

## Correctif TP 6

### Exercice 1

```
*ACP
FACTOR
/VARIABLES v205 v206 v207 v208 v209 v210 v211 v212 v213 v214 v215 v216 v217 v218 v219 v220 v221
v222
/MISSING LISTWISE
/ANALYSIS v205 v206 v207 v208 v209 v210 v211 v212 v213 v214 v215 v216 v217 v218 v219 v220 v221
v222
/PRINT UNIVARIATE INITIAL KMO EXTRACTION ROTATION
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.
```

On déjà réalisé cette ACP → Résultats dans le correctif du TP 4.

Rotation de la matrice des composantes

	Composante				
	1	2	3	4	5
how much confidence in: church (Q63A)	0,228	-0,015	0,036	-0,105	0,734
how much confidence in: armed forces (Q63B)	-0,048	0,313	-0,006	0,167	0,721
how much confidence in: education system (Q63C)	0	-0,115	0,52	0,27	0,301
how much confidence in: the press (Q63D)	0,252	0,178	-0,004	0,67	-0,05
how much confidence in: trade unions (Q63E)	0,087	0,036	0,074	0,761	0,095
how much confidence in: the police (Q63F)	0,301	0,105	0,274	0,317	0,391
how much confidence in: parliament (Q63G)	0,732	0,271	0,08	0,189	0,088
how much confidence in: civil service (Q63H)	0,524	0,162	0,291	0,098	0,21
how much confidence in: social security system (Q63I)	0,235	0,187	0,725	0,078	-0,042
how much confidence in: european union (Q63J)	0,355	0,688	0,197	0,081	0,053
how much confidence in: NATO (Q63K)	0,136	0,863	0,081	0,098	0,209
how much confidence in: united nations organisation (Q63L)	0,189	0,844	0,128	0,157	0,107
how much confidence in: health care system (Q63M)	0,097	0,189	0,785	0	0,03
how much confidence in: justice system (Q63N)	0,498	0,141	0,301	0,174	0,172
how much confidence in: major companies (Q63O)	0,293	0,127	0,114	0,088	0,392
how much confidence in: environmental organizations (Q63P)	0,221	0,093	0,16	0,481	0,082
how much confidence in: political parties (Q63Q)	0,795	0,071	0,008	0,232	0,092
how much confidence in: government (Q63R)	0,824	0,166	0,081	0,135	0,104

Nous allons maintenant réaliser une analyse de fidélité afin de voir si les items corrélés à chaque facteur sont fidèles à cette dimension et peuvent être utilisés en échelle.

```
*Première dimension
RELIABILITY
/VARIABLES=v211 v212 v221 v222 v214 v218 v210
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE
/SUMMARY=TOTAL.
```

L'alpha de Cronbach est 0.828, ce qui est très bon : les items peuvent être utilisés dans le cadre d'une construction d'échelle.

Statistiques de total des éléments				
	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
how much confidence in: parliament (Q63G)	15,52	9,142	,692	,784
how much confidence in: civil service (Q63H)	15,74	10,155	,525	,812
how much confidence in: political parties (Q63Q)	15,16	9,699	,606	,799
how much confidence in: government (Q63R)	15,34	9,208	,694	,784
how much confidence in: european union (Q63J)	15,81	10,032	,488	,818
how much confidence in: justice system (Q63N)	15,67	9,701	,534	,811
how much confidence in: the police (Q63F)	15,98	10,214	,476	,820

En regardant de près la statistique de Cronbach si l'un des items est supprimé de l'échelle : on regarde que l'alpha de Cronbach diminue dans tous les cas. Cela signifie donc que chaque item contribue bien à l'échelle.

Ensuite, SPSS nous renseigne des statistiques sur l'échelle que nous pourrions créer en sommant les variables :

Statistiques d'échelle			
Moyenne	Variance	Ecart type	Nombre d'éléments
18,20	12,881	3,589	7

Cette échelle aurait une moyenne de 18 sur 28 (4 possibilités de réponses sur 7 éléments). Ceci nous indique que, de façon générale, la population européenne n'accorde pas une grande confiance envers les institutions étatiques<sup>1</sup>.

<sup>1</sup> La moyenne est de 18. Or, les possibilités de réponse dans les variables initiales vont de 1(confiance) à 4(pas confiance). Le score minimal possible est donc de 7 et le score maximal de 28. Afin d'avoir une variable plus clair,

Il faut ensuite procéder de la même manière pour les autres dimensions.

\*Deuxième dimension

RELIABILITY

/VARIABLES=v206 v214 v215 v216

/SCALE('ALL VARIABLES') ALL

/MODEL=ALPHA

/STATISTICS=DESCRIPTIVE SCALE

/SUMMARY=TOTAL.

L'alpha de Cronbach est ici moins bon (0.767) mais toujours acceptable selon les critères de Nannuly. Cela est relativement intuitif, vu que les composantes principales ont une valeur propre qui diminue par à la composante principale précédente.

#### Statistiques de total des éléments

	Moyenne de l'échelle en cas de suppression d'un élément	Variance de l'échelle en cas de suppression d'un élément	Corrélation complète des éléments corrigés	Alpha de Cronbach en cas de suppression de l'élément
how much confidence in: armed forces (Q63B)	7,13	3,949	,323	,836
how much confidence in: european union (Q63J)	7,26	3,458	,564	,714
how much confidence in: NATO (Q63K)	7,26	3,036	,731	,620
how much confidence in: united nations organisation (Q63L)	7,31	3,099	,693	,642

Dans ce cas-ci, la variable « Armed forces » semble moins fidèle à la dimension car si on la retire de l'analyse, l'alpha de Cronbach remonte vers un seuil bien meilleur de 0.836. On pourrait donc envisager de créer une échelle avec les seuls trois items « European Union », « NATO » et « United Nations Organization ».

Et ainsi de suite pour les autres dimensions...

Pour créer les échelles, il suffit de créer une variable avec les items concernés en les sommant... Soit en utilisant la fonction « COMPUTE VARIABLE ». Cette variable sera ensuite plus intuitive à interpréter qu'une composante principale car le score ne sera pas relatif au centre de gravité.

## Exercice 2

Changez de base de données !

\*ACP

FACTOR

/VARIABLES V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V21 V23 V24 V22

il faut donc soustraire ce chiffre minimal et diviser par le nombre maximal pour avoir une indication en %.  $(18-7)/(28-7) = 11/21 = 52.3$ . L'échelle étant inversée, un petit score signifiant une grande confiance accordée, on trouvera l'indicateur en % de confiance en soustrayant ce pourcentage à 1. Soit  $1-0,523 = 0,473$ . Les européens ont une confiance de 47% envers les institutions politiques.

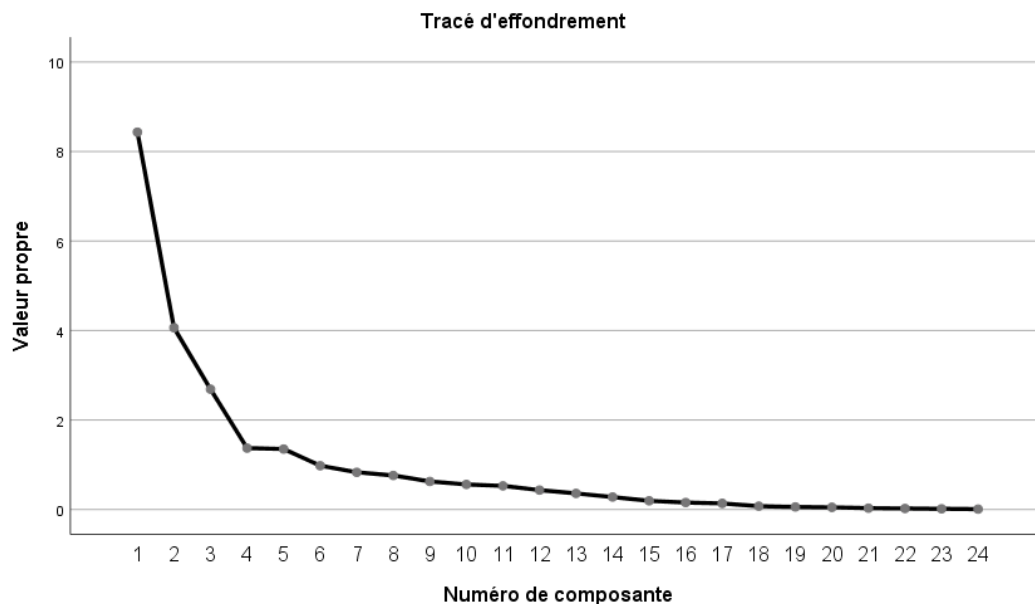
```

/MISSING LISTWISE
/ANALYSIS V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V21 V23
V24 V22
/PRINT UNIVARIATE INITIAL KMO EXTRACTION ROTATION
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.

```

L'indice KMO est de 0.721, p-valeur du test de sphéricité de Bartlett est inférieure à 0.0001 : Nous avons des corrélations suffisamment correctes que pour faire tourner une ACP.

5 facteurs ont une valeur-propre supérieure à 1, reprenant 74% de la variance ! Le gain en concentration d'information est ici colossal ! La méthode du coude nous suggérerait de conserver 4 facteurs (68%). Cette option est envisageable au vu du faible gain d'information apportée par l'axe factoriel 5 par rapport aux autres.



Description des axes :

	Composante				
	1	2	3	4	5
International migrant stock, total	-0,19	0,12	0,93	0,04	0,05
Adjusted net national income per capita (current US\$)	0,47	0,59	0,14	-0,37	0,30
Pupil-teacher ratio, primary	0,08	-0,23	0,08	0,20	0,65
Pupil-teacher ratio, secondary	0,01	-0,03	-0,18	0,03	0,59
Proportion of seats held by women in national parliaments (%)	0,28	0,49	0,00	-0,33	-0,13
Survival to age 65, female (% of cohort)	0,89	0,36	0,05	-0,10	-0,01
Survival to age 65, male (% of cohort)	0,93	0,25	0,02	-0,18	0,11
Suicide mortality rate (per 100,000 population)	-0,86	0,23	0,05	0,10	-0,01
Smoking prevalence, total (ages 15+)	-0,07	-0,07	-0,03	0,81	-0,03
Population, total	-0,27	0,05	0,91	0,11	0,02
Population growth (annual %)	0,43	0,38	0,01	-0,28	0,40
Mortality rate, under-5 (per 1,000 live births)	-0,50	-0,75	-0,04	0,05	0,20
Life expectancy at birth, total (years)	0,85	0,40	0,12	-0,23	0,08

International tourism, number of arrivals	0,24	0,08	0,81	0,05	-0,03
International tourism, number of departures	0,05	0,07	0,79	-0,10	-0,14
International tourism, receipts (current US\$)	0,38	0,14	0,83	-0,09	-0,05
Intentional homicides (per 100,000 people)	-0,76	-0,14	0,01	-0,14	0,00
Military expenditure (% of central government expenditure)	-0,34	-0,12	0,44	0,64	0,08
Armed forces personnel (% of total labor force)	0,00	-0,42	-0,09	0,66	0,20
Vulnerable employment, total (% of total employment)	0,19	-0,91	-0,09	0,06	0,07
GDP per person employed (constant 2011 PPP \$)	0,47	0,65	0,10	-0,25	0,25
Employment in agriculture (% of total employment)	-0,22	-0,91	-0,13	0,02	0,14
Unemployment, total (% of total labor force)	0,04	-0,45	-0,08	0,28	-0,42
Employment in services (% of total employment)	0,32	0,86	0,15	-0,16	-0,01

Facteur 1 : Indicateur de développement démographique

Facteur 2 : Indicateur de développement socio-économique

Facteur 3 : Indicateur de flux des populations sur le territoire

Facteur 4 : Indicateur d'investissement militaire (au détriment des politiques de santé)<sup>2</sup>

Facteur 5 : Indicateur d'augmentation démographique scolaire

### Exercice 3

```
*Application du filtre
COMPUTE filter_$=(Time = 2010).
VARIABLE LABELS filter_$ 'Time = 2010 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

*Analyse en cluster.
DATASET ACTIVATE DataSet1.
CLUSTER  demographics economics populationflow mitilarysmoking educationsaturation
/METHOD WARD
/MEASURE=SEUCLID
/PRINT SCHEDULE
/PLOT DENDROGRAM VICICLE
/SAVE CLUSTER(2,6).
```

Pour rappel : deux manières d'évaluer le nombre de clusters à conserver.

#### 1) Le tableau d'agglomération itératif.

Regarder le gain du coefficient d'inertie par rapport à l'étape précédente de fusion. Possibilité de construire un ratio dans Excel en calculant simplement le quotient entre le ratio de l'étape et celui de l'étape précédente.

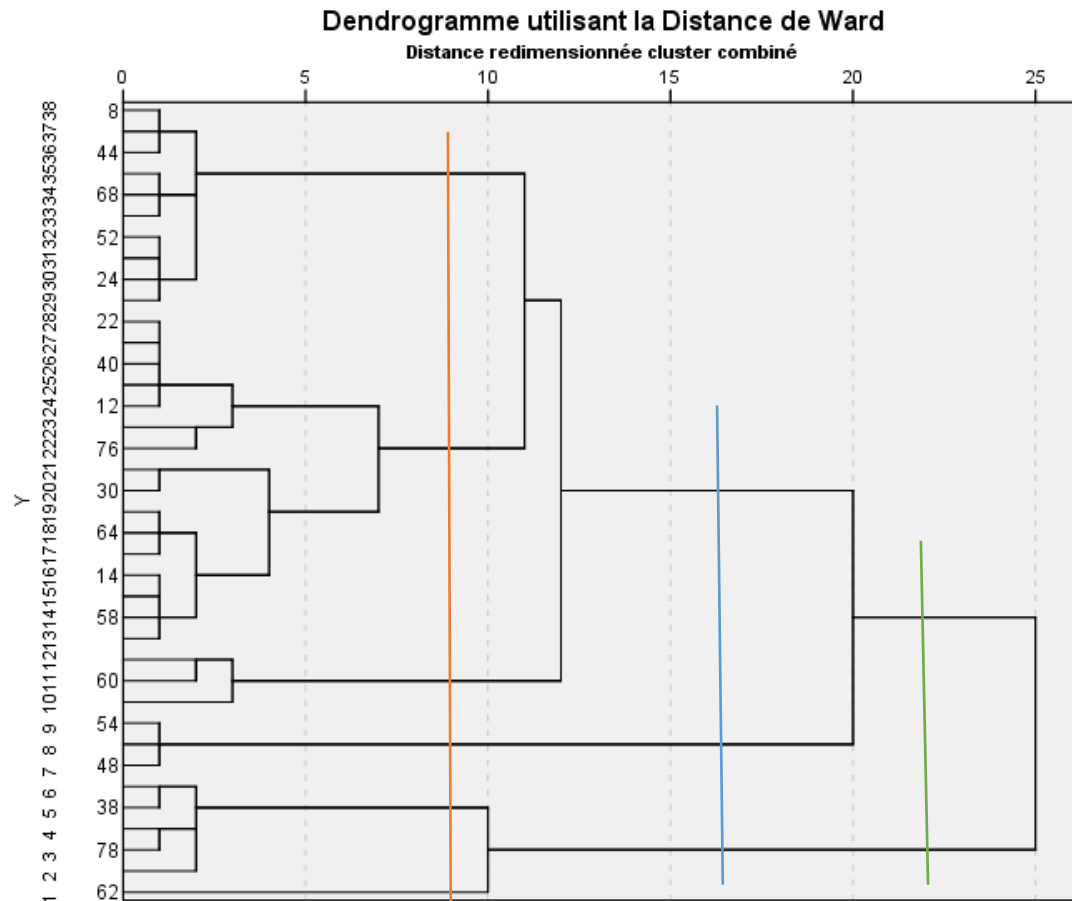
Planning des agglomérations									
Etape	Cluster combiné		Coefficients	Etape de première apparition du cluster		Etape suivante			
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	Ratio			
Etapes 1 à 26...									
27	4	8	26,152	22	16	34	1,11842307		
28	26	28	29,249	18	24	33	1,12085883		
29	2	50	32,784	23	0	35	1,12490849		
30	6	12	36,879	26	20	32	1,1683614		
31	10	16	43,088	25	17	32	1,23510026		
32	6	10	53,218	30	31	34	1,29001466		

<sup>2</sup> L'interprétation de cet axe n'est pas des plus simples : pourquoi est-ce que le tabagisme semble corrélé aux contingents armés et au pourcentage du budget alloué aux dépenses militaires ? Une étude démontre qu'en générale, l'allocation du budget au militaire se fait en général au détriment des politiques de santé (<https://www.tandfonline.com/doi/abs/10.1080/10242694.2017.1303303>).

33	26	62	68,652	28	0	37	1,23817223
34	4	6	85,003	27	32	35	1,2122631
35	2	4	103,046	29	34	36	1,29354851
36	2	48	133,295	35	21	37	1,29782813
37	2	26	172,994	36	33	0	0

Sur ces bases, il semblerait qu'il soit idéal de préserver 2, 3 ou 6 clusters.

2) L'analyse du dendrogramme



L'analyse du dendrogramme semble également nous suggérer de nous diriger vers 2,3 ou 6 clusters. Pour une raison spécifique, nous allons décider d'en conserver 5 : L'analyse en 3 facteurs préserve un groupe très grand et deux petits groupes, ce qui ne permettra sans doute pas d'obtenir des résultats différenciant bien les pays. Nous nous dirigeons donc vers la classification en 6 groupes. Cependant, à la sixième étape avant la fusion totale, l'algorithme sépare un de nos groupes déjà petits en créant un groupe conservant une seule observation, ce qui n'est pas très intéressant. Nous nous arrêterons donc une étape plus loin : 5 clusters.

#### Exercice 4

\*Analyse descriptive bivariable de variables quantitatives par une variable qualitative → Means.

MEANS TABLES=FAC1\_1 FAC2\_1 FAC3\_1 FAC4\_1 FAC5\_1 V2 V11 BY CLU5\_1

/CELLS=MEAN COUNT STDDEV.

\*Variables qualitatives → Crosstabs.

CROSSTABS

/TABLES=CLU5\_1 BY Continent Union

/FORMAT=AVALUE TABLES

/CELLS=COUNT ROW

/COUNT ROUND CELL.

\*ID= Indicateur de développement

		Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
	Effectifs	3	10	16	6	3
Facteurs	ID démographique	-0,21	0,51	-0,06	0,28	0,66
	ID socioéconomique	-1,94	0,74	-0,08	0,18	0,78
	ID flux pop.	-0,31	-0,32	-0,38	2,43	-0,39
	Mil-Health bal.	-1,46	-0,86	0,2	-0,32	0,51
	Education pop.	-0,11	-0,19	-0,89	-0,06	2,62
V2	Population totale (en millions)	8,9	7,3	10,2	76,4	4,4
V11	PIB / habitant (en milliers \$)	3,9	37,7	10,3	27,7	45,9
Subcontinent	% Europe-Est	67	20	44	17	0
	% Europe –Nord	0	30	19	0	33
	% Europe-Sud	33	0	38	33	33
	% Europe-Ouest	0	50	0	50	33
	% Union Européenne	33	100	70	83	33

Groupe 1 : 3 pays du sud et de l'est, très pauvres en moyenne. Tous les indicateurs factoriels sont négatifs. L'indice socio-économique est le plus faible de tous ainsi l'indicateur d'investissement militaire.

Groupe 2 : 10 pays, majoritairement constitués de pays de l'Ouest. Les pays ont en moyenne, une population relativement basse mais une PIB/ habitant élevé. Les indices socio-économiques et démographiques sont par ailleurs positifs, indiquant une certaine modernisation et croissance démographique relativement aux autres groupes. Ce groupe est le seul à être constitués uniquement de pays membres de l'union. C'est un groupe où l'allocation du budget vers le militaire est un des plus faible également.

Groupe 3 : 16 pays provenant majoritairement de l'Est et du Sud, plutôt pauvres. Au niveau des indicateurs, ils sont presque tous proches du centre de gravité à l'exception de la population scolaire qui semble bien se porter dans ces pays relativement aux autres.

Groupe 4 : 6 pays majoritairement d'Europe de l'Ouest qui ont la caractéristique commune d'avoir des populations conséquentes. Ces pays connaissent de fort flux de populations (migrants, tourisme). Ils ont des indicateurs socio-économiques et démographiques légèrement positifs. Il s'agit sans doute des poumons de l'Europe de par leur population.

Groupe 5 : 3 pays qui ont en moyenne une très faible population avec un PIB par habitant très élevé. Les indicateurs sont tous positifs. Ces pays connaissent en outre une forte augmentation de la population scolaire.

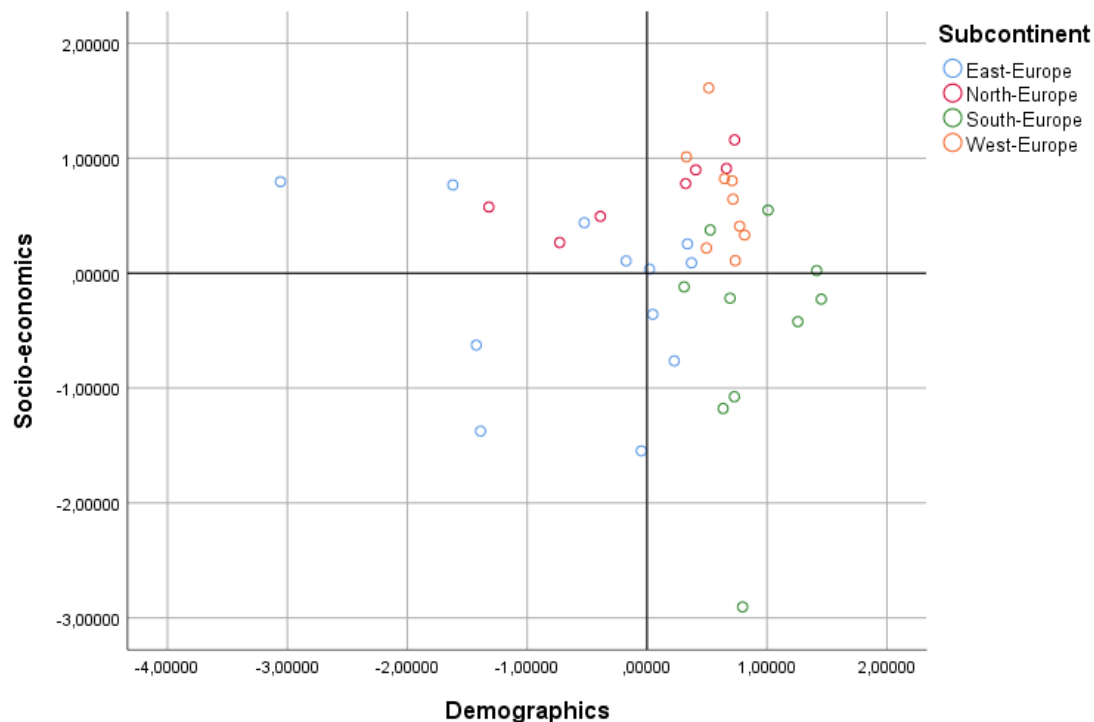
## Exercice 5

\*Scatter dot.

GRAPH

/SCATTERPLOT(BIVAR)=demographics WITH economics BY Continent

/MISSING=LISTWISE.



Ces indicateurs permettent de relativement bien décrire les pays membres d'une région de l'Europe comme défini par l'ONU. Les pays de l'Europe de L'Ouest se trouvent en positif sur les deux axes. Les pays de l'Europe du Nord, jouissent d'un bon indicateur économique mais d'un certain déficit de population, l'inverse pour les pays du Sud. Enfin, les pays de l'Est se trouvent majoritairement avec un déficit de population et plusieurs d'entre eux éprouvent également des difficultés économiques.

N'oubliez pas que les composantes principales sont des variables centrées-réduites. L'interprétation des chiffres se fait donc comparativement aux autres pays. Ainsi, l'analyse que l'on vient de faire pour les pays de l'Est serait sans doute bien différente si l'on incluait dans la base de données des pays en voie de développement.

Bonus :

\*Il suffit d'ajouter un libellé pour les observations! Il faut ensuite sélectionner dans les options le fait d'afficher les libellés  
 GRAPH  
 /SCATTERPLOT(BIVAR)=demographics WITH economics BY Continent BY CountryName (NAME)  
 /MISSING=LISTWISE.

