

Projet-UE

AYMAN YAYA

2023-09-27

Introduction

L'étude "Les réalités sociales françaises à l'aune européenne", publiée par le Centre d'Analyse Stratégique en octobre 2007, avait pour objectif de comparer les réalités sociales françaises avec celles des autres pays de l'Union Européenne. L'étude mettait en évidence des similitudes et des différences entre la France et les autres pays de l'Union Européenne en termes d'indicateurs sociaux tels que le taux de chômage, la pauvreté, l'éducation, le bénévolat et le PIB. Pour cela, il est possible de mener cette étude à l'aide d'une analyse bivariée suivie de l'analyse en Composantes Principales (ACP) et terminer par la Classification Ascendante Hiérarchique (CAH) pour comparer les différentes réalités sociales de la France avec celles des autres pays de l'Union Européenne. L'objectif est de mesurer la corrélation entre les variables explicatives et de déterminer si elles sont liées ou non et ainsi réaliser une réduction de variable et un regroupement des pays selon leurs caractéristiques.

Présentation des données

On s'intéresse ici à une partie de ces données, qui sont disponibles pour tous les pays de l'Europe à 25. Les variables considérées sont:

Table 1: Tableau des variables

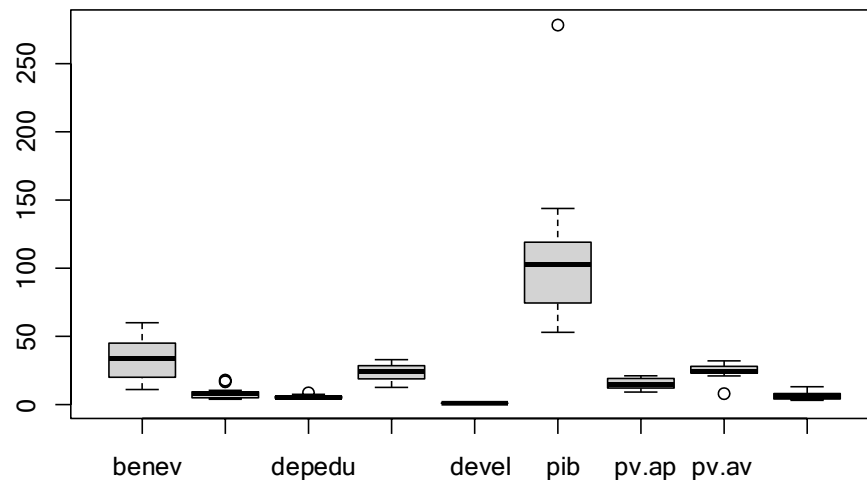
Variables	Description
benev	part des personnes exerçant une activité bénévole dans une association
chom	taux de chômage des personnes entre 15 et 74 ans
depedu	dépenses publiques d'éducation en pourcentage du PIB, tous niveaux confondus
depsoc	dépenses de protection sociale en pourcentage du PIB
devel	indice de développement humain par pays
pib	Produit Intérieur Brut par habitant
pv.ap	taux de pauvreté après transferts sociaux autre que pensions de vieillesse et de survie
pv.av	taux de pauvreté avant transferts sociaux autre que pensions de vieillesse et de survie
trv.pv	taux de pauvreté des travailleurs à temps complet

Table 2: Tableau des Pays

Abbréviation	Nom complet	Abbréviation	Nom complet
at	Autriche	it	Italie
be	Belgique	lt	Lituanie
cy	Chypre	lu	Luxembourg
cz	République Tchèque	lv	Lettonie
de	Allemagne	mt	Malte
dk	Danemark	nl	Pays-Bas
ee	Estonie	pl	Pologne
es	Espagne	pt	Portugal
fi	Finlande	se	Suède
fr	France	si	Slovénie
gr	Grèce	sk	Slovaquie
hu	Hongrie	uk	Royaume Uni
ie	Irlande		

Vous pouvez retrouver dans le tableau ci-dessus, les abréviations des pays ainsi que leur nom complet.

Statistiques descriptives univariées et bivariées

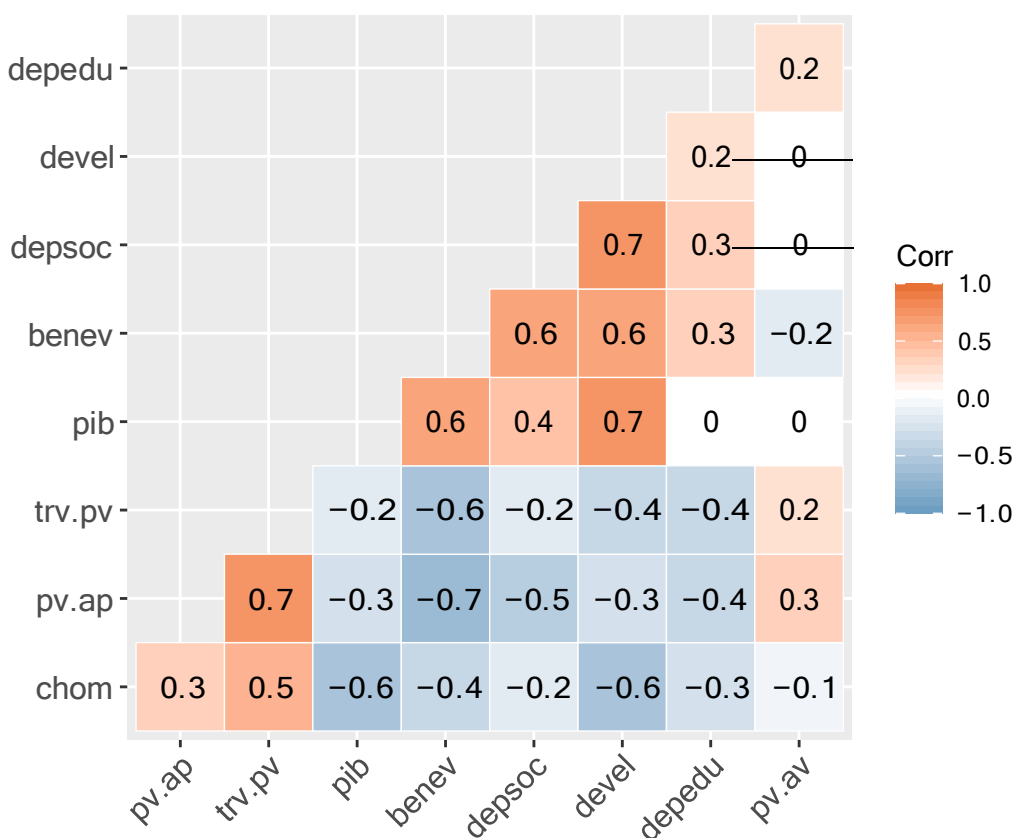


On observe une forte variabilité des personnes exerçant une activité bénévole dans une association (benev) ($\text{var}=210.6767$) et du PIB ($\text{var}=2050.062$) au sein des pays membres de l'union européenne. Notamment, un pays se distingue fortement des autres pays (Luxembourg = Pays riche) au niveau du PIB (valeur aberrante). Les pays sont assez similaires par rapport aux autres variables par conséquent, on peut dire que les valeurs suivantes représentent l'état général des pays de l'UE :

Table 3: Tableau de certaines moyennes

Variables	Moyennes
chom	8.0320
depedu	5.3648
depsoc	23.1400
devel	0.9126
pv.ap	5.4000
pv.av	24.8800
trv.pv	6.6800

Corrélations entre variables:



On observe ici une corrélation positive entre les variables benev, depsoc, pib et devel (en moyenne 0.7) et aussi entre pv.ap et trv.pv (0.7). Ainsi, le niveau de dépenses sociales et la contribution du bénévolat sont très liés au niveau de développement humain et économique des pays de l'UE.

De plus, on a aussi une corrélation négative entre chom et devel/pib (-0.6). L'évolution d'un pays d'un point de vue économique et sociale favorise l'opportunité d'emploi qui contribue à l'augmentation de la population active.

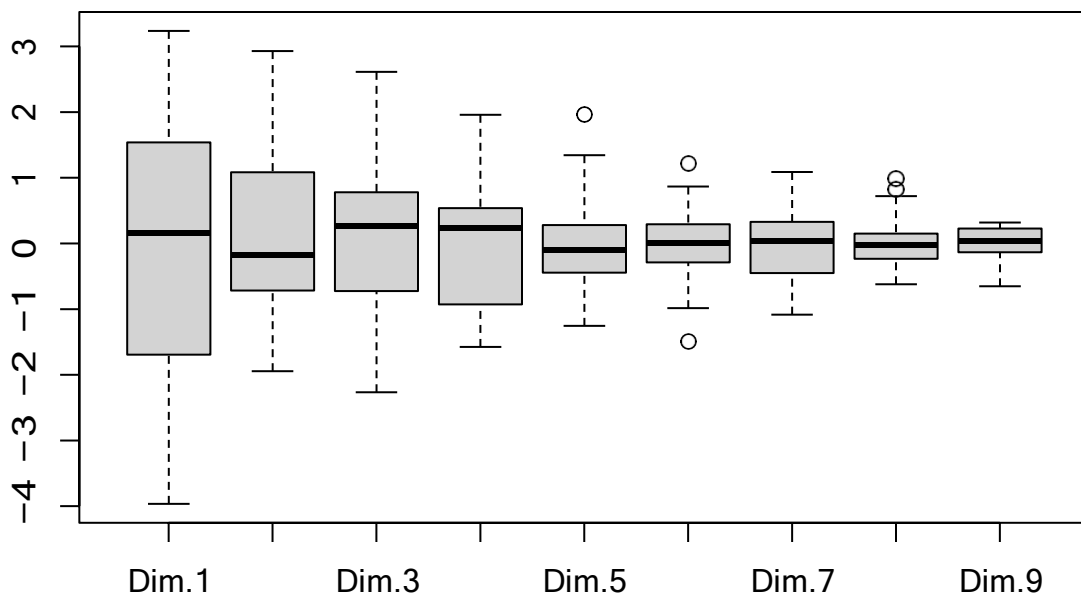
Les variables pv.ap et trv.pv sont corrélées négativement avec benev (en moyenne -0.65). De ce fait, on peut supposer que la présence de nombreux bénévoles au sein de la société contribue à l'amélioration du niveau de vie des populations pauvres.

Méthodes factorielles non-supervisées : réduction de variables

Suite au observation de la partie précédente, l'utilisation d'une ACP normée est fortement recommandée car elle permettra de limiter l'influence des valeurs aberrantes sur les résultats de notre analyse de la composante principale.

Analyse en Composantes Principales (ACP) : variables continues

Distribution empirique



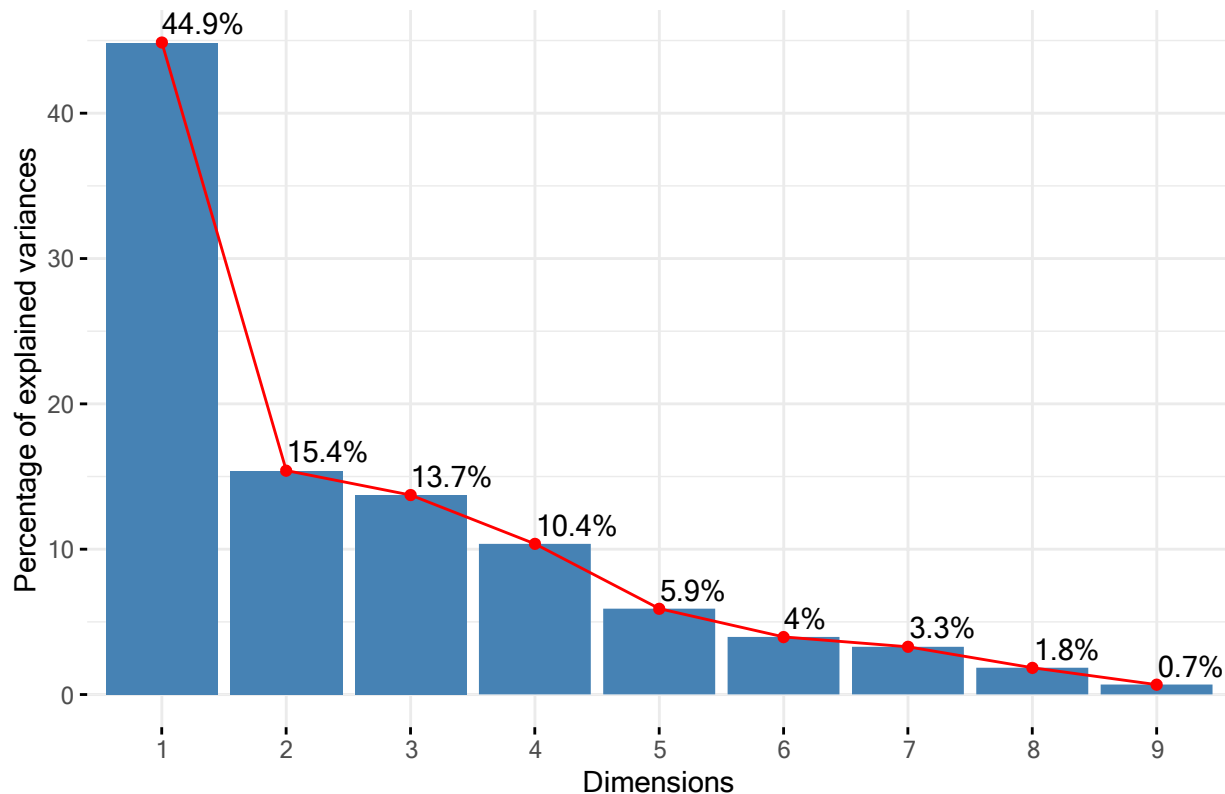
La variabilité des nouvelles variables est décroissante, ce qui était attendu. En effet, l'objectif de l'ACP est de construire de nouvelles variables, comme combinaisons linéaires des variables d'origine, d'importance décroissante, i.e. d'inertie (donc de variabilité) décroissante.

Valeurs propres

Table 4: Tableau des Valeurs Propres

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	4.0373060	44.8589552	44.85896
comp 2	1.3863071	15.4034117	60.26237
comp 3	1.2349456	13.7216179	73.98398
comp 4	0.9328492	10.3649914	84.34898
comp 5	0.5310060	5.9000665	90.24904
comp 6	0.3559071	3.9545232	94.20357
comp 7	0.2954523	3.2828039	97.48637
comp 8	0.1651987	1.8355410	99.32191
comp 9	0.0610280	0.6780893	100.00000

Éboulis des valeurs propres



Le pourcentage d'inertie porté par le plan principal (donc les deux premiers axes) est d'environ 44.9%. La somme des valeurs propres dans une ACP normée est égale au nombre de variables p , donc ici 9. Le critère de Kaiser consiste à ne conserver que les axes dont les valeurs propres sont supérieures à 1. Ici c'est le cas pour les 3 premiers axes, on peut donc les conserver. Cependant, la valeur propre associée au quatrième axe est proche de 1 et on serait donc tenter de rajouter un quatrième axe à notre analyse, mais en observant le pourcentage cumulé de la variance, on constate que les 3 premiers axes consistent à eux seuls plus de 70% de la variance, ce qui corrobore au critère de Kaiser. Il faudra vérifier la bonne représentativité des variables d'origine sur le plan principal afin de voir si on ne considère pas plus de 3 axes.

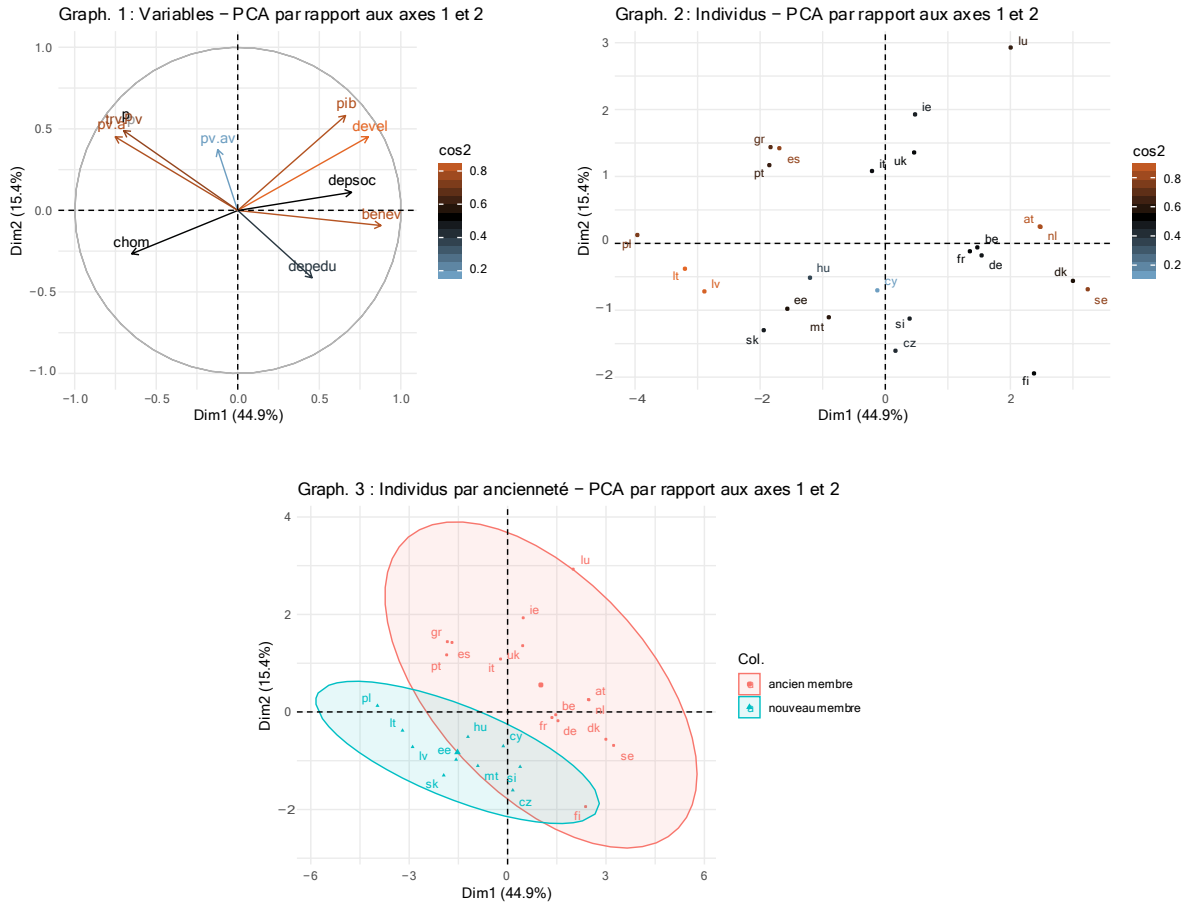
Analyse des individus vis à vis des dimensions retenues

Table 5: Corrélations des individus vis à vis des dimensions retenues

	Dim.1	Dim.2	Dim.3
at	0.8176446	0.0085652	0.0196303
be	0.4859189	0.0008207	0.1819371
cy	0.0034933	0.1048479	0.0860844
cz	0.0040857	0.3984525	0.2056063
de	0.4579424	0.0062616	0.1342968
dk	0.5220354	0.0181892	0.3957041
ee	0.3872208	0.1505126	0.0063963
es	0.4472702	0.3157647	0.0638555
fi	0.2950408	0.1974505	0.2686666
fr	0.4890625	0.0036577	0.0296648
gr	0.4091593	0.2516846	0.0639839
hu	0.2761795	0.0498700	0.1760028
ie	0.0245260	0.4056295	0.0732465
it	0.0179289	0.4592732	0.0835304
lt	0.8832700	0.0123971	0.0206219
lu	0.1863714	0.3985757	0.1969892
lv	0.8218727	0.0504585	0.0072581
mt	0.2273567	0.3384830	0.0947256
nl	0.7936099	0.0079467	0.0847815
pl	0.7258554	0.0007385	0.0278741
pt	0.4292044	0.1709749	0.0509057
se	0.7247618	0.0324972	0.1553140
si	0.0461475	0.3895359	0.0686333
sk	0.3039341	0.1354743	0.1754329
uk	0.0504579	0.4389645	0.3363092

En ce qui concerne la qualité de la représentation des individus, après avoir observé le tableau des mesures de corrélation ci-dessus, on constate que les axes 2 et 3 malgré le fait qu'ils soient pris en compte vis à vis des valeurs propres, les individus sont très mal représentés dans ce plan (2,3). On va donc par la suite porter moins d'importance pour ce dernier.

Analyse des axes 1 et 2

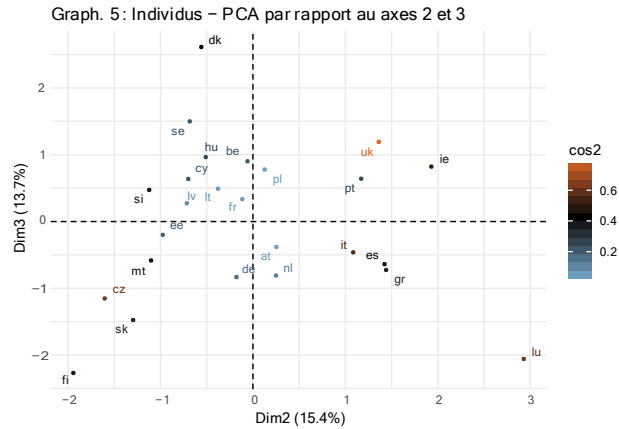
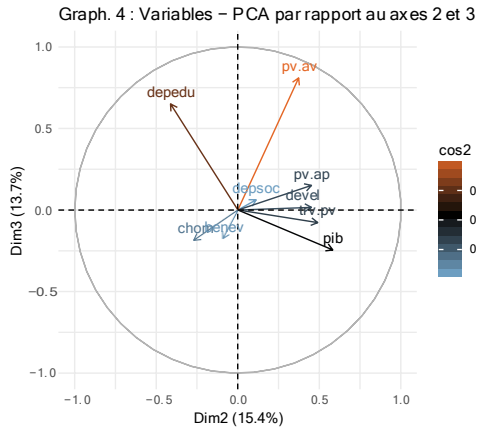


Les flèches associées aux variables pv.ap, trv.pv, pib, devel et benev sont proches du cercle (cos2 vaut en moyenne 0.7): ces variables sont toutes bien représentées et pourront donc être utilisées pour l'interprétation des deux axes.

L'axe 1 (44.9%) est construit essentiellement par la contribution des variables benev, devel, pib. Toutes ces variables contribuent fortement et positivement à la formation de cet axe. Pour les variables trv.pv, pv.ap, elles contribuent négativement à la formation de cet axe. (cf. Graphe 1) D'ailleurs, la variable pib contribue moins à l'axe 1 par rapport aux autres variables, ce qui peut s'expliquer par la surreprésentation du Luxembourg. La Suède et le Danemark contribuent positivement à cet axe, ils sont caractérisés par un pib élevé, un bon indice de développement humain et un grand nombre de personnes exerçant une activité bénévole. Ainsi, ce sont les pays dont le système social et économique sont les plus performants dans notre échantillon. Tandis que la Pologne et la Lituanie (nouveaux membres de l'UE, cf. Graphe 3) contribuent négativement, elles sont caractérisées par un taux de chômage et de pauvreté élevé. Par conséquent, ils font partie des pays ayant le système social et économique les plus défaillants dans l'Union Européenne (cf. Graphe 2). De ce fait, l'axe 1 représente l'indice de performance du système social et économique des pays de l'UE.

L'axe 2 (17.2%) est construit essentiellement par la contribution des variables pib, devel, trv.pv et pv.ap. Ces variables contribuent positivement à la formation de cet axe. Par ailleurs, on observe un angle d'environ 90° entre les couples de variables pib, devel et trv.pv, pv.ap, ce qui signifie que ces deux groupes de variables sont peu corrélés entre eux (cf Graphe 1). Quant aux individus on trouve par ordre : Luxembourg et Finlande qui contribuent respectivement positivement et négativement à l'axe 2. Ainsi, on peut conclure que le 2ème axe désigne l'inégalité au sein de la société des pays de l'UE, cela pourrait se référer à l'écart de revenu.

Analyse des axes 2 et 3



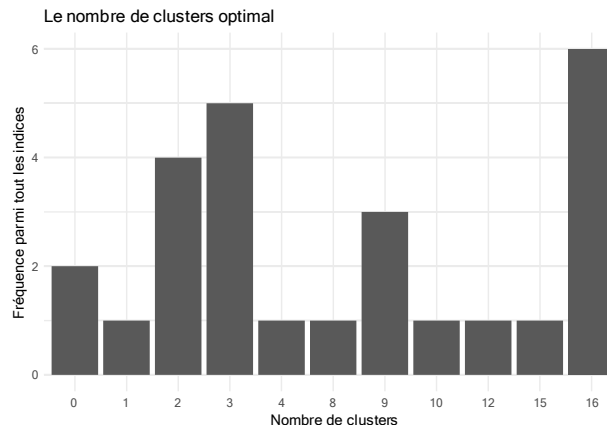
L'axe 3 (13.7%) est construit essentiellement par la contribution de variable pv.av, elle contribue positivement à la formation de cet axe (cf Graphe 4). Le danemark et le Royaume-Uni contribuent positivement à cet axe, ils sont caractérisés par un taux de pauvreté avant transferts sociaux élevé. Tandis que la Finlande ou encore le Luxembourg se caractérisent par un taux de pauvreté avant transferts sociaux plus faible (cf Graphe 5). Cet axe représente le taux de pauvreté avant transferts sociaux, qui peut s'expliquer par un coût de vie relativement élevé dans certains pays.

L'axe 2 (17.2%) ne peut pas être interprété car les autres variables sont mal représentées sur ce plan (2,3).

Méthodes de classification non-supervisée : réduction d'individus

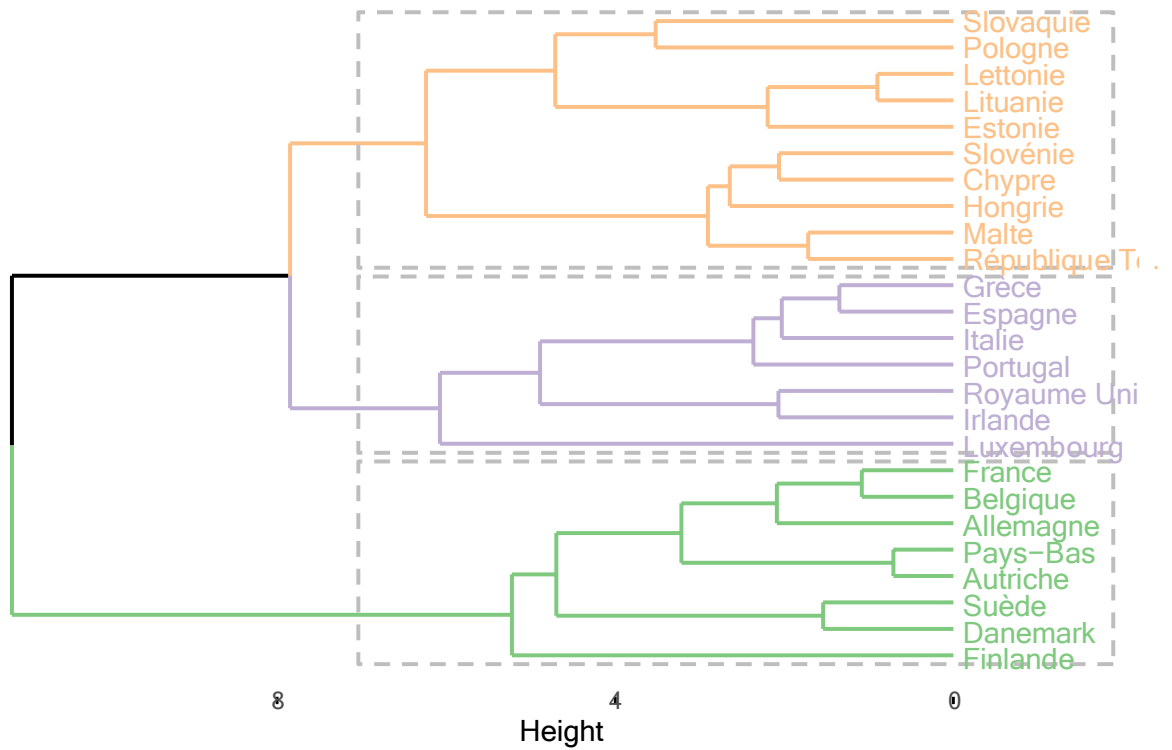
À la suite de notre étude sur les variables, nous n'avons pas distingué de "patterns" dans nos données autre que vis à vis de l'ancienneté des membres. C'est pour cela que nous avons opté pour une classification non supervisée.

Classification Ascendante Hiérarchique (CAH)



Selon les 30 critères de la fonction Nbclust, ces derniers sélectionnent en majorité 16 et 3 groupes. On en choisit 3 ici car 16 est un nombre de clusters relativement élevé dans CAH et indique que les pays sont très différents et qu'il peut être difficile de trouver des motifs ou des tendances claires. On peut donc représenter le dendrogramme avec ces 3 groupes.

Cluster Dendrogram



On constate que tous les nouveaux membres (2004) se sont retrouvés regrouper dans le même groupe. Affirmant l'hypothèse qu'ils ont de grande similitudes entre eux.

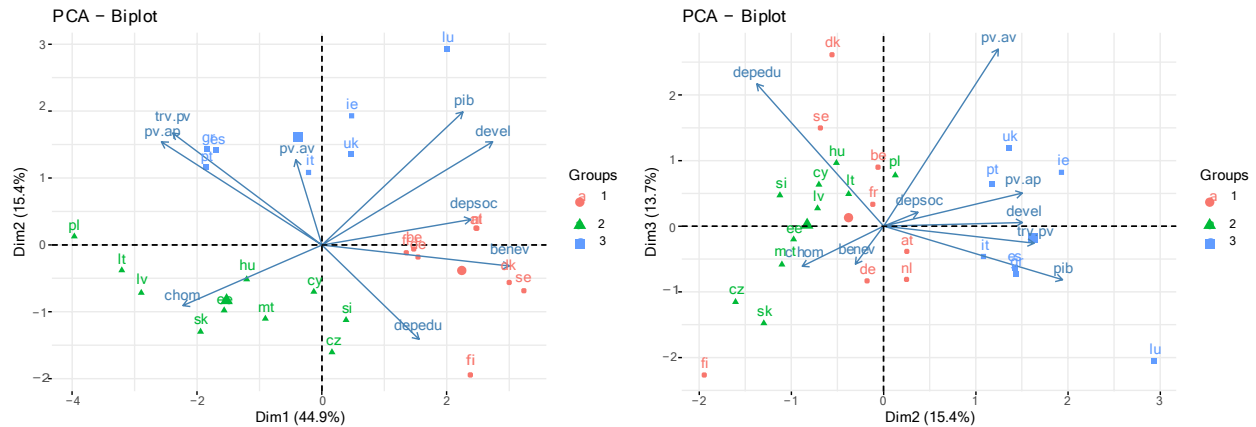


Table 6: Tableau des moyennes des variables par groupe

Group.1	benev	chom	depedu	depsoc	devel	pib	pv.ap	pv.av	trv.pv
1	39.5	6.9	5.6	26.5	0.9	115.2	15.0	24.9	5.9
2	23.9	9.9	5.1	18.5	0.9	68.1	16.2	25.1	7.6
3	45.0	4.2	3.9	22.6	0.9	278.3	13.0	23.0	9.0

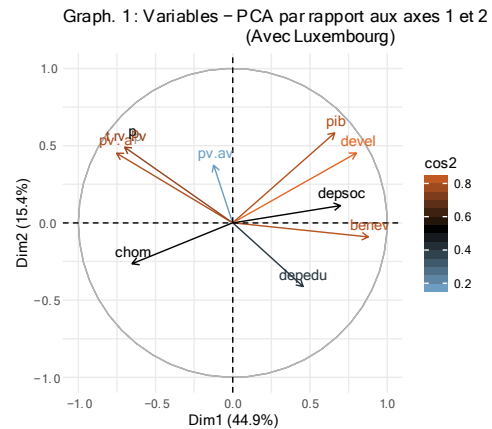
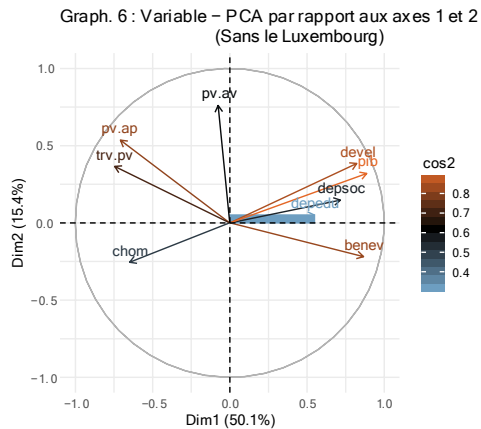
Pour la classification en 3 groupes, chacun se distinguent par leur situation économique et sociale. Cependant, ils sont assez similaires en ce qui concerne l'indice de développement humain (devel).

- Groupe 1: Pays développés économiquement et socialement avec un taux de pauvreté plus faible.
- Groupe 2: Pays en développement (pib et dépenses de protection sociale faibles) avec un taux de chômage plus important. On a essentiellement des pays qui sont nouveaux membres de l'UE (pays de l'Est).
- Groupe 3: Pays développés économiquement (pib élevé) mais présentant des inégalités fortes au sein de la société (taux de pauvreté plus élevé avec et sans aides sociales), se caractérisant également par de faibles dépenses dans l'éducation et la protection sociale.

Ainsi, on voit que l'axe 1 permet de distinguer assez bien ces 3 groupes allant des pays les moins développés (à gauche ou groupe 3) aux pays les plus développés (à droite ou groupe 1). L'axe 2 oppose les groupes 2 et 3 qui se distinguent d'un point de vue social (depsoc).

Alternative

On a remarqué que l'individu lu (Luxembourg) est sur-représenté. En effet, il a un pib très élevé, c'est un pays riche avec peu d'habitants. Pour éviter qu'il ne déforme les axes, on a retiré cet individu et réalisé une nouvelle analyse en composantes principales.

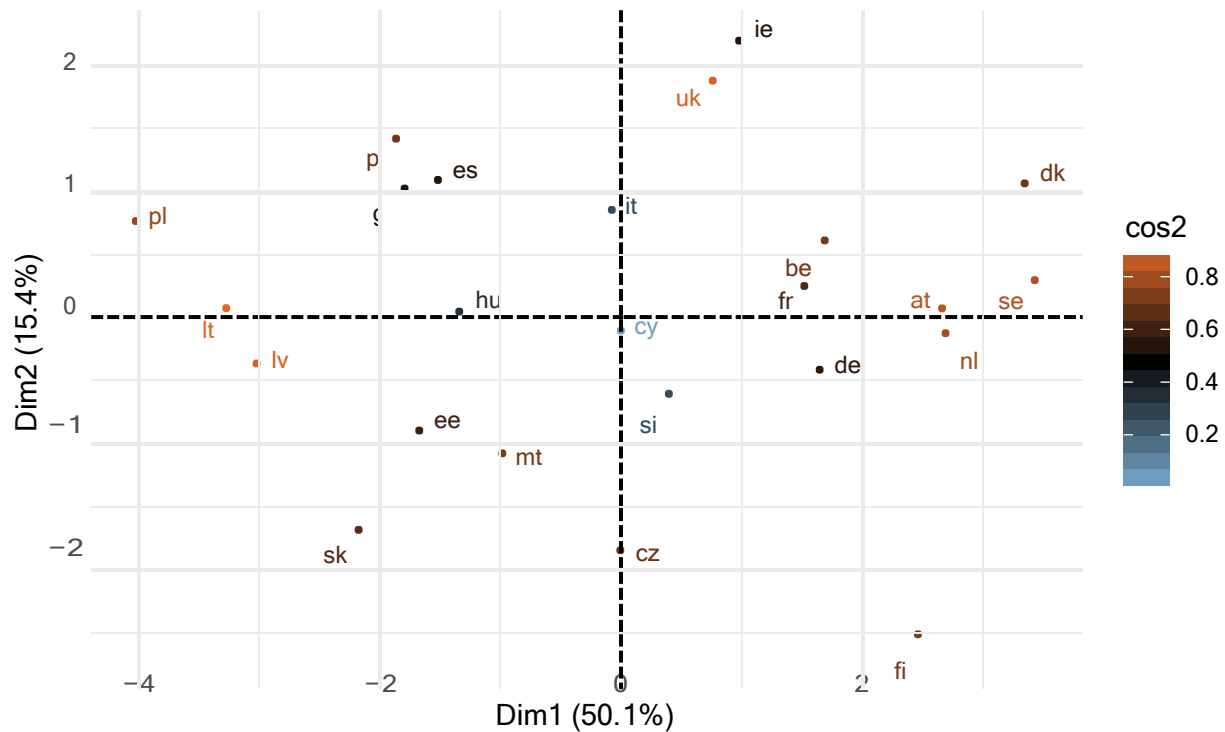


Les flèches associées aux variables pv.ap, pv.ap, trv.pv, pib, devel et benev sont proches du cercle: ces variables sont toutes bien représentées et pourront donc être utilisées pour l'interprétation des deux axes.

L'axe 1 (50.1%) est construit essentiellement par la contribution des variables benev, devel et pib. Toutes ces variables contribuent fortement et positivement à la formation de cet axe. Pour la variables trv.pv, elle contribue négativement à sa formation. Cet axe représente l'indice de performance du système sociale et économique des pays de l'UE. Cet axe oppose la Suède et le Danemark caractérisé par un bon niveau de développement, à la Pologne et la Lituanie (pays de l'Est).

L'axe 2 (15.4%) est construit essentiellement par la contribution de la variable pv.av. Cette variable contribue positivement à la formation de cet axe. Cet axe met en avant les pays où le taux de pauvreté avant transferts sociaux est important. C'est notamment le cas du Royaume-Unie. À l'inverse, La République Tchèque et la Slovaquie ont un taux significativement plus faible. Notons que ces 2 pays formaient autrefois la Tchécoslovaquie. On peut conclure que le 2ème axe désigne l'inégalité au sein de la société des pays de l'UE.

Graph. 7 : Individus – PCA par rapport aux axes 1 et 2
(Sans le Luxembourg)



Les flèches associées aux variables pv.ap, pv.ap, trv.pv, pib, devel et benev sont proches du cercle: ces variables sont toutes bien représentées et pourront donc être utilisées pour l'interprétation des deux axes.

L'axe 1 (50.1%) est construit essentiellement par la contribution des variables benev, devel et pib. Toutes ces variables contribuent fortement et positivement à la formation de cet axe. Pour la variables trv.pv, elle contribue négativement à sa formation. Cet axe représente l'indice de performance du système sociale et économique des pays de l'UE. Cet axe oppose la Suède et le Danemark caractérisé par un bon niveau de développement, à la Pologne et la Lituanie (pays de l'Est).

L'axe 2 (15.4%) est construit essentiellement par la contribution de la variable pv.av. Cette variable contribue positivement à la formation de cet axe. Cet axe met en avant les pays où le taux de pauvreté avant transferts sociaux est important. C'est notamment le cas du Royaume-Unie. À l'inverse, La République Tchèque et la Slovaquie ont un taux significativement plus faible. Notons que ces 2 pays formaient autrefois la Tchécoslovaquie. On peut conclure que le 2ème axe désigne l'inégalité au sein de la société des pays de l'UE.

Conclusion

Au cours de notre analyse, on a donc pu constater que l'Union Européenne est composée de pays qui se distinguent par leur situation économique et sociale. D'un côté les pays les plus développés, que ce soit économiquement ou socialement, tel que la Suède. D'un autre, les pays en cours de développement, comme la plupart des nouveaux membres arrivés en 2004 (Pologne, Lituanie, Lettonie). Cependant, on remarque aussi que certains pays présentent des inégalités au sein de leur population comme le Royaume-Uni.

La France, pays fondateur de l'UE, fait partie des pays développés, avec un pib, un indice de développement et un taux de bénévolat élevé. Le taux de pauvreté et donc les inégalités y sont moins présentes. En effet, d'après la Commission Européenne, la France se place 2ème puissance économique de l'UE.

Pour conclure, le fait qu'un pays soit socialement développé est lié au fait qu'il soit bien développé économiquement, à noter que dans la réalité ce n'est pas toujours appliqué.