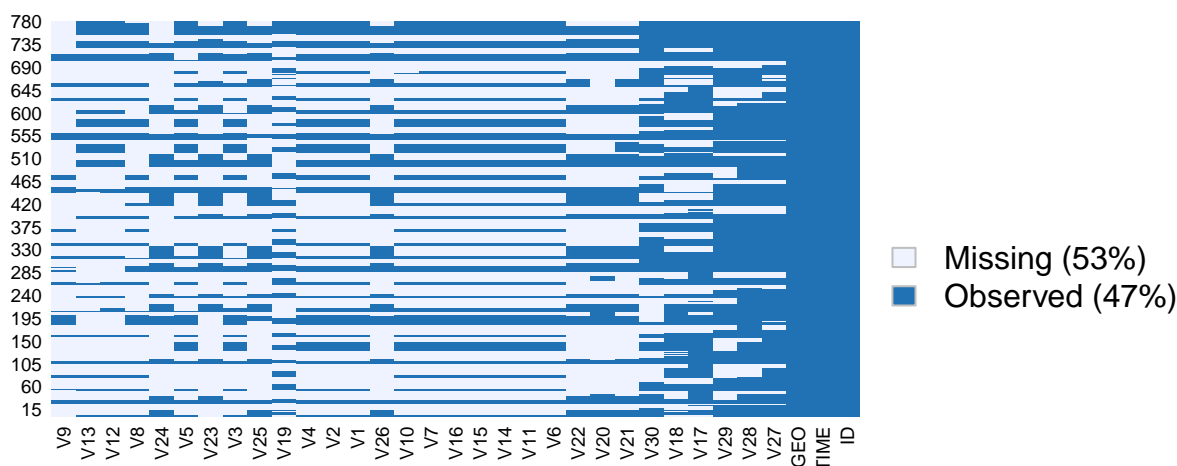
Introduction et motivations	2
I. Traitement des données	2
II. Études des variables	3
1. Analyses Univariées	3
2. Analyse Bivariée	6
III. Analyse exploratoire	7
1. Classification Hierarchique sur l'ACP – (Ensemble des individus)	7
a. Analyse en composante principale	7
b. Clustering Hiérarchique	11
1. Classification Hierarchique sur l'ACP – (Temporalité)	15
a. Analyse en composante principale	15
b. Classification Hierarchique	16
1. Classification Hierarchique sur l'ACP – (Sur 3 années)	18
a. Analyse en composante principale	18
b. Classification Hierarchique	20
IV. Conclusions générales	20

Introduction et motivations

Le financement et le fonctionnement des différents systèmes de santé européens font l'objet d'un débat ardent en ces temps de crise sanitaire. Ce projet vise à utiliser des méthodes d'analyses de données afin de dégager des similitudes et/ou tendances entre ces différents pays. Notre étude a pour objectif d'explorer la richesse d'un certain nombre d'indicateurs afin de savoir si la crise économique de 2008 a impacté durablement la répartition des dépenses de santé des pays européens. Pour ce faire, une première étape sera le traitement des données. En effet, la base de données à notre disposition est issue d'Eurostat, l'Office de Statistique de l'Union Européenne et comporte beaucoup de données manquantes. Dans un second temps, nous procéderons à des études univariées et bivariées de nos variables, afin de comprendre au mieux les données que nous possédons. Enfin, nous pourrions appliquer nos méthodes d'analyse non supervisée sur des données complètes et interprétables, afin de tirer des conclusions à notre problématique.

I. Traitement des données

Notre base de données initiale concentre 30 indicateurs de dépenses de santé, pour 32 pays européens sur une période de 25 années allant de 1992 à 2017. Cependant, cette dernière est composée à 53% de valeurs manquantes. Nous devons donc procéder par tâtonnement pour sélectionner les observations et variables possédant le moins de données manquantes. Pour commencer, nous faisons donc le choix de concentrer notre étude uniquement à partir de 2007.



De plus, la présence de valeurs manquantes pour un certain nombre de pays nous a contraint à retirer des pays de notre étude. Pour ce faire, nous sélectionnons sur la nouvelle période considérée les pays ayant au moins 60% d'information présente. Ces derniers sont présentés ci-dessous :

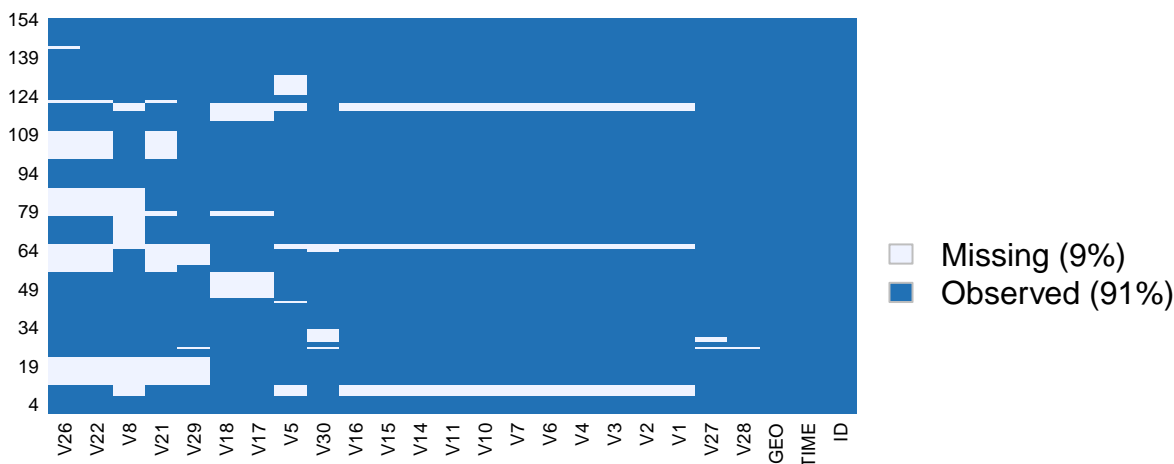
```
## [1] Allemagne Autriche Belgique Chypre Danemark
## [6] Espagne Estonie Finlande France Grece
## [11] Hongrie Irlande Islande Liechtenstein Lituanie
## [16] Luxembourg Malte Norvege Pays-Bas Portugal
## [21] Roumanie Suisse Tchequie
## 30 Levels: Allemagne Autriche Belgique Bulgarie Chypre Croatie ... Tchequie
```

Ces pays restants sont relativement intéressants dans le sens où ils présentent une diversité. En effet, nous avons aussi bien des pays de l'Est, de l'Ouest, du Nord et du Sud de l'Europe. Ce sont également des pays proposant une diversité au niveau du PIB. On retrouve également la France qui est le pays dont les dépenses de santé en pourcentage du PIB sont les plus élevées en Europe. Mais aussi la Grèce, l'un des pays le plus touché par la crise de 2008.

Toujours par soucis de données manquantes, on propose de se concentrer sur les 14 pays suivants : l'Allemagne, l'Autriche, la Belgique, Chypre, l'Espagne, l'Estonie, la Finlande, la France,

la Grece, la Hongrie, la Lituanie, les Pays-Bas, le Portugal et la Roumanie.

Pour finir, nous choisissons de supprimer les variables ayant plus de 35% d'observations manquantes. Ainsi, nous avons réduit à 9% l'information manquante. Ces 9% restants seront remplacés par une régression linéaire par rapport au temps pour chaque pays, si le nombre de valeurs manquantes est inférieur à 4 et si la régression possède un R^2 supérieur à 0.7. On procède à une dernière suppression de variables. Enfin on remplacera les dernières NAs par la médiane annuelle des 14 pays. Il faudra donc prendre en considération les remplacements obtenus dans cette section lorsque nous commenterons les résultats obtenus.



Nous possédons à présent une base de données complète. Dans la section suivante nous pourrions présenter les données que nous avons conservé et étudier le comportement individuel de chaque variable.

II. Études des variables

Notons que nous avons réalisé des études univariée et bivariée avant la suppression et le traitement de données. Cependant, elles ne sont pas essentielles pour le lecteur, c'est la raison pour laquelle nous présenterons ci-dessous seulement une partie de cette analyse sur les variables que nous conservons.

1. Analyses Univariées

Pour avoir un regard d'ensemble sur nos données, nous avons choisi d'observer l'évolution de 2007 à 2017 de la médiane ainsi que de la moyenne des 14 pays considérés.

Tout d'abord la variable V1, représente les dépenses de santé courante en euro par habitant. La moyenne de cet indicateur a une tendance haussière sur les 10 années. Cependant, la médiane révèle une forte chute à partir de 2009, indiquant ainsi que la hausse de la moyenne est portée principalement par certains pays.

Il en est de même, pour la variable V2 qui porte sur les dépenses en euros par habitant en soins curatifs.

Les variables V3 à V6 portent toutes également sur les soins curatifs, respectivement (en milieu hospitalier, en hospitalisation de jour, ambulatoires, et à domicile). Ces dernières ont des tendances à la hausse aussi bien pour la médiane que pour la moyenne.

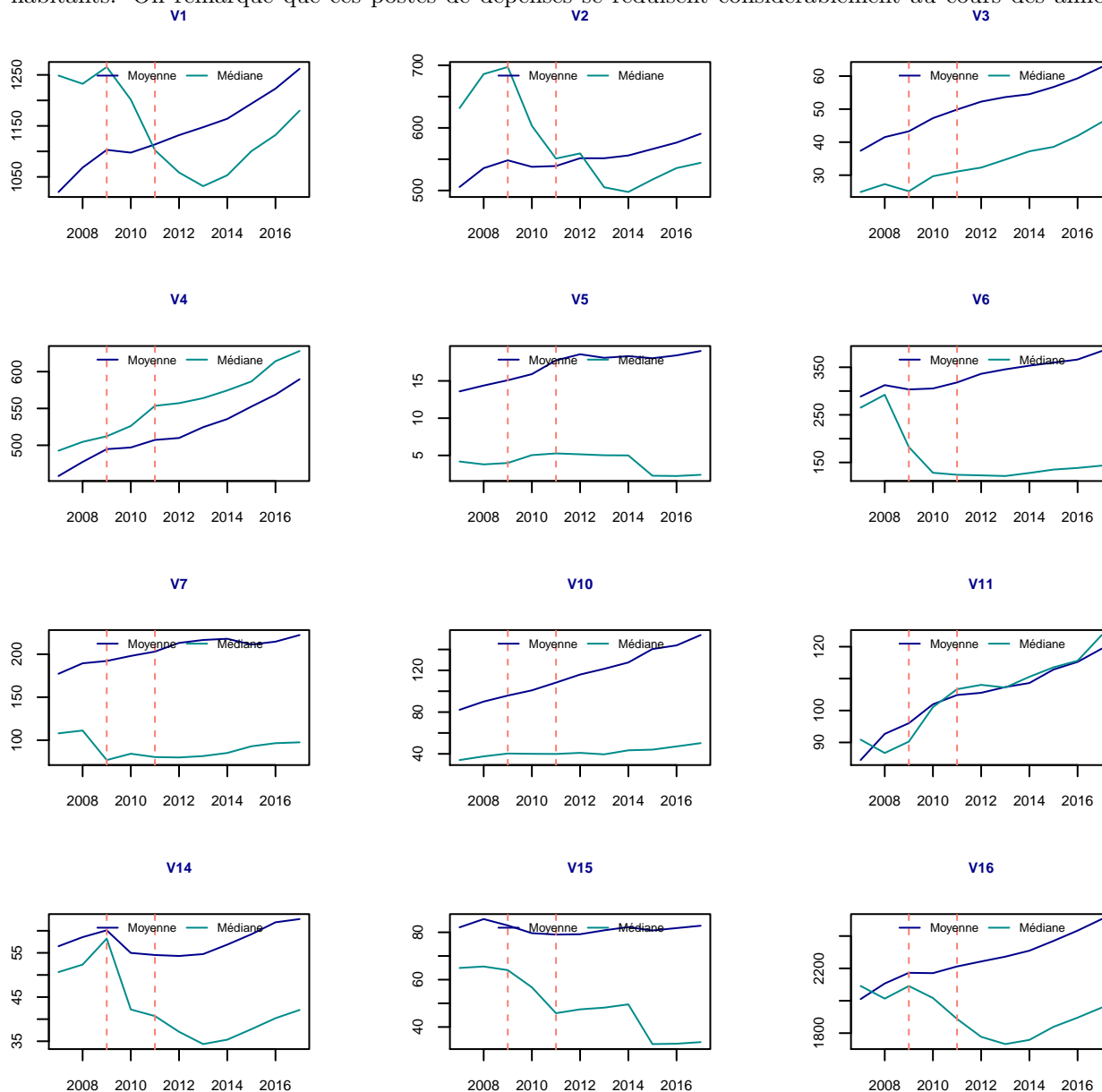
De même, pour les variables V7 à V11 qui portent sur les dépenses en soins de longue durée en euro par habitant. (V10 en ambulatoire, V11 à domicile)

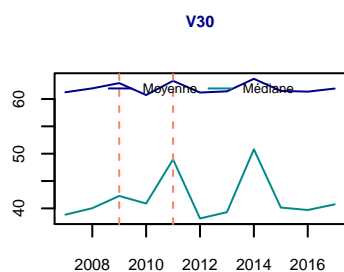
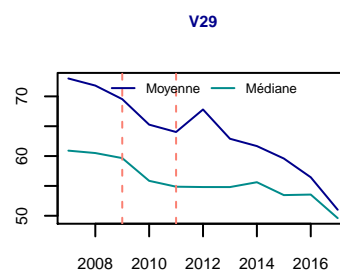
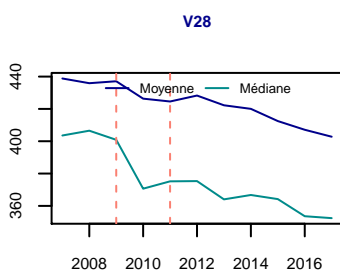
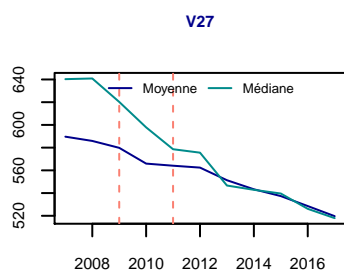
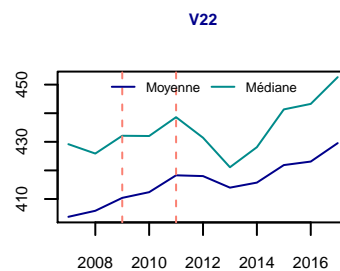
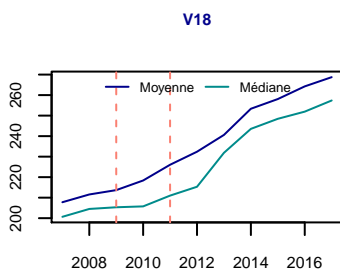
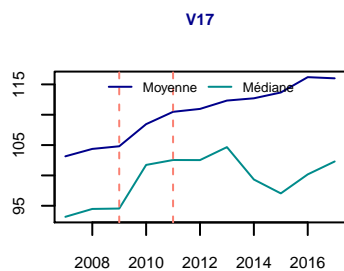
La variable V14 représente les dépenses de médicaments sans ordonnance. On remarque une baisse considérable à partir de 2009 pour la médiane, et une baisse plus faible pour la moyenne. V15 quant à elle porte sur les soins préventifs, et une baisse de la médiane des 14 pays est également constatée. Les dépenses de gouvernance et d'administration des systèmes de santé sont représentées par V16, dont la moyenne présente une tendance à la hausse alors que la médiane baisse fortement à partir de 2009. Le nombre de praticiens en médecine

générale et de praticien en médecine généralisée par 100 000 habitants sont représentés respectivement par V17 et V18. Leur médiane et moyenne ont des tendances haussières sur la période considérée.

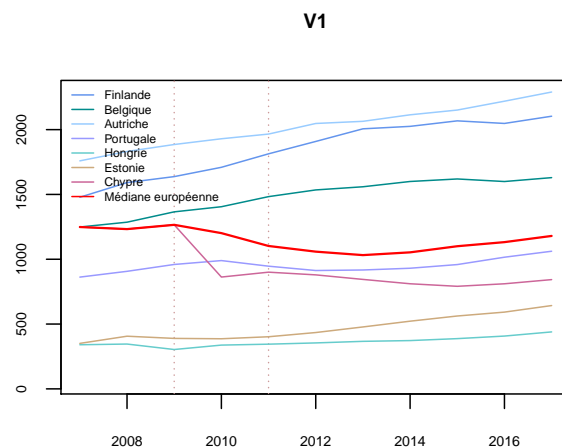
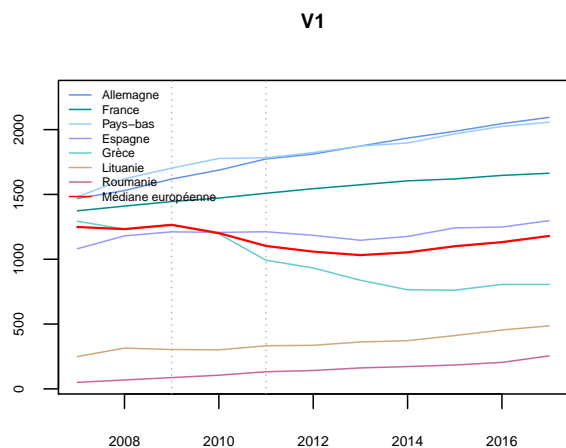
De même, pour V22 qui est le nombre de personnels infirmiers par 100 000 habitants, même si on enregistre une forte baisse à partir de 2011 pour la médiane.

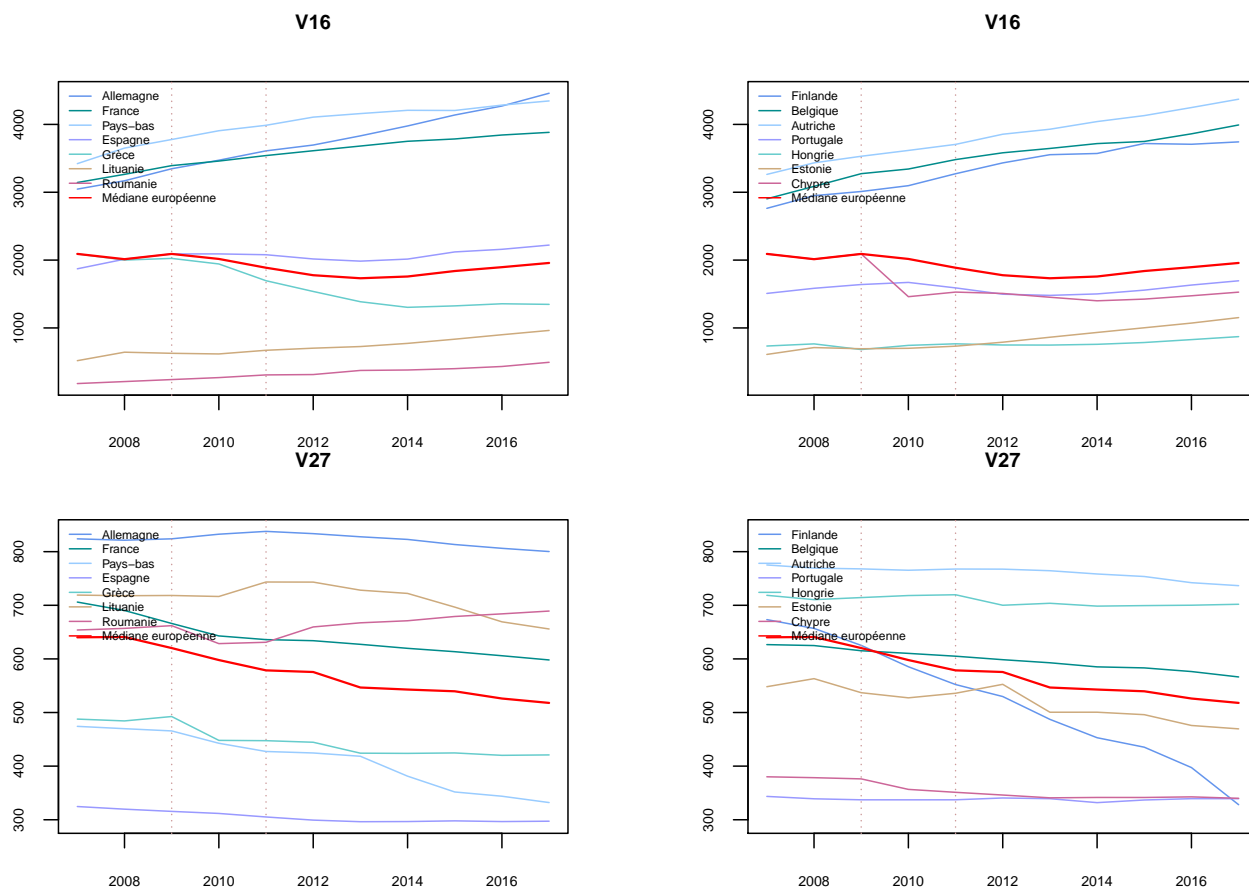
Les quatre dernières variables que nous observons informent sur le nombre de lits d'hôpital par 100 000 habitants. On remarque que ces postes de dépenses se réduisent considérablement au cours des années.





Cette vision d'ensemble étant faite, nous décidons de nous attarder plus en détail sur trois variables, afin de voir l'évolution de chacun des pays dans ces postes de dépense.





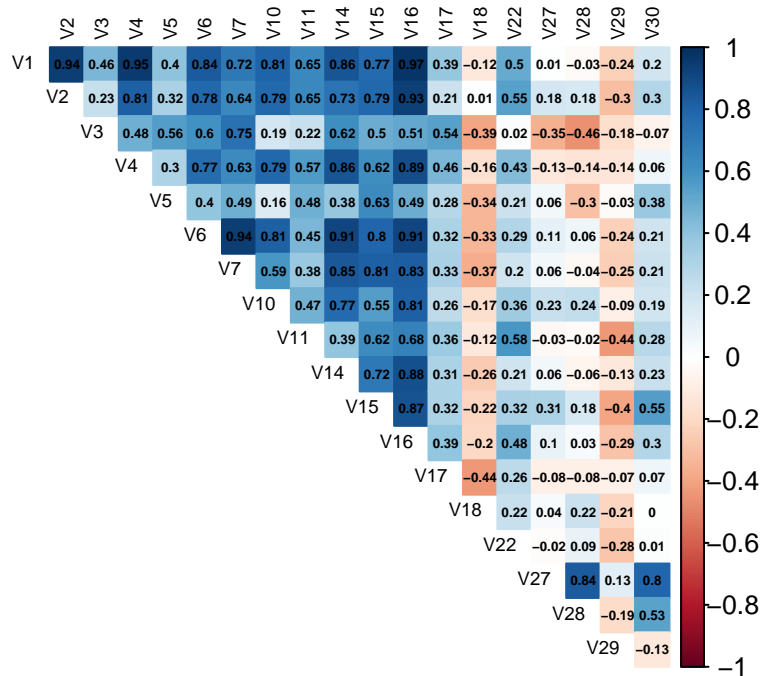
Premièrement pour la variable de dépense courante (V1) où la médiane chutait à partir de 2009, on remarque que cette baisse a été induite principalement par la Grèce et Chypre mais aussi par l'Estonie et le Portugal dans une moindre mesure. Nous pouvons supposer que cette baisse est due à la crise de 2008 car ce sont des pays qui ont été assez impactés. Pour les 10 autres pays la variable V1 présente une certaine constance, voir une légère hausse.

Pour la variable V16, nous avons les mêmes conclusions. La baisse est portée par ces mêmes pays.

Pour finir nous souhaitons analyser la baisse de la moyenne et la médiane de V27. On remarque une baisse très marquée pour la Finlande, dont le nombre de lits a baissé de plus de 50% sur la période considérée. Cette baisse peut surprendre car pour les indicateurs la Finlande fait partie des pays dont les dépenses sont élevées et ont tendance à augmenter. Cependant, on sait notamment que depuis les années 2000 la Finlande a débuté des changements dans ses politiques de dépenses de santé, principalement en réduisant le nombre de lits d'hôpitaux. Néanmoins, la baisse est également portée par la Grèce, les Pays-Bas, la France, la Belgique, et Chypre.

2. Analyse Bivariée

Ici, nous étudierons les liens qui peuvent exister entre les différentes variables de notre modèle afin de savoir celles qui interagissent le plus entre-elles. En effet, une forte corrélation (au sens absolue) entre deux variables pourrait expliquer un lien entre celles-ci dans certains cas. La matrice de corrélation ci-dessous prend en compte toutes les variables sauf ID, TIME et GEO.



Cette matrice de corrélation permet d'affirmer qu'il n'existe pas ou peu de corrélation entre **V27**, **V28**, **V29**, **V30** et toutes les autres variables du modèle. Il s'agit des variables indiquant le nombre de lits d'hôpitaux, et nous avons vu précédemment que leur variation au cours du temps était atypique. Néanmoins elles ne seront pas automatiquement supprimées de notre étude car présentant un aspect particulier. À l'inverse, les variables **V1**, **V2**, **V3**, **V4**, **V5**, **V6**, **V7**, **V10**, **V11**, **V14**, **V15**, **V16**, **V17** et **V22** dans une moindre mesure sont très fortement corrélées entre elles

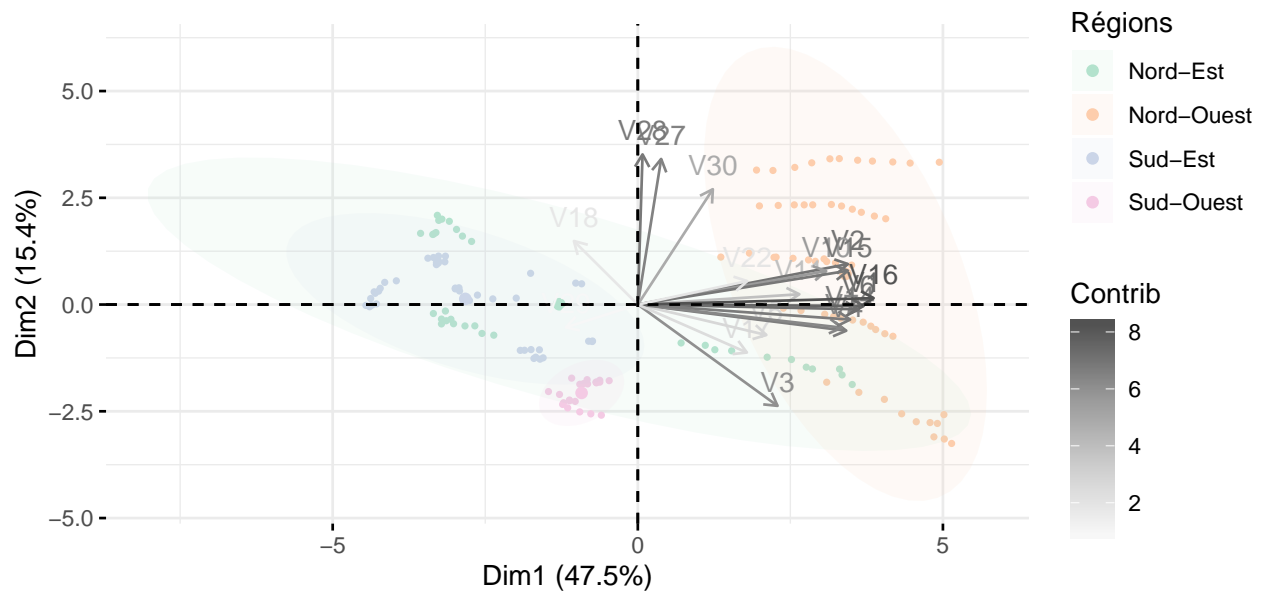
III. Analyse exploratoire

1. Classification Hierarchique sur l'ACP – (Ensemble des individus)

a. Analyse en composante principale

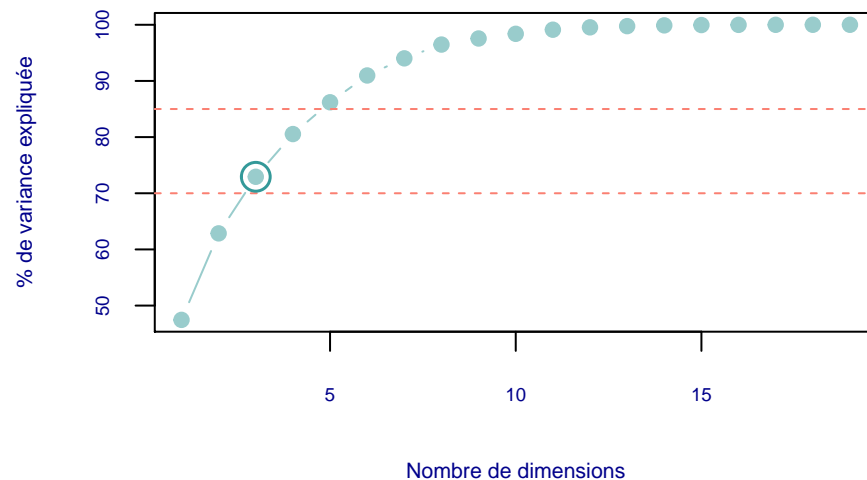
Nous réalisons une première ACP sur nos données après normalisation. Nous décidons de superposer les individus au cercle de corrélations, par soucis de concision. On fait également apparaître les régions Nord, Sud, Ouest, Est, afin de faire apparaître de potentielles similitudes.

Avant de pouvoir interpréter des résultats de cette ACP, une première étape est de définir les dimensions que nous allons conserver.



Choix des dimensions

On affiche le graphique du pourcentage cumulé de variance expliquée en fonction du nombre de dimensions de l'ACP.



Nous constatons que les deux premières dimensions contiennent environ 63% de la dispersion des données. Comme cela est relativement peu et qu'il est compliqué d'analyser nos résultats au delà de 3 dimensions, nous décidons de garder les trois premières, qui expliquent environ 73% de la variance.

Regardons maintenant la contribution des variables et des observations aux axes ainsi que leur cosinus carré. En effet, il est très important de passer par cette étape pour savoir ce que représentent les axes, et une observation mal projetée dans le plan (c'est-à-dire avec un cosinus carré faible) ne sera pas interprétable.

On passe donc de l'étude de 20 variables à celle de 3 composantes principales.

Commentaires des résultats

On constate que les variables contribuant le plus à la dimension 1 sont **V1, V2, V4, V6, V7, V14, V15, V16**. En effet, leur contribution sont toutes supérieure à 8%. De plus, sur le graphique de l'ACP, on remarque également que la première composante principale est corrélée positivement avec toutes ces variables, et ces corrélations sont assez proches de 1. Cet axe est donc déterminé par les dépenses en médicaments, en gouvernance et administration des systèmes de santé, en soins curatifs et en santé courante. Cela signifie que

les observations ayant une forte coordonnée sur cet axe sont les observations ayant le plus dépensé dans ces domaines là.

Celles contribuant le plus à la dimension 2 sont **V3**, **V27**, **V28**, **V30**. Celles-ci sont corrélées positivement avec cette deuxième composante principale, exceptée **V3**. Ainsi, nous en déduisons que sur cet axe, les lits d'hôpitaux disponibles et les dépenses en soins curatifs en milieu hospitaliers sont opposés.

Enfin, celles contribuant le plus à la dimension 3 sont **V18**, **V22**, respectivement à 20% et 18%.

```
##      Dim.1 Dim.2 Dim.3
## V1    9.99  0.00  2.25
## V2    8.64  2.02  3.87
## V3    3.83 12.64  4.67
## V4    8.51  0.84  2.66
## V5    3.24  1.14  7.64
## V6    9.51  0.03  0.59
## V7    8.17  0.66  3.29
## V10   6.95  1.38  0.77
## V11   5.11  0.13  4.12
## V14   8.83  0.28  0.55
## V15   8.64  1.41  1.87
## V16  10.89  0.05  0.30
## V17   2.33  2.83  1.45
## V18   0.80  4.97 19.92
## V22   2.34  0.68 18.05
## V27   0.11 26.12  9.89
## V28   0.00 27.78  0.26
## V29   1.01  0.64  7.50
## V30   1.10 16.39 10.35
```

En ce qui concerne les individus, il est souhaitable que les individus contribuent de manière équivalente à l'inertie de chaque axe. Une contribution excessive est une source d'instabilité des résultats. Dans notre cas, les contributions de chaque individu sont relativement proches. Voici tout de même ci-dessous les données qui contribuent le plus à chaque axe.

```
## [1] "Contribution DIM 1"
## [1] "Pays-Bas2013" "Pays-Bas2014" "Pays-Bas2016" "Allemagne2017"
## [5] "Pays-Bas2017"
## [1] "Contribution DIM 2"
## [1] "Allemagne2007" "Allemagne2008" "Allemagne2009" "Allemagne2010"
## [5] "Allemagne2011" "Allemagne2012" "Allemagne2013" "Pays-Bas2013"
## [9] "Allemagne2014" "Allemagne2015" "Pays-Bas2015" "Allemagne2016"
## [13] "Pays-Bas2016" "Allemagne2017" "Pays-Bas2017"
## [1] "Contribution DIM 3"
## [1] "Roumanie2007" "Roumanie2008" "Roumanie2009" "Roumanie2010" "Roumanie2011"
## [6] "Roumanie2012" "Roumanie2013" "Grece2014" "Roumanie2014" "Grece2015"
## [11] "Roumanie2015" "Grece2016" "Roumanie2016" "Grece2017"
```

Le pourcentage d'inertie expliquée est un critère global et il est le plus souvent complété par un critère local, défini pour chaque individu. Ainsi, il est essentiel d'étudier la qualité de représentation des individus. Si le cosinus d'un individu est grand, en valeur absolue, alors on peut examiner sa position par rapport à un autre individu. Nous calculons la somme des cosinus carré sur les 3 premières dimensions. On remarque que pour la quasi totalité des observations le \cos^2 est élevé. Nous faisons afficher ci-dessous la liste de ceux dont la somme des \cos^2 est supérieure à 0.75.

```

## [1] "somme des cos2 des observations sur les 3 dimensions"

## [1] "Allemagne2007" "Estonie2007" "Hongrie2007" "Lituanie2007"
## [5] "Roumanie2007" "Allemagne2008" "Autriche2008" "Estonie2008"
## [9] "Hongrie2008" "Lituanie2008" "Roumanie2008" "Allemagne2009"
## [13] "Autriche2009" "Estonie2009" "Hongrie2009" "Lituanie2009"
## [17] "Pays-Bas2009" "Roumanie2009" "Allemagne2010" "Autriche2010"
## [21] "Estonie2010" "Hongrie2010" "Lituanie2010" "Pays-Bas2010"
## [25] "Roumanie2010" "Allemagne2011" "Autriche2011" "Espagne2011"
## [29] "Estonie2011" "Grece2011" "Hongrie2011" "Lituanie2011"
## [33] "Pays-Bas2011" "Roumanie2011" "Allemagne2012" "Autriche2012"
## [37] "Espagne2012" "Estonie2012" "Hongrie2012" "Lituanie2012"
## [41] "Pays-Bas2012" "Roumanie2012" "Allemagne2013" "Autriche2013"
## [45] "Espagne2013" "Estonie2013" "Grece2013" "Hongrie2013"
## [49] "Lituanie2013" "Pays-Bas2013" "Roumanie2013" "Allemagne2014"
## [53] "Autriche2014" "Espagne2014" "Estonie2014" "Hongrie2014"
## [57] "Lituanie2014" "Pays-Bas2014" "Roumanie2014" "Allemagne2015"
## [61] "Autriche2015" "Chypre2015" "Espagne2015" "Estonie2015"
## [65] "Hongrie2015" "Lituanie2015" "Pays-Bas2015" "Roumanie2015"
## [69] "Allemagne2016" "Autriche2016" "Espagne2016" "Estonie2016"
## [73] "Hongrie2016" "Lituanie2016" "Pays-Bas2016" "Roumanie2016"
## [77] "Allemagne2017" "Autriche2017" "Espagne2017" "Estonie2017"
## [81] "Hongrie2017" "Lituanie2017" "Pays-Bas2017" "Roumanie2017"

## Dim.1 Dim.2 Dim.3
## V1 0.90 0.00 0.04
## V2 0.78 0.06 0.07
## V3 0.34 0.37 0.09
## V4 0.77 0.02 0.05
## V5 0.29 0.03 0.15
## V6 0.86 0.00 0.01
## V7 0.74 0.02 0.06
## V10 0.63 0.04 0.01
## V11 0.46 0.00 0.08
## V14 0.80 0.01 0.01
## V15 0.78 0.04 0.04
## V16 0.98 0.00 0.01
## V17 0.21 0.08 0.03
## V18 0.07 0.15 0.38
## V22 0.21 0.02 0.35
## V27 0.01 0.76 0.19
## V28 0.00 0.81 0.00
## V29 0.09 0.02 0.14
## V30 0.10 0.48 0.20

```

Les pays tels que l'**Allemagne**, la **Roumanie**, l'**Autriche**, l'**Espagne**, l'**Estonie**, la **Hongrie** et la **Lituanie** ont un bon \cos^2 sur toutes les années, ils sont donc bien représentés dans le plan, nous pourrions donc facilement interpréter leur proximité sur le plan. Les pays tels que les **Pays-Bas** (de 2008 à 2017), la **Grèce** (de 2009 à 2016), la **Belgique** (de 2012 à 2017) et **Chypre** (de 2013 à 2017) ont un bon cosinus carré sur certaines années, nous pourrions donc également facilement interpréter leur résultats sur ces années là. En revanche, les pays comme la **Finlande**, la **France** et le **Portugal** ont un \cos^2 plutôt faible pour toutes les années, il nous sera alors difficile d'interpréter leurs résultats, nous devrions le faire avec précaution.

En ce qui concerne les variables, celles ayant un \cos^2 supérieur à 0,65 sont **V1**, **V2**, **V3**, **V4**, **V6**, **V7**, **V10**, **V14**, **V15**, **V16**, **V27**, **V28** et **V30**. En revanche, la somme des \cos^2 de **V5**, **V17** et **V29** est relativement faible. Nous devrions faire attention lors de l'interprétation de ces dernières variables, car elles sont mal

projetées dans le plan.

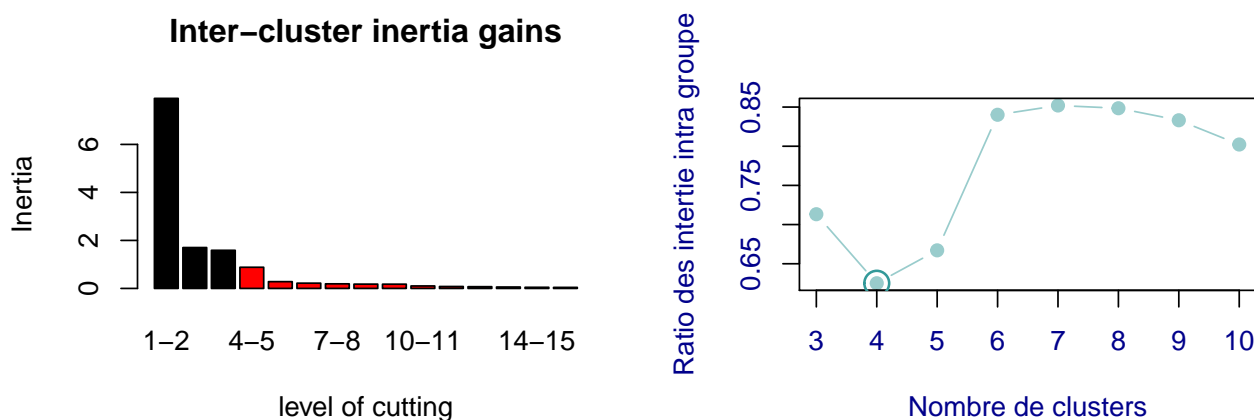
Interprétation: Les premiers résultats que nous pouvons apporter suite à cette ACP sont les suivants. Premièrement, les pays du Nord-Ouest de l'Europe étant représentés à l'extrême droite du premier plan factoriel, ils sont donc caractérisés par des dépenses importantes en gouvernance et administration des systèmes de santé, en soins curatifs et dépenses de santé courante. Alors que les pays du Sud-Est et Sud-Ouest présentent des dépenses plus faibles dans ces secteurs, puisqu'ils sont plutôt à gauche. Pour les pays du Nord-Est on constate une certaine hétérogénéité sur ces mêmes indicateurs. Les pays du Sud-Ouest semblent se caractériser par des dépenses plus importantes en soins curatifs en milieu hospitaliers. Parmi les pays du Nord-Ouest l'Allemagne investit plus que les autres dans les lits d'hôpitaux et notamment de 2017 elle privilégie les postes de dépenses en gouvernance et administration des systèmes de santé, en soins curatifs, lits d'hôpitaux et dépenses de santé courante, il est en de même pour les Pays-Bas. À l'inverse des pays comme la Roumanie et la Grèce ont privilégié le nombre de praticiens en médecine généralisée et le nombre de personnels infirmiers. (On pourrait afficher l'ACP sur la dimension 3 pour mieux voir)

Les régions ne permettent pas de définir des groupes de dépenses précis, car on constate des hétérogénéités dans la manière de dépenser au sein des régions. On propose ainsi de réaliser un clustering hiérarchique sur les résultats de notre ACP pour remarquer des groupes de pays sur les 10 années considérées.

b. Clustering Hiérarchique

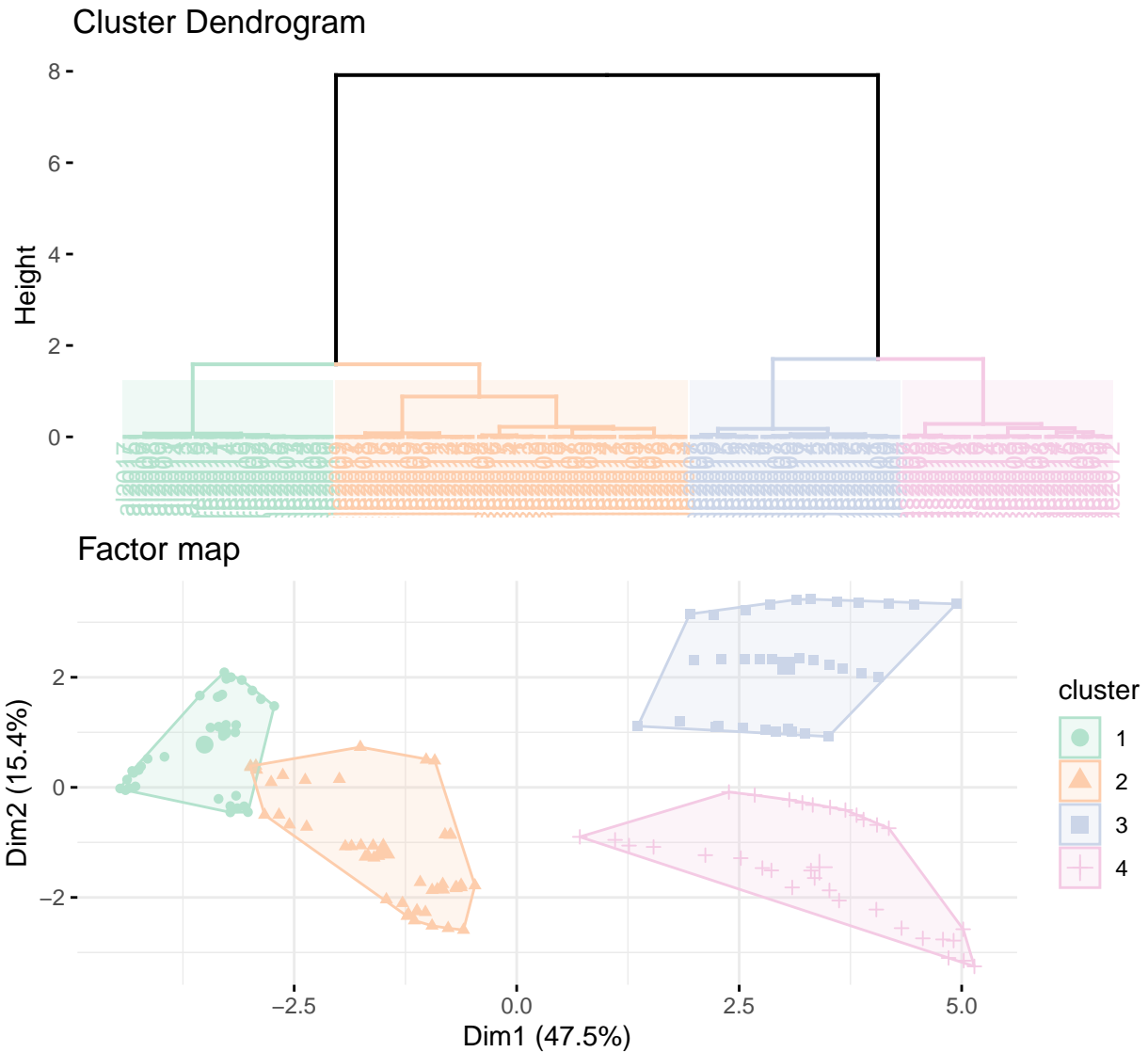
Détection du nombre de cluster - Étude de l'inertie

La classification hiérarchique étant une méthode instable, une première étape est de déterminer le nombre de cluster. Pour se faire, on étudie la perte d'inertie inter-classes, et le gain d'inertie intra-classe. Ces informations sont également apportées par la hauteur des branches du dendrogramme qui représentent la perte d'inertie inter-classe (ou le gain d'inertie intra). Seulement pour rendre la lecture plus facile nous affichons les graphiques suivants. Chaque fois que l'on augmente le nombre de cluster on perd de l'inertie inter-classe et on gagne de l'inertie-intra. L'objectif est de trouver le meilleur compromis. Sur l'histogramme on repère que le saut qui permet de passer de 4 à 5 clusters apporte un gain d'inertie trop faible. On choisit donc de ne pas aller au-delà de 4 clusters. De même, le graphique présente le ratio des inerties intra-groupe (le premier point environ 0.72 est le ratio du gain d'inertie de 3 clusters sur celui de 2 clusters). Il s'agit d'un critère à minimiser qui confirme que nous devons identifier 4 clusters.



Résultats

Nous pouvons à présent afficher le dendrogramme de notre clustering qui sera coupé de manière à définir 4 clusters, ainsi que le graphique de nos clusters.



Nous pouvons à présent mettre en avant les caractéristiques de nos clusters, afin de déterminer quels individus les composent et à quels postes de dépense les associes-t-on.

```
## $`1`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.3  5.792470      1.0938184  9.031151e-16      1.2437054  1.383581
## Dim.2  3.328041      0.7771138  1.844980e-15      0.7864929  1.710879
## Dim.1 -8.557913     -3.5070839 -1.838356e-16      0.5005292  3.002628
##          p.value
## Dim.3 6.935860e-09
## Dim.2 8.745915e-04
## Dim.1 1.149287e-17
##
## $`2`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.1 -4.154477     -1.498665 -1.838356e-16      0.7766367  3.002628
## Dim.2 -5.675444     -1.166557  1.844980e-15      0.9884393  1.710879
## Dim.3 -7.748111     -1.287915  9.031151e-16      0.6978083  1.383581
##          p.value
```

```

## Dim.1 3.260323e-05
## Dim.2 1.383291e-08
## Dim.3 9.326979e-15
##
## $`3`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.2  8.328626          2.2058841  1.844980e-15    0.9230231    1.710879
## Dim.1  6.517600          3.0295579 -1.838356e-16    0.7918735    3.002628
## Dim.3 -2.328938         -0.4988301  9.031151e-16    0.4933558    1.383581
##          p.value
## Dim.2 8.178580e-17
## Dim.1 7.144128e-11
## Dim.3 1.986235e-02
##
## $`4`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.1  7.317405          3.401329 -1.838356e-16    1.1587666    3.002628
## Dim.3  4.885060          1.046320  9.031151e-16    0.9488244    1.383581
## Dim.2 -5.478571         -1.451031  1.844980e-15    0.9767928    1.710879
##          p.value
## Dim.1 2.528119e-13
## Dim.3 1.033975e-06
## Dim.2 4.287743e-08

```

Pour le cluster 1, la moyenne de la dimension 1 est plus basse que la moyenne pour l'ensemble des observations, et inversement pour les dimensions 2 et 3, mais le plus grand écart correspond à celui pour la dimension 1. Donc, vu les résultats obtenus lors de l'ACP, on peut dire que les pays dans ce cluster dépensent moins que les pays des autres clusters en matière de médicaments, gouvernance et administration des systèmes de santé, soins curatifs et de santé courante, et dépensent plutôt en lits d'hôpitaux.

Pour le cluster 2, les moyennes des trois dimensions pour ce cluster sont plus faibles que les moyennes de ces trois dimensions pour l'ensemble des observations. Donc, vu les résultats obtenus lors de l'ACP, on peut dire que les observations dans ce cluster dépensent moins que les observations des autres clusters (mais plus que ceux du cluster 1) en matière de médicaments, gouvernance et administration des systèmes de santé et de santé courante, et dépensent plutôt en terme de soins curatifs en milieu hospitalier.

Pour le cluster 3, la moyenne des dimensions 1 et 2 pour les observations de ce cluster sont plus élevées que les moyennes pour toutes les observations (et l'écart est assez conséquent pour la dimension 1). En revanche, c'est l'inverse pour la dimension 3. Par rapport aux résultats obtenus lors de l'ACP, les dépenses pour les observations de ce cluster portent plutôt sur les médicaments, la gouvernance et administration des systèmes de santé, les soins curatifs, les lits en hôpitaux et sur la santé courante.

Pour le cluster 4, les moyennes des dimensions 1 et 3 sont plus élevées pour ce cluster que pour l'ensemble des observations, surtout pour la première dimension, et c'est l'inverse pour la dimension 2. A la vue des résultats obtenus lors de l'ACP, les dépenses pour les observations du cluster 4 sont portées plutôt sur les médicaments, la gouvernance et administration des systèmes de santé, les soins curatifs et sur la santé courante.

```

## Cluster: 1
## Hongrie2017 Hongrie2015 Hongrie2014 Hongrie2016 Hongrie2013
## 0.5783524 0.5874125 0.5916502 0.6121147 0.6896584
## -----
## Cluster: 2
## Chypre2011 Chypre2010 Chypre2012 Chypre2013 Chypre2014
## 0.1699513 0.1862745 0.2061853 0.3363432 0.6480091
## -----
## Cluster: 3

```

```
## Autriche2011 Autriche2012 Autriche2010 Autriche2013 Autriche2009
##      0.3244902      0.3360453      0.4095520      0.4280307      0.5216872
## -----
## Cluster: 4
## Pays-Bas2007 Pays-Bas2008 Pays-Bas2009 Finlande2012 France2016
##      0.5217773      0.6882894      1.0313549      1.2690716      1.2743682
```

On constate que le cluster 1 est surtout représenté par la Hongrie qui le plus proche du centre du cluster. Pour le cluster 2 Chypre est le pays le mieux représenté. L'Autriche est le pays le plus proche du centre du cluster 3. Et enfin, le cluster 4 est surtout représenté par les Pays-Bas, la Finlande en 2012 et la France en 2016.

Conclusions

Comme nous l'avons déduit lors de l'ACP, les régions ne permettent pas de définir des groupes de comportement. Comme on voit ci-dessous le cluster 1 possède à la fois des pays du Nord-Est et du Sud-Est. Les pays qui le composent sont la **Roumanie**, la **Lituanie**, la **Hongrie**, et l'**Estonie** (seulement jusqu'en 2013). Ces pays sont donc caractérisés par des dépenses qui sont en moyenne plus faibles que les autres pays en matière de médicaments, gouvernance et administration des systèmes de santé, soins curatifs et de santé courante. À l'inverse, en moyenne ils possèdent plus de lits d'hôpitaux.

Le cluster 2 possède autant de pays du Sud-Est que de pays du Sud-Ouest, et un pays du Nord-Est. Les pays qui le composent sont l'**Espagne**, la **Grèce**, **Chypre**, le **Portugal** et l'**Estonie** (à partir de 2014). Ces derniers dépensent en moyenne moins que les observations des autres clusters (mais plus que ceux du cluster 1) en matière de médicaments, gouvernance et administration des systèmes de santé et de santé courante. Cependant, en terme de soins curatifs en milieu hospitalier leur dépense est plus élevée.

Le cluster 3 est composé uniquement de pays du Nord-Ouest, qui sont : l'**Allemagne**, l'**Autriche** et la **Belgique**. Ces derniers sont caractérisés par des dépenses importantes dans les médicaments, la gouvernance et administration des systèmes de santé, les soins curatifs, les lits en hôpitaux et sur la santé courante.

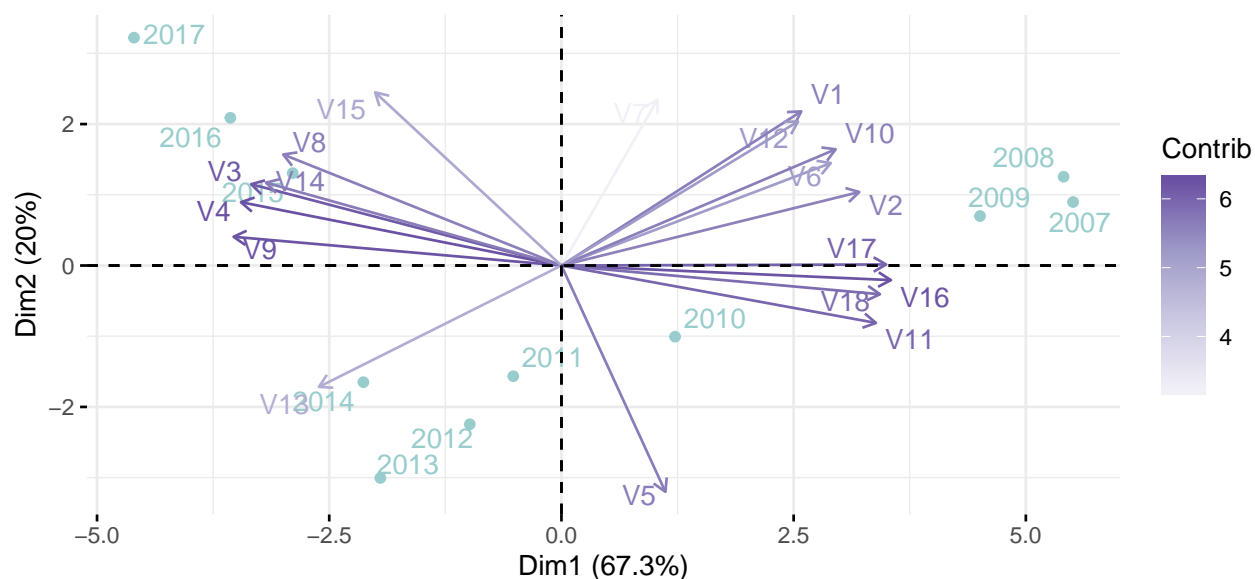
Et enfin le cluster 4 est composé de la **Finlande**, la **France** et la **Pays-Bas**. Dont les dépenses sont en moyenne plus élevées pour les médicaments, la gouvernance et administration des systèmes de santé, les soins curatifs et sur la santé courante.

```
##
##           1  2  3  4
## Nord-Est  18  4  0 11
## Nord-Ouest 0  0 33 22
## Sud-Est   22 22  0  0
## Sud-Ouest  0 22  0  0
```

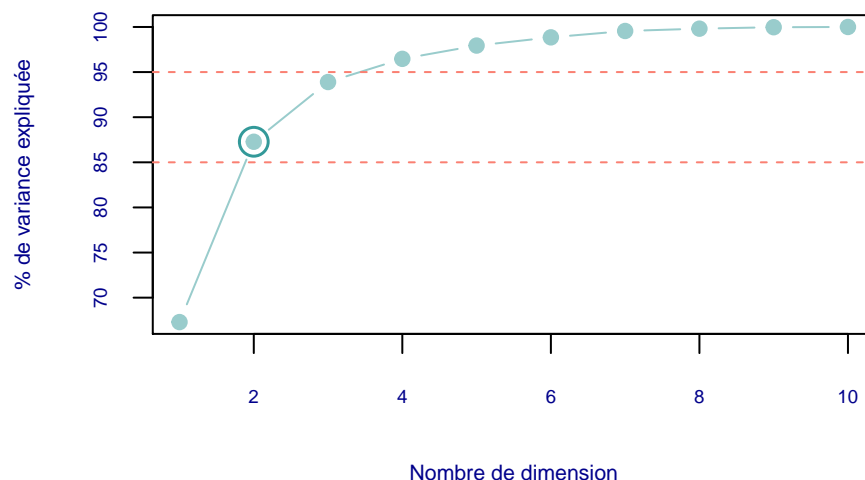
Suite à cette étude sur la totalité de notre base de données, nous nous rendons compte qu'il existe des différences notables en termes de dépenses par les pays au cours du temps. Nous décidons donc d'effectuer une analyse sur les dépenses médianes européennes au cours de la période 2007-2017

1. Classification Hierarchique sur l'ACP – (Temporalité)

a. Analyse en composante principale



Choix des dimensions



On constate que la première composante principale contient 67,3% de la dispersion des données, et la deuxième composante principale en contient 20%. Ainsi, Les deux premières composantes principales expliquent 87,3% de la dispersion totale, ce qui est suffisant pour ne pas altérer la qualité de représentation des données. Nous décidons donc de garder les deux premières composantes principales.

Comme expliqué précédemment pour l'ACP sur l'ensemble des données, il est essentiel d'étudier la qualité de représentation des individus et des variables. Pour cela, nous affichons ci-dessous les résultats des contributions des individus et des variables aux axes ainsi que leurs cosinus carrés dans le plan.

Résultats

Les variables contribuant le plus à l'axe 1 sont V9, V16, V17, V18, V4, V11, V3. En effet, leur contribution sont toutes supérieures à 7%. De plus, sur le graphique de l'ACP, on constate que V16, V17, V18 et V11 sont corrélées positivement avec cet axe et V9, V4 et V3 négativement. Ainsi, on en déduit que sur cet axe, les dépenses en matière de praticiens en médecine et de soins à domicile sont opposés aux dépenses en matière de soins curatifs et aux soins de longue durée en milieu hospitalier. Donc, les années représentées avec une forte coordonnée sur l'axe 1 seront les années où en moyenne les 14 pays sélectionnés ont plutôt dépensé en

matière de gouvernance et administration des systèmes de santé et en matière de praticiens en médecine, alors que les années ayant une faible coordonnée sur l'axe 1 seront les années où les 14 pays sélectionnés ont en moyenne plutôt dépensé en matière de soins curatifs et aux soins de longue durée en milieu hospitalier.

Les variables contribuant le plus à l'axe 2 sont V5, V15, V1 et V12 (elles ont une contribution supérieure à 7%). Sur le graphique de l'ACP, on constate que V1, V15 et V12 ont une corrélation positive avec cet axe, alors que V5 et l'axe 2 sont corrélés négativement. On en conclut donc que sur cet axe, les dépenses en matière de soins curatifs ambulatoires sont opposées aux dépenses en matière de soins préventifs et de services auxiliaires. Toutes les variables présentent un cosinus carré supérieur à 0,75, excepté V7 dont le cosinus carré vaut 0,51. Donc globalement, les variables présentent toutes une très bonne qualité de représentation sur le plan, nous pouvons donc facilement interpréter leurs résultats sur le plan.

Nous savons qu'il est souhaitable que les individus contribuent de manière équivalente à l'inertie de chaque axe. Voici ci-dessus les résultats de contributions de chaque année aux axes. Les cosinus carrés des années sont globalement tous élevés (supérieurs à 0,7), il sera alors facile d'interpréter leurs résultats dans le plan, excepté pour 2010 et 2011, dont les cosinus carrés valent respectivement 0,42 et 0,55 dans le plan.

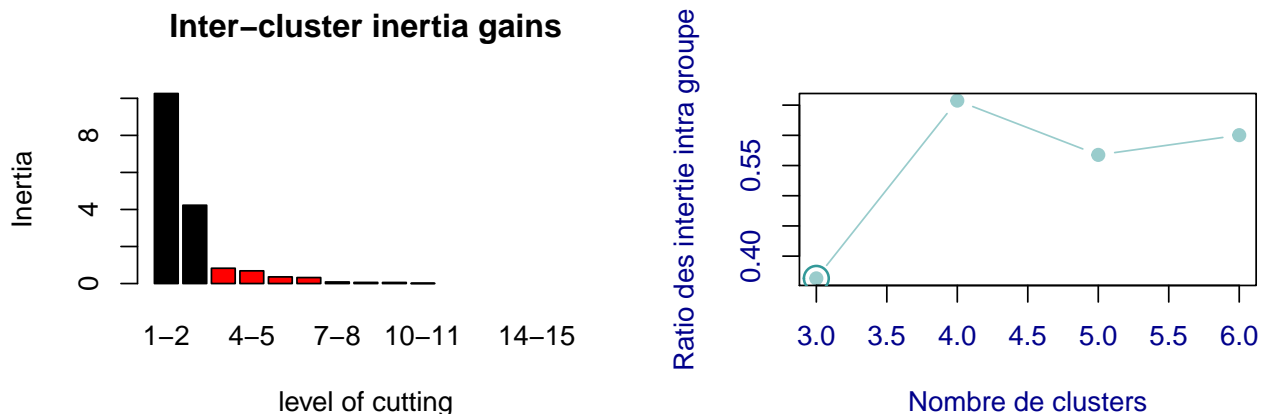
D'après le graphique de l'ACP et les résultats des contributions et des cosinus carrés, nous pouvons penser que les années 2007 à 2009, pour la moyenne des 14 pays européens sélectionnés, sont caractérisées par des dépenses en matière de gouvernance et administration des systèmes de santé et de praticiens en médecine, contrairement aux années 2015 à 2017 qui sont plutôt caractérisées par des dépenses en matière de soins curatifs et de soins de longue durée en milieu hospitalier, de soins préventifs et de services auxiliaires. Les années 2010 à 2014 sont quant à elles représentées par des dépenses en matière de soins curatifs ambulatoires.

On remarque donc 3 groupes d'années : celui pendant la crise (2007 à 2009), celui juste après crise (2010 à 2014), et les années 2015 à 2017. Pour s'assurer de bien diviser correctement nos observations, nous pouvons passer au clustering hiérarchique sur les résultats de cette ACP afin d'analyser les inertie inter et intra groupe.

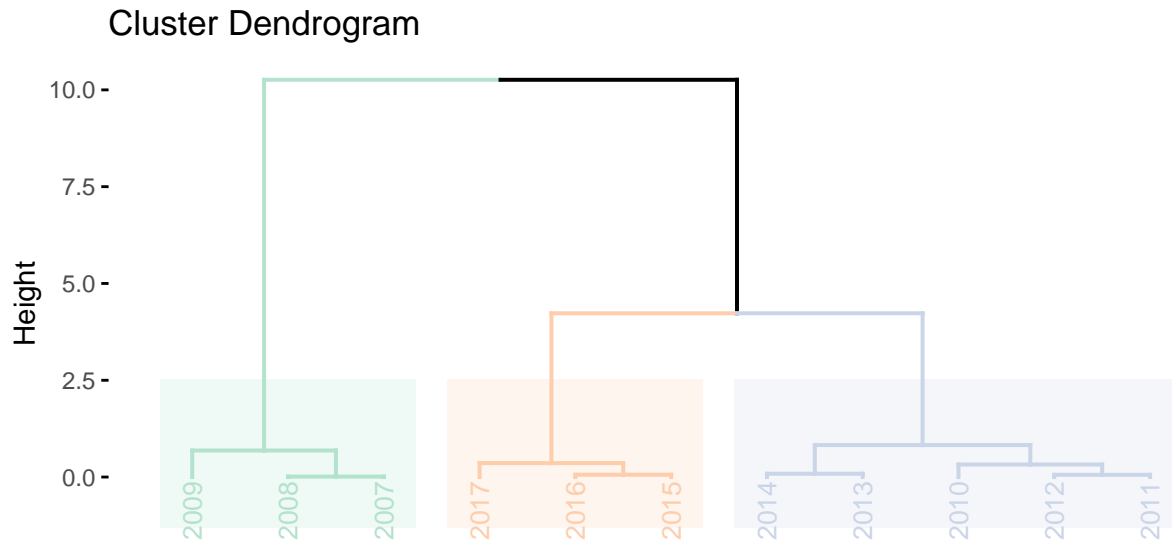
b. Classification Hierarchique

Détection du nombre de cluster - Inertie

Pour les même raison que précédemment, nous identifions 3 clusters.



Résultats



Sur le graphiques ci-dessus, on peut distinguer 3 clusters différents, c'est bien le nombre de clusters qui maximise l'inertie inter groupe et minimise l'inertie intra groupe. Regardons maintenant par quelles composantes principales sont caractérisé chaque cluster.

```
## $`1`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.2  2.252110          2.205816 -8.679925e-16    0.7862057    1.896686
## Dim.1 -2.051342          -3.686740  1.019387e-15    0.7030279    3.480328
##          p.value
## Dim.2 0.02431533
## Dim.1 0.04023367
##
## $`2`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.2 -2.882755          -1.89406 -8.679925e-16    0.6798365    1.896686
##          p.value
## Dim.2 0.003942142
##
## $`3`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.1  2.860462          5.140919  1.019387e-15    0.450184    3.480328
##          p.value
## Dim.1 0.004230239
```

Conclusions :

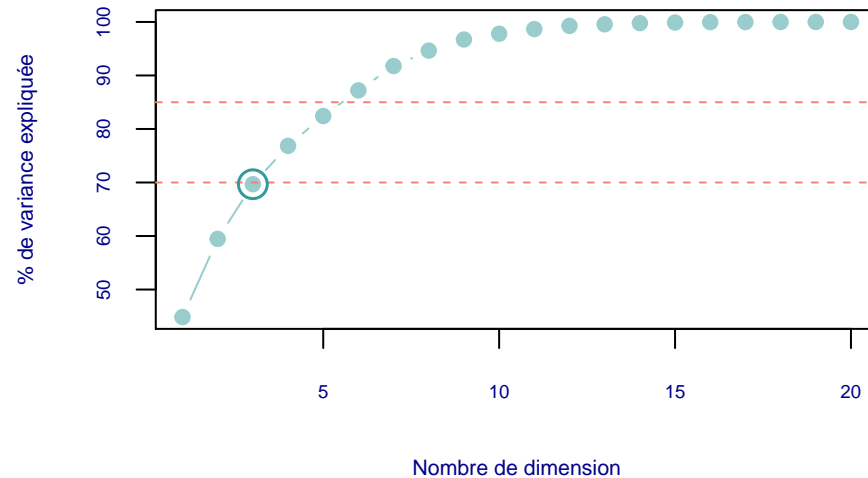
- Dans le **cluster 1**, la moyenne de la dimension 1 est supérieure à celles des autres clusters, et inversement pour la dimension 2. Donc, vu les résultats obtenus lors de l'ACP, on peut dire que les années de ce cluster sont caractérisées par des dépenses en matières de soins curatifs et de longue durée en milieu hospitalier, de soins préventifs et de services auxiliaires.
- Le **cluster 2** quant à lui est marqué par une moyenne de la dimension 2 plus faible que celles des autres. Par conséquent, les années de ce cluster correspondent à celles où les pays Européens dépensaient plutôt pour des soins ambulatoires.
- La moyenne de la dimensions 1 dans le **cluster 3** est beaucoup plus élevée que celle des autres groupes d'années. On en déduit donc que les dépenses se concentraient en priorité sur la gouvernance et l'administration des systèmes de santé ainsi que les praticiens en médecine.

La présence de ces différents clusters représentant des années particulières met en exerce les conséquences

que la crise a pu avoir sur l'économie des différents pays européens sur la période 2007 - 2017. Nous décidons donc d'effectuer une analyse sur 3 années en particulier(2007 - 2012 - 2017) en prenant en compte la totalité des pays et des variables qui nous avons au sein de notre base de données.

1. Classification Hierarchique sur l'ACP – (Sur 3 années)

a. Analyse en composante principale



Suite à l'analyse du graphique ci-contre, nous décidons de garder les trois premières dimensions, qui expliquent environ 70% de la variance. Une représentation des pays Européens durant les années 2007, 2012 & 2017 est donc justifiée et la qualité de représentation ne sera pratiquement pas altérée par la réduction de la dimension de vingt(20) à trois(3).

Analysons à présent les contributions des différentes variables et individus aux axes ainsi que leur cosinus carré. En effet, il est très important de passer par cette étape pour savoir ce que représentent les axes, et une observation mal projetée dans le plan (c'est-à-dire avec un cosinus carré faible) ne sera pas interprétable.

```
##      Dim.1 Dim.2 Dim.3
## TIME  0.03  0.02  0.21
## V1    0.91  0.00  0.03
## V2    0.78  0.06  0.06
## V3    0.34  0.35  0.08
## V4    0.78  0.03  0.03
## V5    0.28  0.03  0.14
## V6    0.85  0.00  0.01
## V7    0.71  0.01  0.07
## V10   0.63  0.03  0.02
## V11   0.47  0.00  0.07
## V14   0.79  0.00  0.02
## V15   0.76  0.05  0.04
## V16   0.98  0.00  0.00
## V17   0.21  0.09  0.02
## V18   0.05  0.09  0.53
## V22   0.21  0.01  0.27
## V27   0.00  0.80  0.14
## V28   0.00  0.81  0.00
## V29   0.09  0.01  0.20
## V30   0.08  0.53  0.10

## [1] "somme des cos2 des observations sur les 3 dimensions"
```

```
## [1] "Allemagne2007" "Estonie2007" "Hongrie2007" "Lituanie2007"
## [5] "Pays-Bas2007" "Roumanie2007" "Allemagne2012" "Autriche2012"
## [9] "Espagne2012" "Estonie2012" "Grece2012" "Hongrie2012"
## [13] "Lituanie2012" "Pays-Bas2012" "Roumanie2012" "Allemagne2017"
## [17] "Autriche2017" "Belgique2017" "Chypre2017" "Espagne2017"
## [21] "Estonie2017" "Grece2017" "Hongrie2017" "Lituanie2017"
## [25] "Pays-Bas2017" "Roumanie2017"
```

Contribution des variables :

On constate que les variables contribuant le plus à la **dimension 1** sont V1, V2, V4, V6, V7, V14, V15, V16. Sur le graphique de l'ACP, on constate également que la première composante principale est corrélée positivement avec toutes ces variables, et ces corrélations assez proches de 1. Cet axe est donc déterminé par les dépenses en médicaments, en gouvernance et administration des systèmes de santé, en soins curatifs et en santé courante.

Les variables contribuant le plus à la **dimension 2** sont V3, V27, V28, V30. Celles-ci sont corrélées positivement avec cette deuxième composante principale, exceptée V3. Elle permettra donc de mettre en opposition les quantités de lits d'hôpitaux aux dépenses en matières de soins curatifs en milieu hospitalier.

Les variables qui contribuent le plus à la dimension 3 sont V18, V22, V5 et V29.

Contribution des individus :

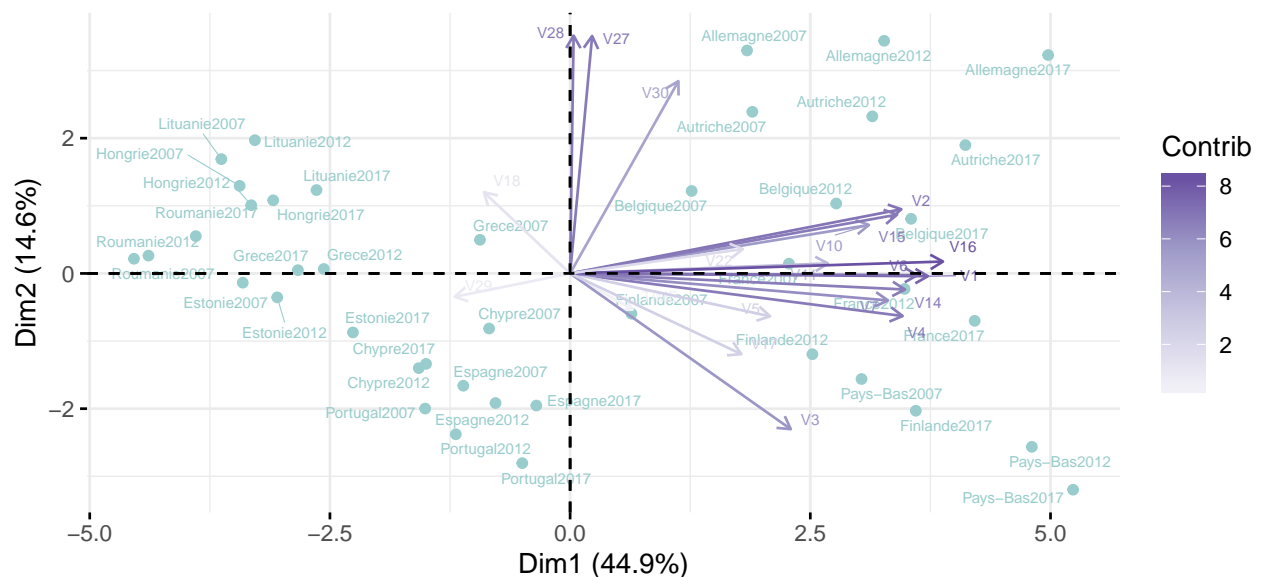
Les pays tels que l'Allemagne, la Roumanie, l'Autriche, l'Espagne, l'Estonie, la Hongrie, la Lituanie et les Pays-Bas ont un bon \cos^2 sur toutes les années, ils sont donc bien représentés dans le plan, leur proximité seront donc plus aisément interprétable.

En revanche, les pays restant n'ont pas un bon \cos^2 pour ces années, il nous sera alors difficile d'interpréter leurs résultats et le faire avec précaution.

Concernant les variables, celles ayant un $\cos^2 > 0,65$ sont V1, V2, V3, V4, V6, V7, V10, V14, V15, V16, V18, V27, V28 et V30. Nous devrons faire attention lors de l'interprétation de ces variables ayant un $\cos^2 < 0.5$ car elles sont mal projetées dans le plan.

Interprétation :

Le graphique confirme nos interprétations faites ci-dessus :



L'axe 1 isole les dépenses en médicaments, en gouvernance et administration des systèmes de santé, en soins curatifs et en santé courante. En réalité, il oppose l'Estonie, la Hongrie, la Roumanie et la Lituanie prise

sur les années 2007, 2012 et 2017 aux **Pays-Bas**, à l'**Allemagne** et à l'**Autriche** (toutes prises uniquement sur les années 2012 et 2017) marqués par de forte dépenses dans ces agrégats.

Sur l'**axe 2**, on constate que l'**Allemagne** et l'**Autriche** caractérisés par de grandes quantités de lits d'hopitaux sont en opposition aux **Pays-Bas**, au **Portugal** ainsi qu'à l'**Espagne** marqués par des dépenses en termes de soins curatifs.

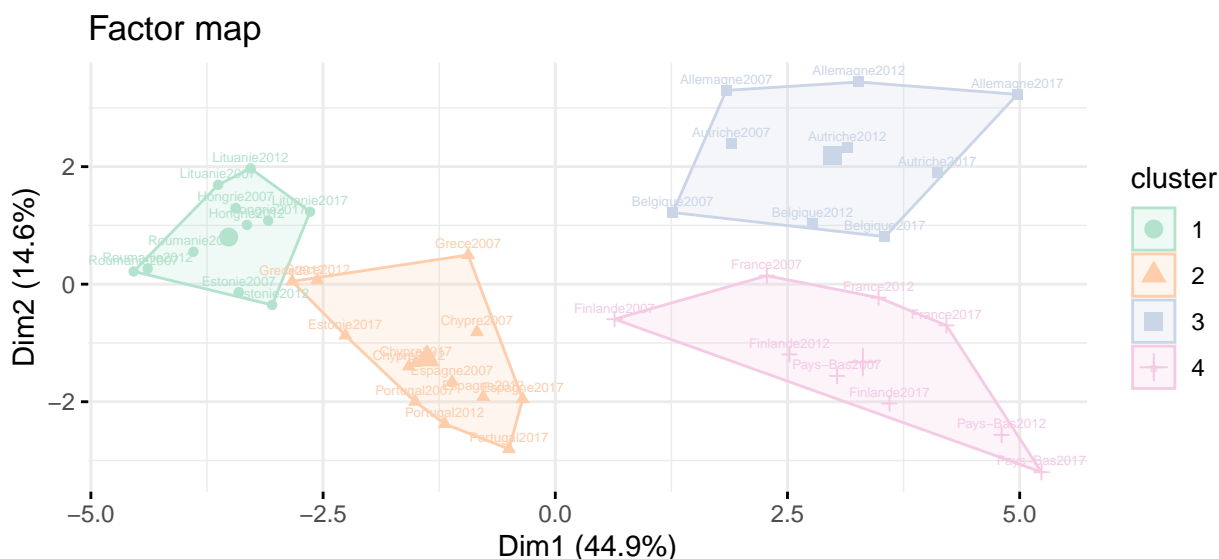
Néanmoins, il existe des groupes se détachant de par leur type de dépenses :

- **Finlande, France** : caractérisé par beaucoup de dépenses en matières de santé mais sans différenciation particulière dans le choix des agrégats.
- **Belgique** : beaucoup de dépenses en matières de santé avec une légère préférence pour les soins en tout genre.

Le regroupement ci-dessous nous permet de constater qu'en général:

- Les pays qui en 2007 dépensaient dans les médicaments, les soins curatifs, la santé courante et la gouvernance et administration des systèmes de santé soins curatifs et en santé courante n'ont cessé d'augmenter leur moyens dans ces secteurs.
- L'**Estonie** à compter de 2012 tend à changer de groupe du fait de l'augmentation de ses dépenses en matières de santé en général et particulièrement celles en soins curatifs.
- La **Grèce** qui en 2007 était au centre des pays a considérablement baissé ses dépenses de santé au cours des années qui ont suivi la crise financière de 2008.

b. Classification Hierarchique



IV. Conclusions générales

Tout d'abord, contrairement à ce que l'on pourrait croire les différentes régions Nord, Sud, Est et Ouest ne permettent de définir des "groupes" de comportements en matières de dépenses de santé. Ensuite l'étude de la médiane européenne entre les années 2007 et 2017, nous amène à la conclusion qu'il existe 3 grandes étapes dans son évolution au cours de cette période. En effet, les dépenses en matières de santé des années autour de la crise (2007-2009) sont différentes de celles qui suivent la crise. De plus, les indicateurs principaux de dépenses durant les années où l'Europe se remettait de la crise (2010-2014) sont différents de ceux à partir de 2015. A la question de savoir si la crise économique de 2008 a impacté durablement la répartition des dépenses au sein de l'Union Européenne, la réponse est par conséquent oui. Cependant il existe un biais dans

les résultats de notre étude car ceux-ci se basent uniquement sur les données de 14 pays qui n'étaient pas complètes.