

# TP2-ACP

Christophe tiet  
18 avril 2019

## 1-Justification de l'ACP

	EVH	EVF	POP	TEMP	PIBH	TINF	TACT	TCHOM	TCHOMLD	MARIAG	TEL
EVH	1.00	0.90	0.31	0.06	0.70	-0.04	0.38	-0.52	-0.55	0.22	0.39
EVF	0.90	1.00	0.41	-0.08	0.65	-0.06	0.22	-0.37	-0.44	0.10	0.49
POP	0.31	0.41	1.00	-0.17	0.14	-0.29	-0.08	0.16	0.10	-0.15	0.96
TEMP	0.06	-0.08	-0.17	1.00	0.13	-0.13	0.83	-0.24	-0.28	0.25	-0.19
PIBH	0.70	0.65	0.14	0.13	1.00	-0.12	0.47	-0.64	-0.64	-0.02	0.20
TINF	-0.04	-0.06	-0.29	-0.13	-0.12	1.00	-0.11	-0.18	-0.08	-0.05	-0.24
TACT	0.38	0.22	-0.08	0.83	0.47	-0.11	1.00	-0.67	-0.71	0.23	-0.02
TCHOM	-0.52	-0.37	0.16	-0.24	-0.64	-0.18	-0.67	1.00	0.96	-0.21	0.04
TCHOMLD	-0.55	-0.44	0.10	-0.28	-0.64	-0.08	-0.71	0.96	1.00	-0.23	0.00
MARIAG	0.22	0.10	-0.15	0.25	-0.02	-0.05	0.23	-0.21	-0.23	1.00	-0.15
TEL	0.39	0.49	0.96	-0.19	0.20	-0.24	-0.02	0.04	0.00	-0.15	1.00

n= 25

P

	EVH	EVF	POP	TEMP	PIBH	TINF	TACT	TCHOM	TCHOMLD	MARIAG	TEL
EVH		0.0000	0.1259	0.7855	0.0001	0.8332	0.0574	0.0079	0.0045	0.2804	0.0526
EVF	0.0000		0.0436	0.7156	0.0004	0.7588	0.2866	0.0688	0.0278	0.6361	0.0137
POP	0.1259	0.0436		0.4191	0.5154	0.1613	0.7164	0.4368	0.6386	0.4665	0.0000
TEMP	0.7855	0.7156	0.4191		0.5387	0.5251	0.0000	0.2561	0.1713	0.2213	0.3624
PIBH	0.0001	0.0004	0.5154	0.5387		0.5735	0.0176	0.0006	0.0005	0.9241	0.3289
TINF	0.8332	0.7588	0.1613	0.5251	0.5735		0.6115	0.3901	0.7076	0.8062	0.2463
TACT	0.0574	0.2866	0.7164	0.0000	0.0176	0.6115		0.0002	0.0000	0.2754	0.9130
TCHOM	0.0079	0.0688	0.4368	0.2561	0.0006	0.3901	0.0002		0.0000	0.3114	0.8358
TCHOMLD	0.0045	0.0278	0.6386	0.1713	0.0005	0.7076	0.0000	0.0000		0.2659	0.9888
MARIAG	0.2804	0.6361	0.4665	0.2213	0.9241	0.8062	0.2754	0.3114	0.2659		0.4712
TEL	0.0526	0.0137	0.0000	0.3624	0.3289	0.2463	0.9130	0.8358	0.9888	0.4712	

D'après notre tableau de corrélation, nous remarquons que les variables EVH, EVF, PIBH, TACT sont positivement corrélées entre elles. On observe également une corrélation négative entre les variables PIBH, TACT, TEMP et les variables TCHOM, TCHOMLD. Enfin on peut voir une corrélation logique entre POP et TEL puisque plus un pays a d'habitants et plus le nombre d'abonnés aux services de téléphonie mobile est susceptible d'être élevé.

De plus, nous avons un jeu de données comportant 25 individus et 14 variables. L'ACP me paraît donc bien justifié afin de réduire ces 14 variables en essayant de perdre le moins d'informations possibles.

## 2-Choix des composantes principales

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	4.19930765	38.1755241	38.17552
comp 2	2.68386434	24.3987667	62.57429
comp 3	1.42753950	12.9776318	75.55192
comp 4	0.97162717	8.8329743	84.38490
comp 5	0.76418215	6.9471104	91.33201
comp 6	0.56105730	5.1005209	96.43253
comp 7	0.23004299	2.0912999	98.52383
comp 8	0.08363118	0.7602834	99.28411
comp 9	0.03890899	0.3537181	99.63783
comp 10	0.02187735	0.1988850	99.83671
comp 11	0.01796139	0.1632854	100.00000

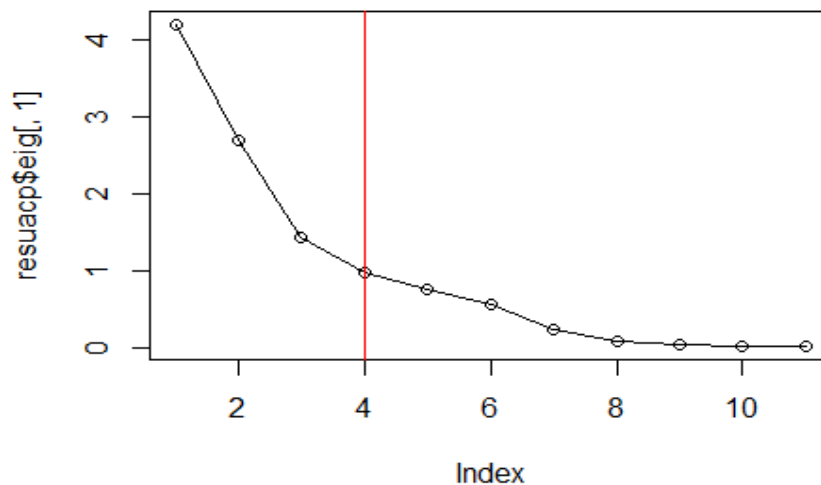
### Critère de Kaiser

Nous ne gardons ici que les trois premières composantes principales étant données qu'elles sont les seules à avoir une valeur propre  $> 1$ .

### Critère de différence

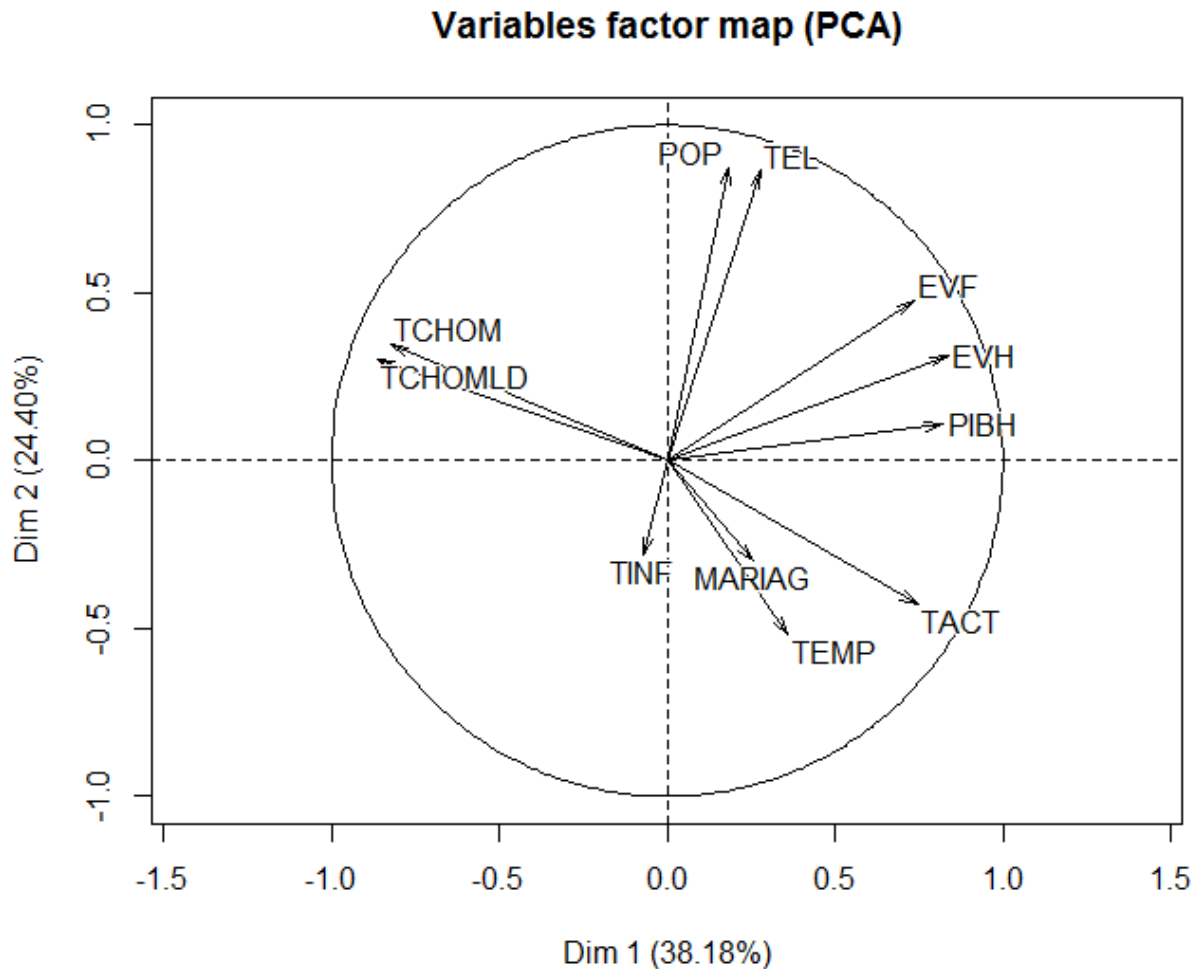
Nous ne sélectionnons que deux composantes principales puisque la différence entre les valeurs propres est de: 1.52 pour la composante 1 1.25 pour la composante 2 0.46 pour la composante 3.

D'après ces deux critères, ils semblent donc cohérents de garder trois composantes principales. Le graphe ci-dessous permet de nous rassurer sur ce choix.



### 3-Interprétation des composantes principales

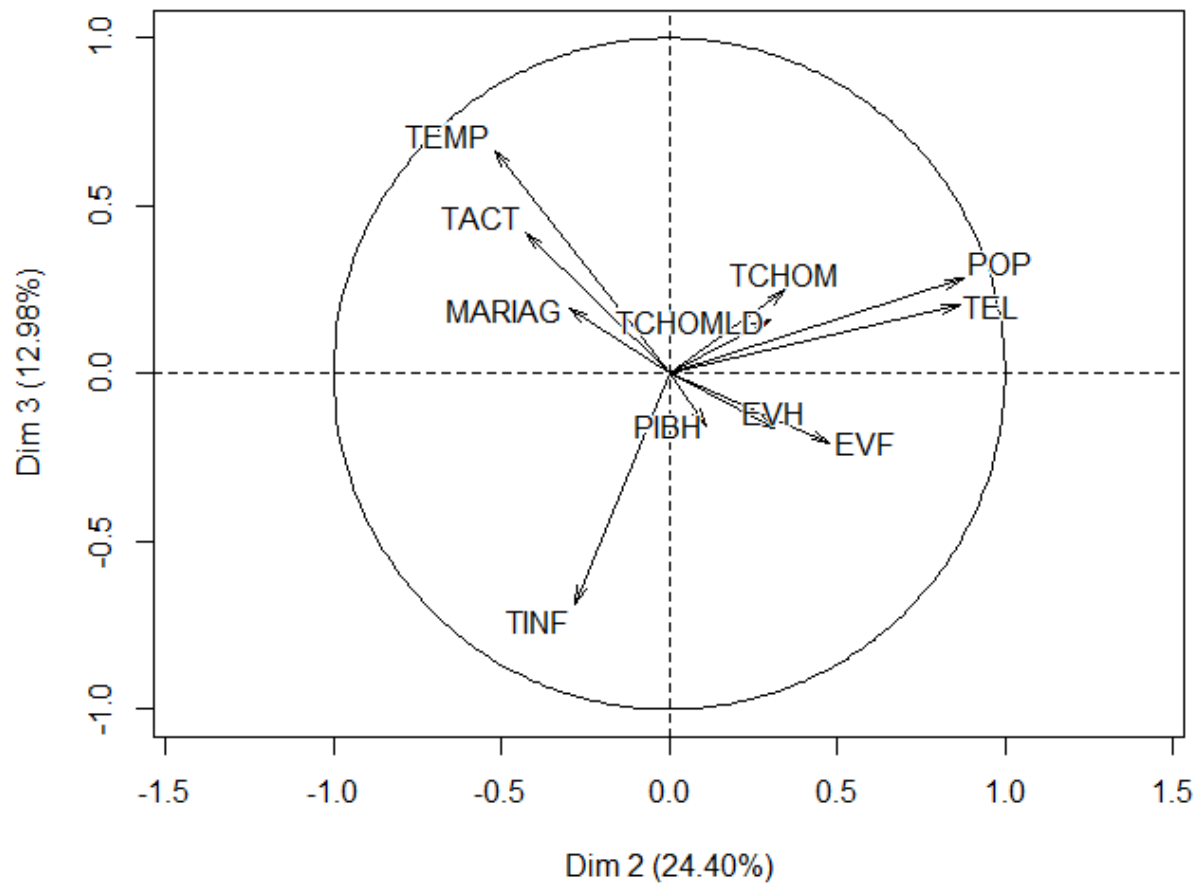
Afin d'interpréter nos trois composantes principales; nous allons nous aider des graphiques des variables sur les trois différents plans(C1/C2,C2/C3,C1/C3).



#### Composante 1:

La composante 1 a une forte corrélation positive avec l'espérance de vie (Hommes et Femmes) et le produit intérieur par habitant. On note aussi une corrélation négative avec le taux de chômage (courte et longue durée). Cette composante permet donc de mesurer le potentielle économique par rapport à l'espérance de vie d'une population. Si cette variable est grande alors l'espérance de vie est élevée, le pibh est grand et le taux de chômage est bas.

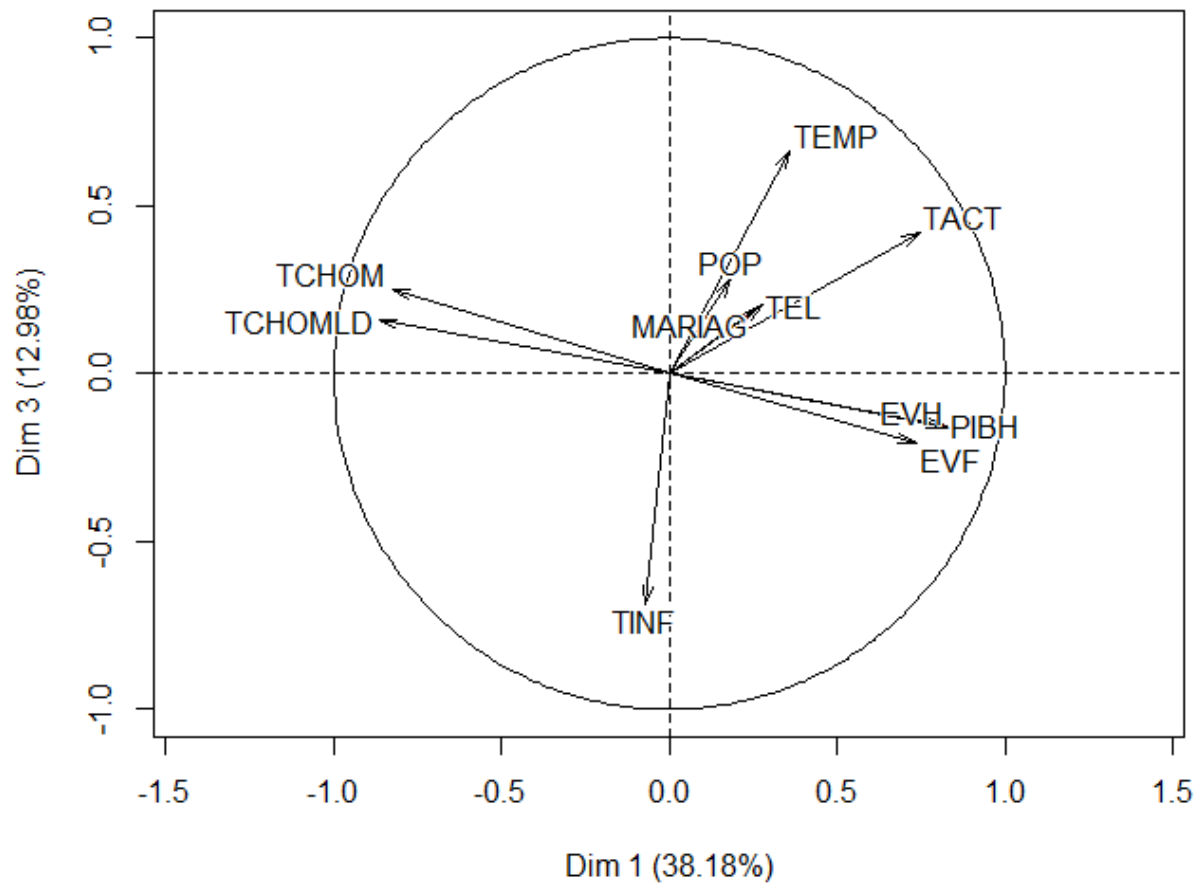
**Variables factor map (PCA)**



**Composante 2:**

La composante 2 a une forte corrélation positive avec la population d'un pays et son nombre d'abonnées aux services de téléphonie mobile. C'est composante permet de garder les informations concernant ces deux variables.

**Variables factor map (PCA)**



**Composante 3:**

La composante 3 oppose le taux d'activité et le taux d'emploi et le taux d'inflation. Plus cette composante est grande et plus le pays est actif.

#### 4-Qualité de représentation des pays.

	Dim.1	Dim.2	Dim.3
Belgique	0.004237840	0.1088965037	0.2248240027
Rep-tche	0.098233170	0.2287960146	0.3620023808
Danemark	0.404038614	0.4071776952	0.1590899843
Allemagne	0.080645377	0.6794905883	0.1502900132
Estonie	0.635206652	0.2112150578	0.0004399479
Grèce	0.174725892	0.0928860916	0.5424275720
Espagne	0.013824844	0.7146677814	0.0866228289
France	0.186138919	0.7602915839	0.0129193256
Irlande	0.330651372	0.2557234771	0.2455376102
Italie	0.008753715	0.9071638310	0.0440483887
Chypre	0.146400104	0.1945725913	0.0208837905
Lettonie	0.728480936	0.1124378843	0.0692083392
Lituanie	0.684986770	0.0320017502	0.1825275157
Luxembou	0.347530374	0.0141998426	0.1309516018
Hongrie	0.309962408	0.0477110306	0.3079377448
Malte	0.056180050	0.0035394921	0.4270458840
Pays-Bas	0.643281575	0.1604836980	0.0142533541
Autriche	0.687400262	0.0503106106	0.0009406521
Pologne	0.726351365	0.1438102835	0.0458574014
Portugal	0.209423668	0.4512611534	0.0026448778
Slovénie	0.010147408	0.1382042686	0.4032351492
Slovaqui	0.772696040	0.0005157353	0.0250872115
Finlande	0.285647746	0.1582238933	0.1406399474
Suede	0.668032096	0.0657602494	0.0354333985
Royaume-	0.407506606	0.1699173624	0.2614943618

Un pays est bien représenté si sa norme au carré est >0.5 sur le 1er axe, >0.25 sur le 2eme axe et >0.15 sur le 3eme axe.

Axe 1: Estonie, Lettonie, Lituanie, Pays-Bas, Autriche, Pologne, Slovaquie, Suède.

Axe 2: Danemark, Allemagne, Espagne, France, Italie et Portugal.

Axe 3: Belgique, République Tchèque, Danemark, Allemagne, Grèce, Irlande, Lituanie, Hongrie, Malte, Slovénie et Royaume-Uni.

## 5- Contribution des pays au 1er axe

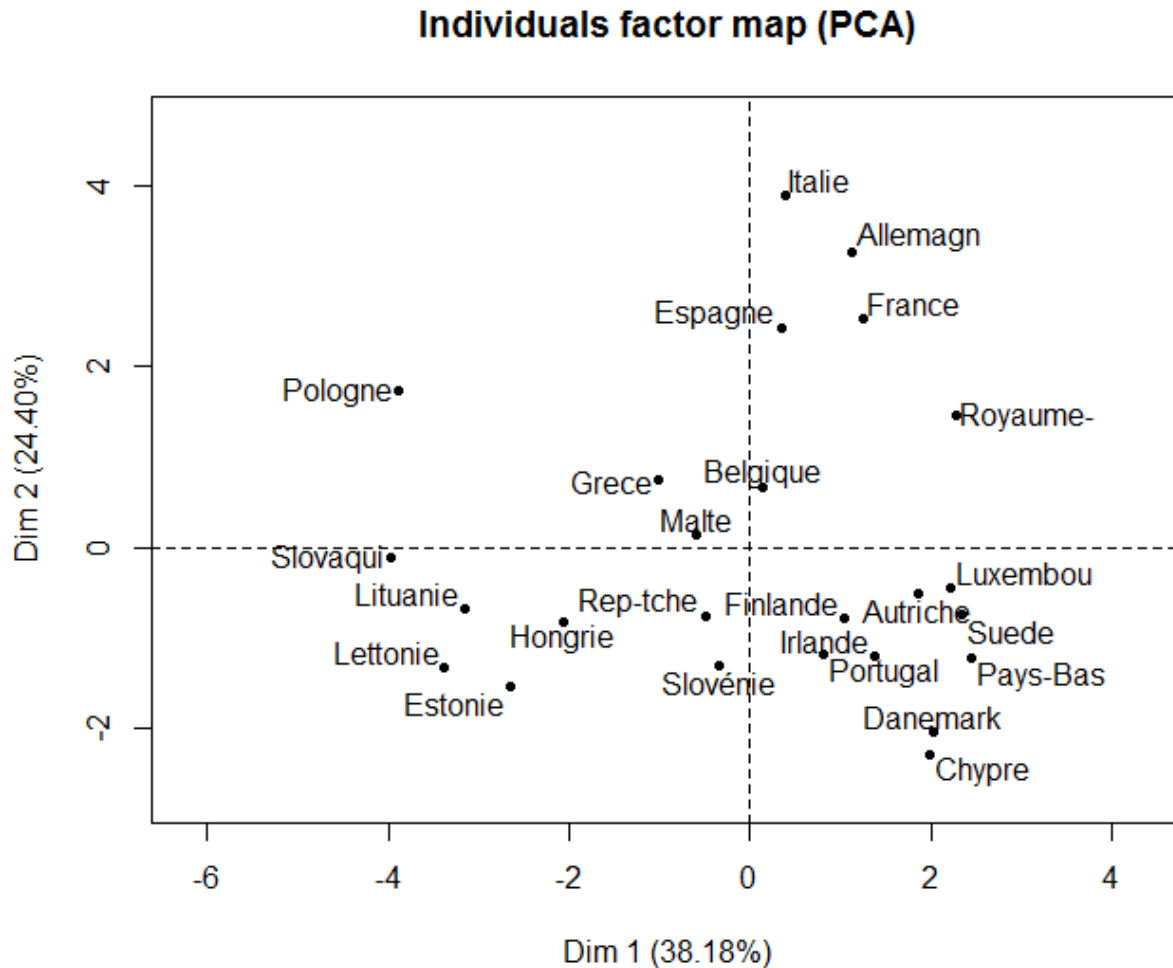
##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Belgique	0.01631346	0.65589232	2.54585528	0.094619380	13.0596719
## Rep-tche	0.23931102	0.87210771	2.59421235	0.090279275	0.7391875
## Danemark	3.88106377	6.11968431	4.49531682	0.029612627	0.0487154
## Allemagn	1.21239455	15.98324933	6.64636485	0.600126723	3.3180687
## Estonie	6.70649601	3.48917412	0.01366380	2.569392617	1.3873684
## Grece	0.98948154	0.82303491	9.03611498	2.824528141	0.1752429
## Espagne	0.10808475	8.74229615	1.99216864	1.460560785	1.6270037
## France	1.48641477	9.49948673	0.30348150	0.003077988	0.0365843
## Irlande	1.77161617	2.14381369	3.86996370	0.772563960	0.3601540
## Italie	0.13999982	22.70063329	2.07230989	0.669310234	0.3555683
## Chypre	3.72491146	7.74592637	1.56305211	66.069928007	2.1029733
## Lettonie	10.84139168	2.61816277	3.02980607	2.353120380	0.1442491
## Lituanie	9.44714871	0.69057274	7.40519453	0.107270234	5.5558504
## Luxembou	4.65864464	0.29782934	5.16377358	1.958431204	24.2439583
## Hongrie	4.07072564	0.98038970	11.89639270	1.105106074	4.0959992
## Malte	0.33153248	0.03268150	7.41323991	5.116384933	5.4425874
## Pays-Bas	5.68991520	2.22102138	0.37086113	2.399108005	4.6487999
## Autriche	3.28478693	0.37616140	0.01322258	0.929071462	4.2743707
## Pologne	14.41767478	4.46637870	2.67761091	2.372186560	0.1666798
## Portugal	0.61887579	2.08652128	0.02299180	0.144360717	3.2589455
## Slovénie	0.11802609	2.51513600	13.79656033	2.974488269	18.9143601
## Slovaqui	15.03314095	0.01569947	1.43576378	0.364914954	0.2689622
## Finlande	1.02711614	0.89018025	1.48760287	0.020437096	1.9174926
## Suede	5.23845724	0.80683859	0.81735071	1.426555135	1.1670699
## Royaume-	4.94647643	3.22712794	9.33712517	3.544565243	2.6901364

Une forte contribution est une contribution  $>4$  ( $1/25 \times 100$ ).

Quatre pays se démarquent avec une très forte contribution  $>9$ : la Lettonie, la Lituanie, la Pologne et la Slovaquie. Viennent ensuite les pays avec une forte contribution: Estonie, Luxembourg, Hongrie, Pays-Bas et Suède.



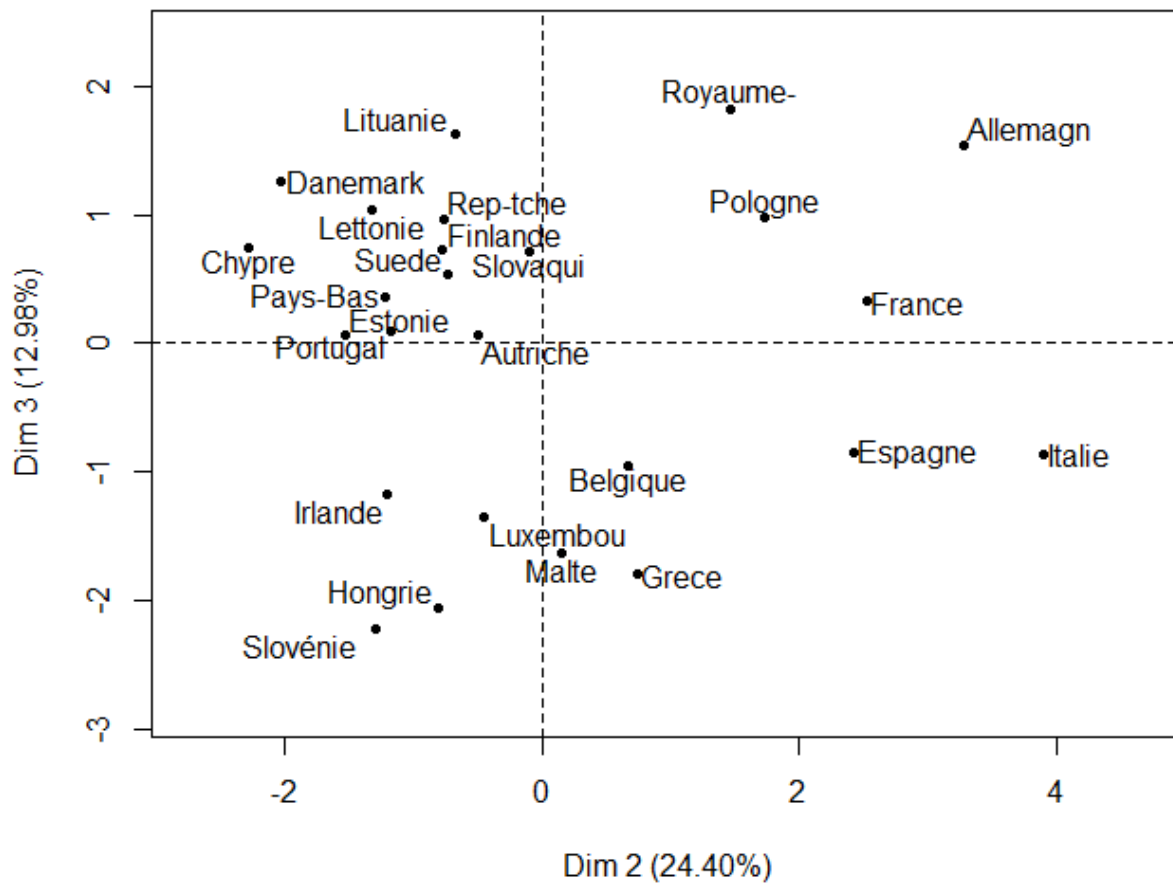
## 6-Commentaire sur l'ACP



### Axe 1

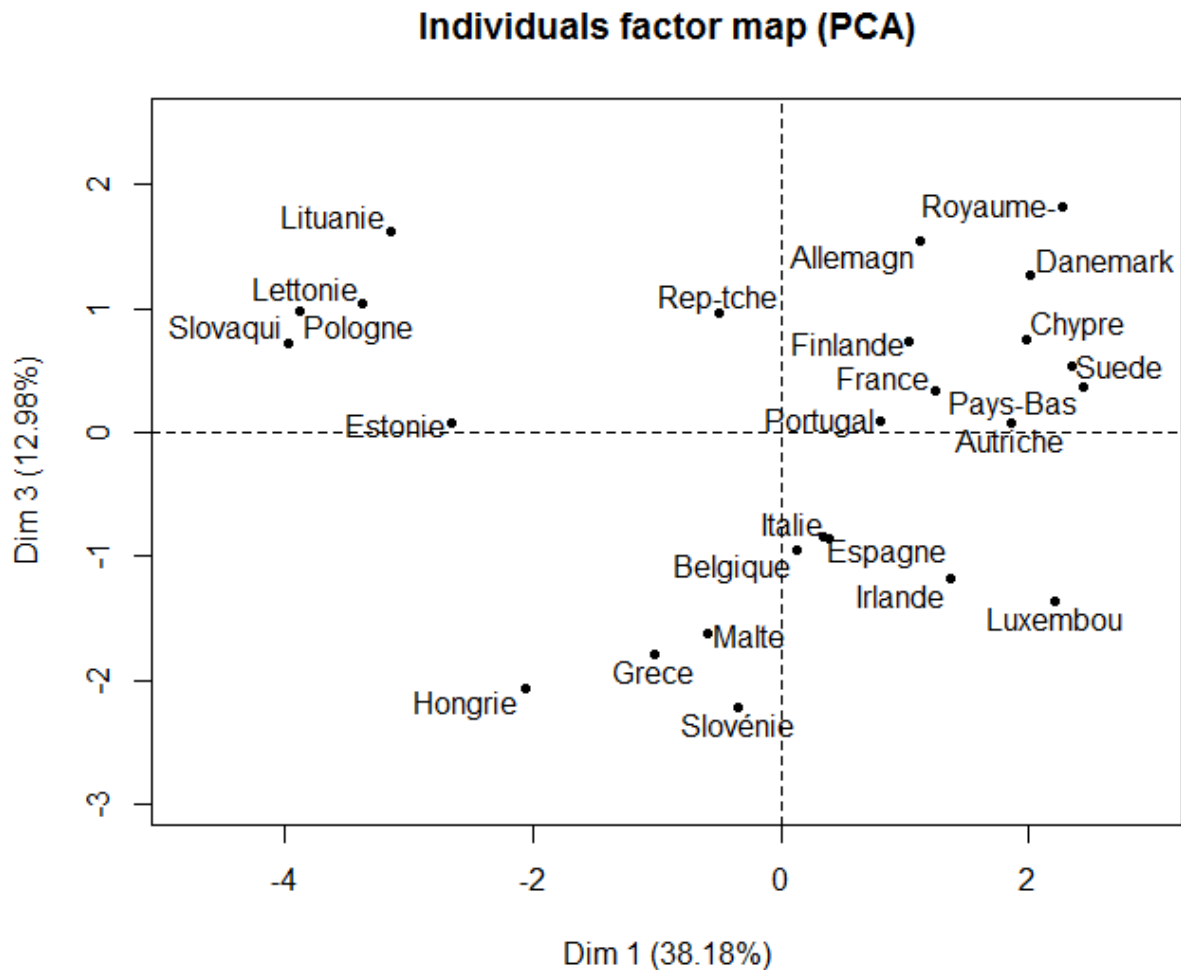
On remarque que deux groupes de pays s'opposent. Celles avec une espérance de vie élevée et un taux de chômage bas (Pays-Bas, Autriche, Suède, Luxembourg, Royaume-Uni) et celles avec une espérance de vie basse et un taux de chômage élevé (Estonie, Lettonie, Hongrie, Pologne et Slovaquie). Dans le 1er cas, les pays font parti de l'Europe des 15. Dans le 2nd cas, les pays font parti de l'Europe des 25.

**Individuals factor map (PCA)**



### Axe 2

De même, deux groupes s'opposent sur cette axe. D'un côté l'Italie, l'Allemagne, la France et l'Espagne. De l'autre, le Danemark et le Portugal.



### Axe 3

Il est difficile de commenter cette axe, les corrélations ne sont pas significatives et il y a beaucoup de variables qui rentrent en compte dans cette Axe. Cette axe sert surtout à ne pas perdre trop d'informations dans notre analyse.

**En gardant ces trois composantes principales, nous avons réussi à garder les 3/4 de l'informations (Inertie de 75.55%) et nous avons réduits nos 11 variables quantitatives à 3. Nous avons pu constaté que les pays baltes, la Pologne et la Slovaquie sont loin derrière en terme économique et d'espérance de vie par rapport à la plupart des pays de l'Europe de 15.**

## ANNEXE

```
library(FactoMineR)
library(Hmisc)
row.names(pays_eu)=pays_eu$PAYS

#Matrice de corrélation
rcorr(as.matrix(pays_eu[,3:13]),type="pearson")

#ACP
resuacp=PCA(pays_eu[,3:13])
resuacp$eig
plot(resuacp$eig[,1])
lines(resuacp$eig[,1])
abline(v=4,col="red")

#Graphique des variables et des individus
resuacp=PCA(pays_eu[,3:13],ncp=3,axes=c(1,3))
resuacp=PCA(pays_eu[,3:13],ncp=3,axes=c(1,2))
resuacp=PCA(pays_eu[,3:13],ncp=3,axes=c(2,3))

#Qualité de représentation
resuacp$ind$cos2

#Contribution des individus
resuacp$ind$contrib
```