

# Analyse des Données

## Projet n°1 : ACP

Killian STEUNOU - Théo DRUILHE

Magistère d'Economiste Statisticien

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Présentation des données . . . . .	1
1.2	Objectifs . . . . .	1
1.3	Plan du rapport . . . . .	1
<b>2</b>	<b>Statistiques</b>	<b>1</b>
2.1	Espérance de vie (Hommes et Femmes) . . . . .	1
2.1.1	Résumé des principaux indicateurs . . . . .	1
2.1.2	Distribution . . . . .	2
2.2	Population et nombre d'abonnés aux services de téléphonie mobile . . . . .	2
2.2.1	Résumé des principaux indicateurs . . . . .	2
2.2.2	Distribution . . . . .	3
2.3	Taux d'activité et taux d'emploi . . . . .	4
2.3.1	Résumé des principaux indicateurs . . . . .	4
2.3.2	Distribution . . . . .	4
2.4	Produit intérieur brut par habitant . . . . .	5
2.4.1	Résumé des principaux indicateurs . . . . .	5
2.4.2	Distribution . . . . .	6
2.5	Taux d'inflation . . . . .	6
2.5.1	Résumé des principaux indicateurs . . . . .	6
2.5.2	Distribution . . . . .	7
2.6	Taux d'emploi et taux de chômage longue durée . . . . .	7
2.6.1	Résumé des principaux indicateurs . . . . .	7
2.6.2	Distribution . . . . .	8
2.7	Nombre de mariages . . . . .	8
2.7.1	Résumé des principaux indicateurs . . . . .	8
2.7.2	Distribution . . . . .	9
2.8	Analyse des corrélations . . . . .	10
<b>3</b>	<b>L'Analyse en Composantes Principales (ACP)</b>	<b>11</b>
3.1	Justification de l'emploi de L'ACP . . . . .	11
3.2	Choix de la dimension . . . . .	11
3.3	Interprétation des composantes retenues . . . . .	12
3.4	Graphiques des individus sur les composantes retenues . . . . .	12
3.4.1	Composante 1 et Composante 2 . . . . .	12
3.4.2	Composante 3 . . . . .	13
3.4.3	Composante 4 . . . . .	14
3.5	Pays ayant fortement contribué à la première composante principale . . . . .	15
<b>4</b>	<b>Conclusion</b>	<b>15</b>

# 1 Introduction

## 1.1 Présentation des données

Nos données concernent 25 pays de l'Union Européenne (source : Eurostat 2002). Ces pays sont caractérisés par différents indicateurs de l'économie, de la démographie et de la société pendant l'année 2002.

Les variables quantitatives considérées sont les suivantes :

- **EVH** : espérance de vie à la naissance pour un homme (en années)
- **EVF** : espérance de vie à la naissance pour une femme (en années)
- **POP** : population (en milliers d'habitants)
- **TACT** : taux d'activité (en pourcentage)
- **PIBH** : produit intérieur brut par habitant (en standards de pouvoir d'achat)
- **TINF** : taux d'inflation (en pourcentage)
- **TEMP** : taux d'emploi (en pourcentage)
- **TCHOM** : taux de chômage (en pourcentage)
- **TCHOMLD** : taux de chômage longue durée (en pourcentage)
- **MARIAG** : nombre de mariages (pour 1000 personnes)
- **TEL** : nombre d'abonnés aux services de téléphonie mobile (en milliers)

## 1.2 Objectifs

Nous disposons de 11 variables quantitatives, ce qui est un nombre assez élevé, rendant difficile l'analyse des données.

L'objectif est donc de résumer l'information contenue dans les variables initiales de notre tableau de données. Pour cela, nous allons réaliser une Analyse en Composantes Principales (ACP), dans le but de créer un petit nombre de variables non corrélées entre elles, et qui conservent au maximum l'information contenue dans les 11 variables initiales.

## 1.3 Plan du rapport

Dans un premier temps nous allons réaliser des statistiques univariées et bivariées. Nous présenterons succinctement les variables les plus intéressantes et nous donnerons la matrice des corrélations.

Dans un second temps, nous allons réaliser l'Analyse en Composantes Principales (ACP).

# 2 Statistiques

Dans cette partie, nous allons étudier chaque variable et interpréter quelques statistiques simples.

## 2.1 Espérance de vie (Hommes et Femmes)

### 2.1.1 Résumé des principaux indicateurs

Indicateurs	EVH	EVF
Minimum	64.80	76
1er Quartile	72.10	78.70
Médiane	75.10	80.70
Moyenne	73.41	80.23
3ème Quartile	75.80	81.50
Maximum	77.70	83.50

TABLE 1 – Principaux indicateurs des espérances de vie

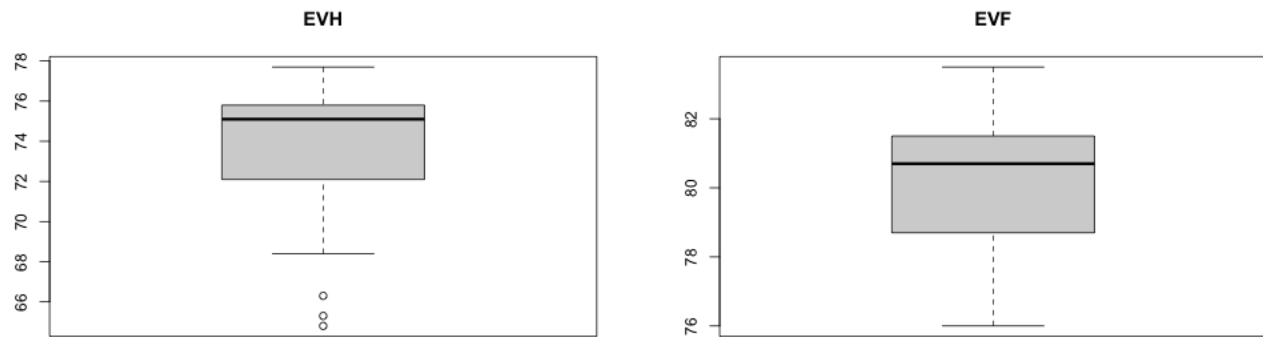


FIGURE 1 – Boîtes à moustache de l'espérance de vie des Hommes et des Femmes

Nous pouvons remarquer que l'espérance de vie des femmes est significativement plus élevée que celle des hommes, en effet, tous ses indicateurs sont plus élevés (voir tableau ci-dessus). En outre, les femmes vivent en moyenne 7 ans de plus que les hommes dans ces 25 pays européens.

### 2.1.2 Distribution

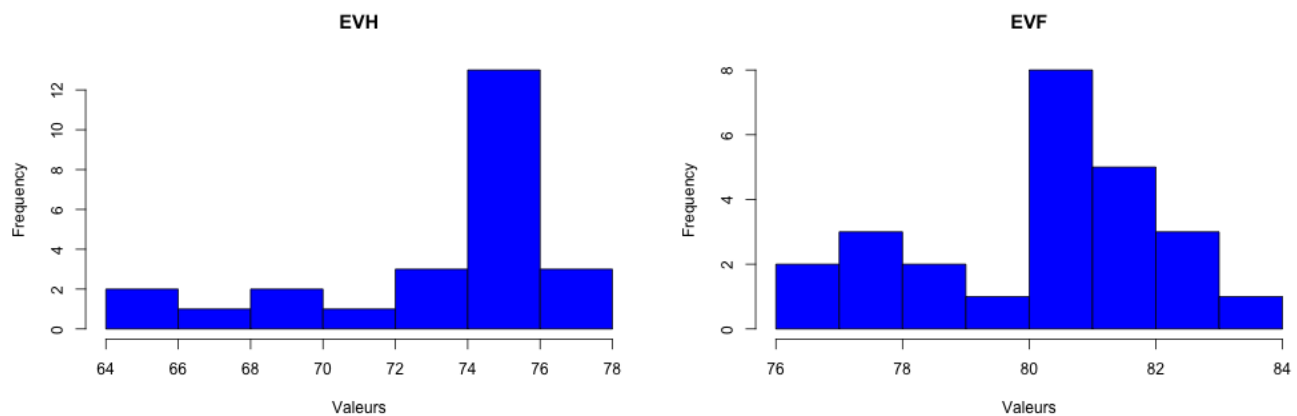


FIGURE 2 – Histogrammes de l'espérance de vie des Hommes et des Femmes

L'observation de la distribution nous confirme que l'espérance de vie des femmes est plus élevée que celle des hommes, avec un mode entre 80 et 81 ans pour les femmes contre un mode entre 74 et 76 ans pour les hommes.

## 2.2 Population et nombre d'abonnés aux services de téléphonie mobile

### 2.2.1 Résumé des principaux indicateurs

Indicateurs	POP	TEL
Minimum	394.6	340
1er Quartile	3475.6	1632
Médiane	8909.1	7949
Moyenne	18120.9	14209
3ème Quartile	16105.3	13898
Maximum	82440.3	59128

TABLE 2 – Principaux indicateurs de la population et du nombre d'abonnés aux services de téléphonie mobile

Nous pouvons remarquer que le nombre médian d'habitants (resp. d'abonnés aux services téléphoniques) est plus de deux fois inférieur au nombre moyen d'habitants (resp. d'abonnés aux services téléphoniques). De plus, 75% des pays comptent moins de 16 millions d'habitants (resp. moins de 13 millions d'abonnés), le maximum valant 82 millions d'habitants (resp. 59 millions d'abonnés) : nous supposons donc l'existence de valeurs extrêmes qui influencent les moyennes pour ces deux variables.

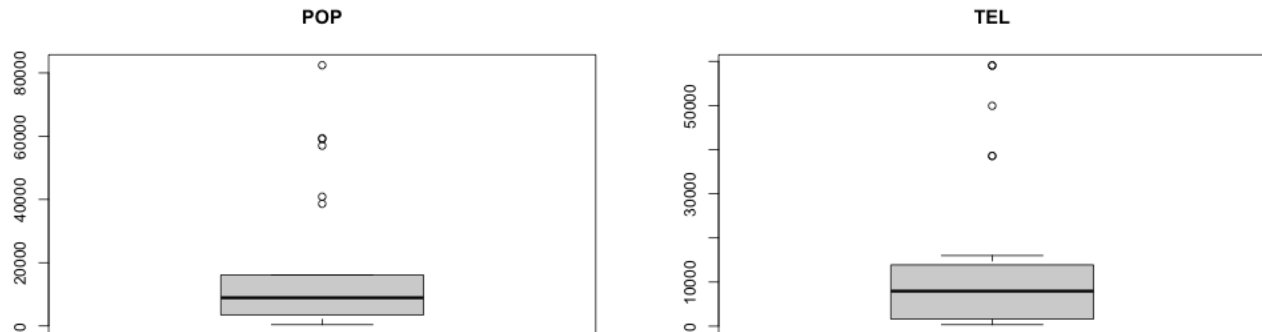


FIGURE 3 – Boîtes à moustache de la population et du nombre d'abonnés aux services de téléphonie mobile

L'observation des boîtes à moustache nous montre qu'il y a en effet quelques valeurs extrêmes bien plus hautes que toutes les autres.

### 2.2.2 Distribution

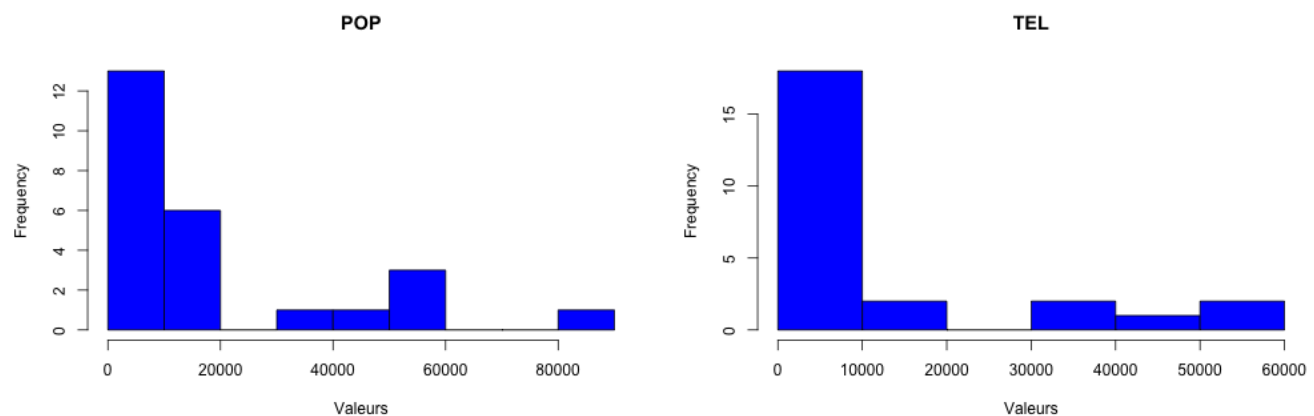


FIGURE 4 – Histogrammes de la population et du nombre d'abonnés au service téléphonique

L'observation de la répartition de la population et du nombre d'abonnés met clairement en lumière la disparité de ces variables : de nombreux pays ont entre 0 et 20 millions d'habitants (resp. entre 0 et 20 millions d'abonnés), seulement 6 d'entre eux ont un nombre d'habitants plus grand que 30 millions (resp. un nombre d'abonnés plus grand que 30 millions), et nous retrouvons bien le nombre maximum d'habitants à plus de 80 millions, et le nombre maximum d'abonnés à presque 60 millions. Enfin, le mode de la population est entre 0 et 10 millions d'habitants, comme celui du nombre d'abonnés aux services de téléphonie mobile.

## 2.3 Taux d'activité et taux d'emploi

### 2.3.1 Résumé des principaux indicateurs

Indicateurs	TACT	TEMP
Minimum	51.5	48.7
1er Quartile	58.4	54.8
Médiane	63.4	58.3
Moyenne	63.5	57.49
3ème Quartile	68.2	61.9
Maximum	75.9	65.6

TABLE 3 – Principaux indicateurs du taux d'activité et du taux d'emploi

Nous pouvons voir que les pays ont un taux d'activité compris entre 50% et 76%, avec une moyenne et une médiane presque confondues autour de 63%. Pour le taux d'emploi également, la moyenne et la médiane sont presque confondues autour de 58%.

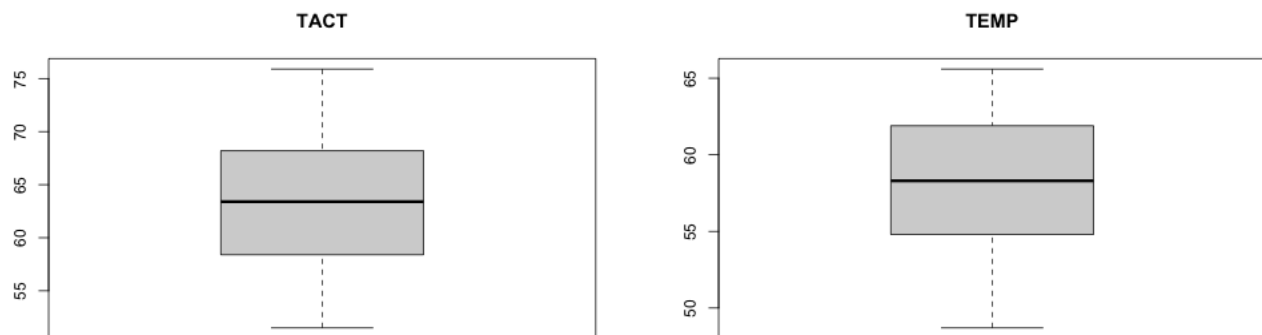


FIGURE 5 – Boîtes à moustache du taux d'activité et du taux d'emploi

Nous pouvons voir sur les boîtes à moustache que les taux d'activité et d'emploi n'ont pas de valeur extrême, et que les valeurs sont assez bien réparties.

### 2.3.2 Distribution

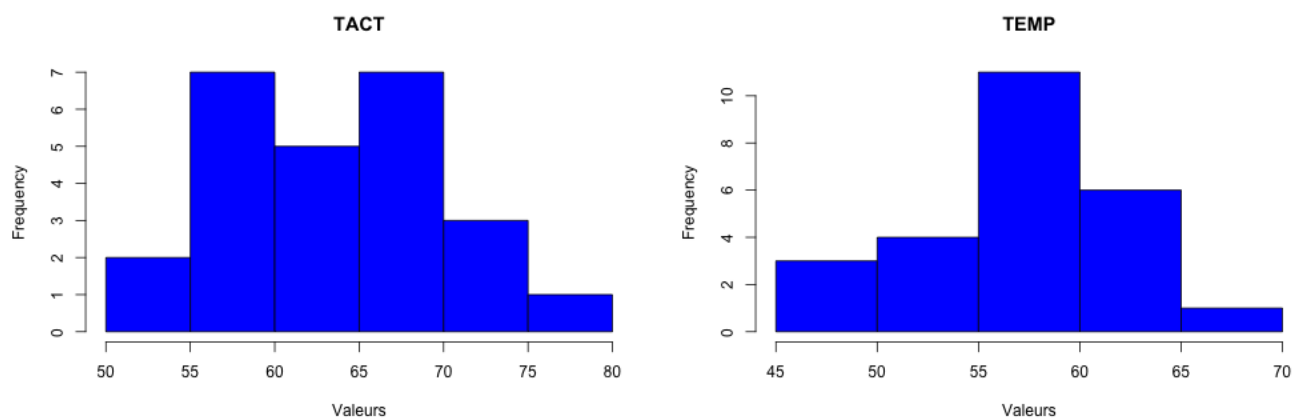


FIGURE 6 – Histogrammes du taux d'activité et du taux d'emploi

Nous remarquons que la plupart des pays ont taux d'activité compris entre 55% et 70%, et qu'un faible nombre de pays a un taux d'activité en dehors de cet intervalle (entre 50% et 55% et entre 70% et 75%). Les modes du taux d'activité sont entre 55% et 60% et entre 65% et 70%.

Le mode du taux d'emploi est entre 55% et 60%.

## 2.4 Produit intérieur brut par habitant

### 2.4.1 Résumé des principaux indicateurs

Indicateurs	PIBH
Minimum	8300
1er Quartile	14300
Médiane	19900
Moyenne	19716
3ème Quartile	24600
Maximum	43900

TABLE 4 – Principaux indicateurs du PIB par habitant

Les PIB par habitant moyen et médian sont quasi confondus, aux alentours des 20 000 Standards de Pouvoir d'Achat (SPA). De plus, 75% des pays ont un PIB par habitant compris entre 8 300 et 24 600 SPA. On observe également que le maximum est de 43 900 SPA, soit bien plus élevé que le 3ème quartile.

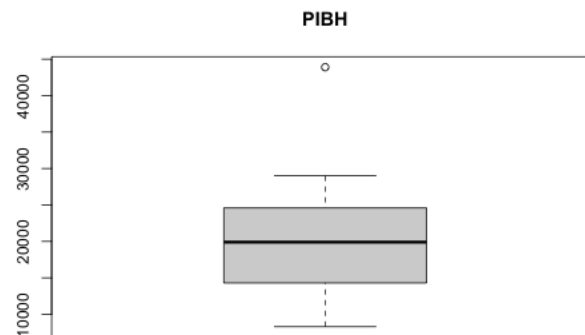


FIGURE 7 – Boîte à moustache du PIB par habitant

La boîte à moustache rend compte de la distribution équilibrée autour de la médiane. Nous remarquons également que le maximum se situe bien au-dessus de la moustache supérieure.

## 2.4.2 Distribution

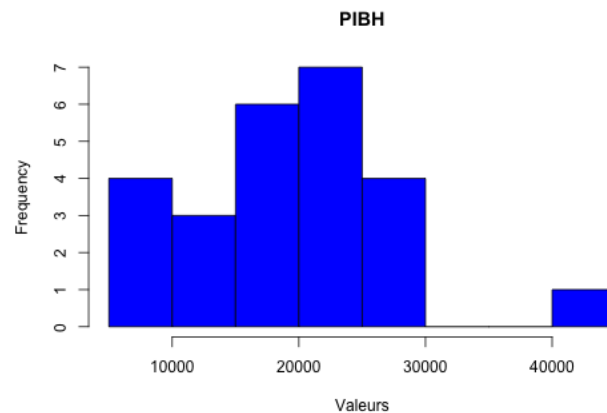


FIGURE 8 – Histogramme du PIB par habitant

Le PIB par habitant de chaque pays semble assez également réparti entre 8 300 SPA et 30 000 SPA, et nous retrouvons la valeur extrême à 43 900 SPA, avec un mode entre 20 000 SPA et 25 000 SPA.

## 2.5 Taux d'inflation

### 2.5.1 Résumé des principaux indicateurs

Indicateurs	TINF
Minimum	0.4
1er Quartile	1.9
Médiane	2.4
Moyenne	2.784
3ème Quartile	3.6
Maximum	7.5

TABLE 5 – Principaux indicateurs du taux d'inflation

Le taux d'inflation varie entre 0,4% et 3,6% pour 75% des pays. Il est en moyenne de 2,784% dans ces pays de l'Union Européenne. Cependant, nous pouvons remarquer que le maximum est très élevé comparé aux autres indicateurs (7,5%).

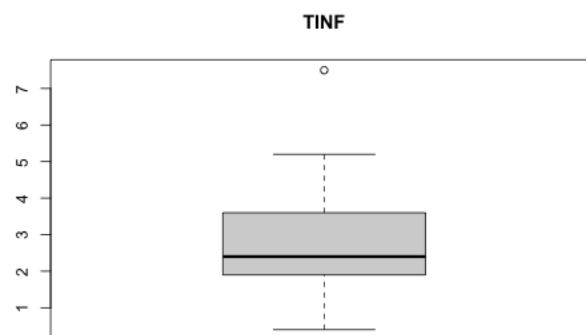


FIGURE 9 – Boîte à moustache du taux d'inflation



Nous voyons sur la boîte à moustache que les valeurs sont pour la plupart regroupées entre 0,4% et  $\approx 5,5\%$ .

## 2.5.2 Distribution

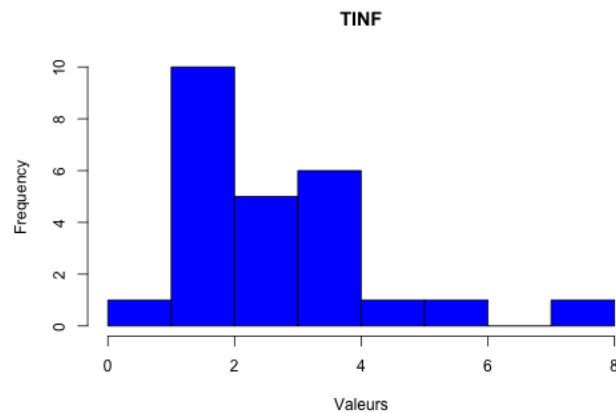


FIGURE 10 – Histogramme du taux d'inflation

Nous remarquons que la majorité des valeurs d'inflation sont comprises entre 1% et 4%. En effet, seulement 4 pays ont un taux d'inflation en dehors de cet intervalle. Le mode est entre 1% et 2%.

## 2.6 Taux d'emploi et taux de chômage longue durée

### 2.6.1 Résumé des principaux indicateurs

Indicateurs	TCHOM	TCHOMLD
Minimum	2.7	0.7
1er Quartile	4.9	1.1
Médiane	7.3	3.3
Moyenne	8.096	3.612
3ème Quartile	9.5	5
Maximum	19.8	12.2

TABLE 6 – Principaux indicateurs des taux de chômage courte durée et longue durée

Nous remarquons dans un premier temps que tous les indicateurs du taux de chômage de court-terme sont plus élevés que ceux du chômage longue durée. Dans les deux cas, la moyenne et la médiane sont assez proches et le maximum est très élevé comparé à ces dernières.

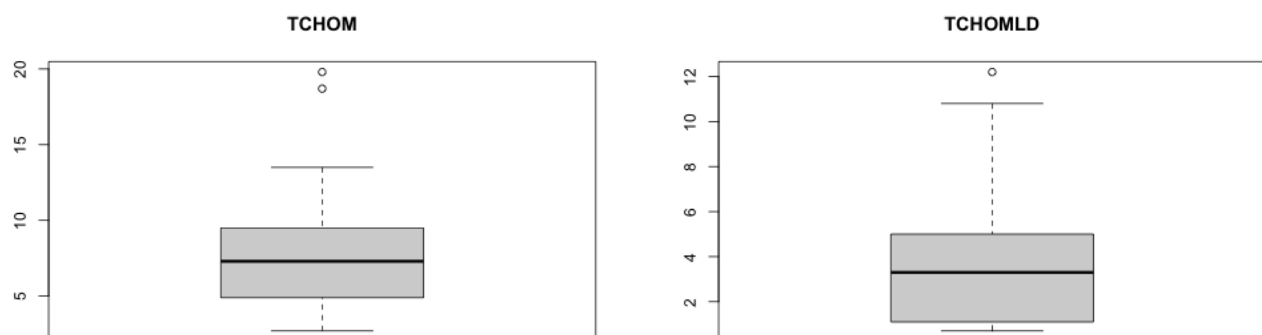


FIGURE 11 – Boîtes à moustache des taux de chômage courte durée et longue durée

Pour le taux de chômage courte durée, la boîte est relativement centrée entre ses deux moustaches. Cependant, pour le taux de chômage de longue durée, la boîte est très proche de sa moustache inférieure. En effet, 75% des données sont inférieures à 5%, alors que 75% des données sont inférieures à  $\approx 10\%$  pour le chômage courte durée. Nous voyons pour les deux taux des valeurs extrêmes.

### 2.6.2 Distribution

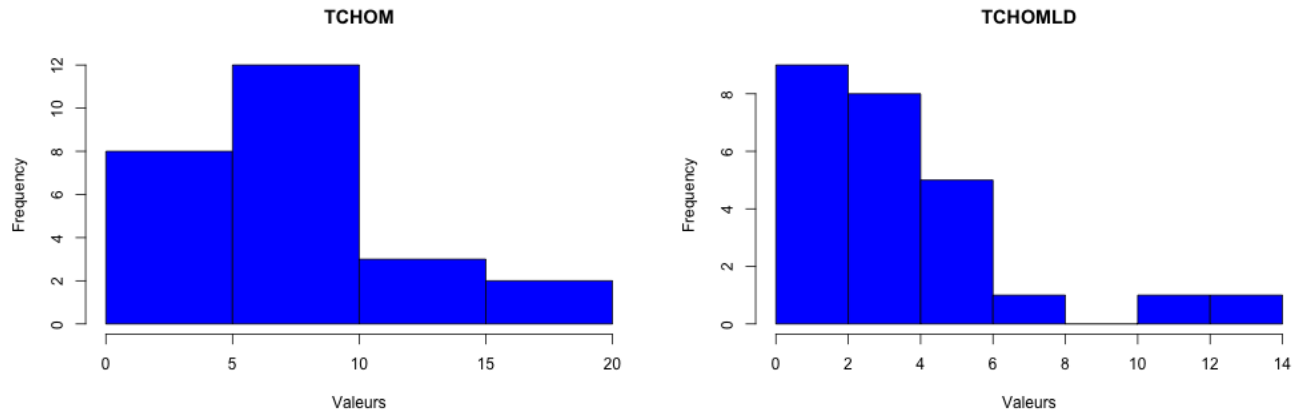


FIGURE 12 – Histogrammes des taux de chômage courte durée et longue durée

Nous voyons sur les histogrammes que la grande majorité des données est comprise entre 0% et 10% pour le taux de chômage, avec un mode entre 5% et 10%, et un petit nombre de valeurs entre 10% et 20%. Pour le taux de chômage longue durée, la majorité des données est comprise entre 0% et 6%, avec un mode entre 0% et 2%, et seulement 3 valeurs supérieures à 6%.

## 2.7 Nombre de mariages

### 2.7.1 Résumé des principaux indicateurs

Indicateurs	MARIAG
Minimum	3.5
1er Quartile	4.5
Médiane	4.7
Moyenne	5.22
3ème Quartile	5.2
Maximum	14.5

TABLE 7 – Principaux indicateurs du nombre de mariages

Nous pouvons observer que le nombre moyen de mariages pour 1 000 personnes est de 5,22, le nombre médian à 4,7. De plus, 75% des valeurs pays comptent entre 3,5 de 5,2 mariages, et le maximum est à 14,5 mariages, ce qui est une valeur très extrême comparé à l'ensemble des autres indicateurs.



FIGURE 13 – Boîte à moustache du nombre de mariages

Nous pouvons voir sur la boîte à moustache que toutes les valeurs sont comprises entre 3.5 et environ 7, sauf pour le maximum avec une valeur de 14,5 mariages pour 1 000 personnes.

### 2.7.2 Distribution

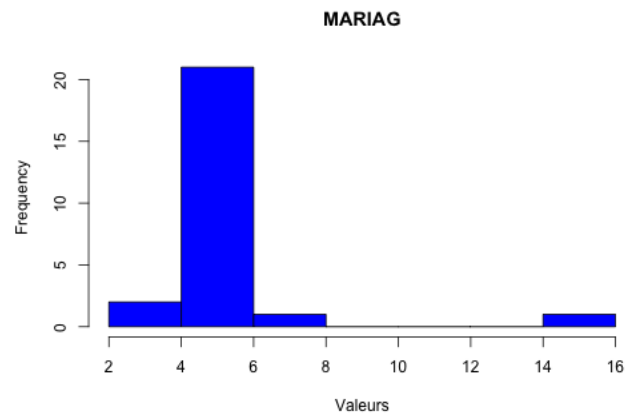


FIGURE 14 – Histogramme du nombre de mariages

L'histogramme nous montre que le mode de cette variable est entre 4 et 6 mariages, contenant plus de 20 pays. Les valeurs sont très peu dispersées avec seulement 4 valeurs en dehors du mode, dont la valeur très extrême.

## 2.8 Analyse des corrélations

Variabes	EVH	EVF	POP	TEMP	PIBH	TINF	TACT	TCHOM	TCHOMLD	MARIAG	TEL
EVH	1.0	0.9	0.31	0.06	0.7	-0.04	0.38	-0.52	-0.55	0.22	0.39
EVF	<b>0.9</b>	1.0	0.41	-0.08	0.65	-0.06	0.22	-0.37	-0.44	0.1	0.49
POP	0.31	0.41	1.0	-0.17	0.14	-0.29	-0.08	0.16	0.1	-0.15	0.96
TEMP	0.06	-0.08	-0.17	1.0	0.13	-0.13	0.83	-0.24	-0.28	0.25	-0.19
PIBH	<b>0.7</b>	<b>0.65</b>	0.14	0.13	1.0	-0.12	0.47	-0.64	-0.64	-0.02	0.2
TINF	-0.04	-0.06	-0.29	-0.13	-0.12	1.0	-0.11	-0.18	-0.08	-0.05	-0.24
TACT	0.38	0.22	-0.08	<b>0.83</b>	0.47	-0.11	1.0	-0.67	-0.71	0.23	-0.02
TCHOM	<b>-0.52</b>	<b>-0.37</b>	0.16	-0.24	<b>-0.64</b>	-0.18	<b>-0.67</b>	1.0	0.96	-0.21	0.04
TCHOMLD	<b>-0.55</b>	<b>-0.44</b>	0.1	-0.28	<b>-0.64</b>	-0.08	<b>-0.71</b>	<b>0.96</b>	1.0	-0.23	0.0
MARIAG	0.22	0.1	-0.15	0.25	-0.02	-0.05	0.23	-0.21	-0.23	1.0	-0.15
TEL	0.39	0.49	<b>0.96</b>	-0.19	0.2	-0.24	-0.02	0.04	0.0	-0.15	1.0

TABLE 8 – Matrice des corrélations

### — Corrélations linéaires positives remarquables :

- **0,96** entre le taux de chômage et le taux de chômage longue durée, ce qui est assez évident ;
- **0,96** entre le nombre d'abonnés au service téléphonique et le nombre d'habitants, en effet plus il y a d'habitants, plus il y a de besoins en service téléphonique et donc d'abonnés ;
- **0,90** entre l'espérance de vie des hommes et celle des femmes ;
- **0,83** entre le taux d'activité et le taux d'emploi ;
- **0,70** (respectivement **0,65**) entre le PIB par habitant et l'espérance de vie des hommes (respectivement l'espérance de vie des femmes) : plus l'on est riche plus l'on vit longtemps.

### — Corrélations linéaires négatives remarquables :

- **-0,71** entre le taux de chômage longue durée et le taux d'activité, et **-0,67** entre le taux de chômage et le taux d'activité (évident : une baisse du taux de chômage implique forcément une hausse du taux d'activité) ;
- **-0,64** entre le taux de chômage et le PIB par habitant, et entre le taux de chômage longue durée et le PIB par habitant (moins il y a de chômage et plus la population est riche) ;
- **-0,55** entre le taux de chômage longue durée et l'espérance de vie des hommes. Nous remarquons également des corrélations similaires entre le taux de chômage et l'espérance de vie des hommes et des femmes (**-0,52** entre TCHOM et EVH, **-0,44** entre TCHOMLD et EVF).

### 3 L'Analyse en Composantes Principales (ACP)

#### 3.1 Justification de l'emploi de L'ACP

L'étude de la matrice des corrélations nous montre l'existence de fortes corrélations entre différentes variables, ce qui justifie l'emploi de l'ACP. Nous allons donc la réaliser sur nos données puis retenir un nombre de composantes adéquat, qui résume bien l'information initiale contenue dans nos données.

#### 3.2 Choix de la dimension

Nous réalisons donc l'ACP sur R (à l'aide du package "FactoMineR"), qui va nous donner toutes les informations nécessaires au choix de la dimension. Nous commençons par étudier l'ébouil des valeurs propres, qui représente le pourcentage de variance totale (c'est à dire l'information) contenue dans chaque composante.

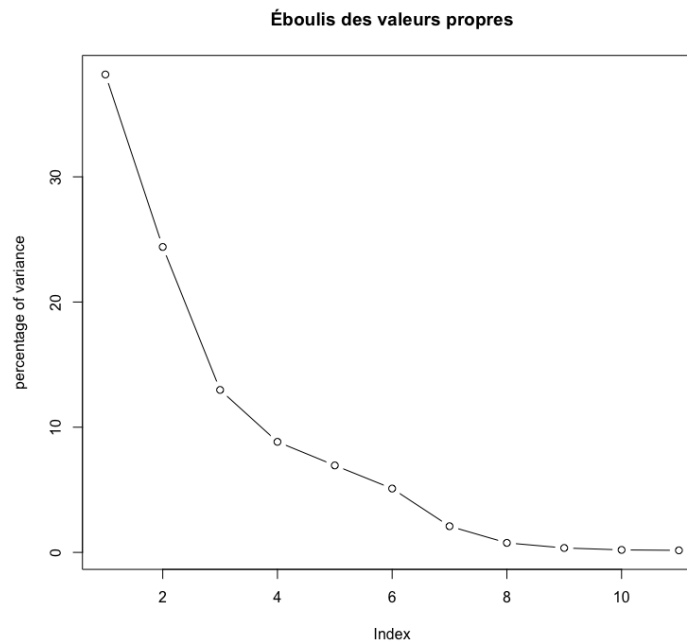


TABLE 9 – Ébouil des valeurs propres

On remarque un coude à 4 ou 5 composantes. Pour choisir le bon nombre, nous allons regarder les corrélations entre chaque variable et les composantes principales.

Variables	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
EVH	<b>0.83516</b>	0.31512	-0.16029	0.23337	-0.05005
EVF	<b>0.73441</b>	0.47893	-0.2059	0.1948	-0.08283
POP	0.18248	<b>0.87463</b>	0.28259	-0.0655	0.28338
TEMP	0.3572	-0.51996	<b>0.66532</b>	-0.12422	0.20995
PIBH	<b>0.81965</b>	0.11026	-0.15474	-0.1844	-0.32565
TINF	-0.07018	-0.28142	<b>-0.68539</b>	-0.05865	0.62708
TACT	<b>0.74457</b>	-0.42826	0.42018	-0.20534	0.1537
TCHOM	<b>-0.82576</b>	0.34535	0.25	0.11864	-0.04979
TCHOMLD	<b>-0.86451</b>	0.29887	0.1589	0.0946	-0.01888
MARIAG	0.25375	-0.29944	0.19341	<b>0.86641</b>	0.12905
TEL	0.27713	<b>0.8642</b>	0.2038	-0.07875	0.29673

TABLE 10 – Corrélations entre variables et composantes

Ce tableau représente les corrélations entre les variables et chaque composante. Plus la corrélation est élevée (en valeur absolue), plus la variable a contribué à la composante. Nous pouvons donc voir que, en valeur absolue, pour

chaque variable, chaque valeur de la composante 5 est inférieure à au moins une autre composante. Donc les 4 premières composantes contiennent suffisamment d'informations.

**Conclusion :** On retient 4 composantes principales.

### 3.3 Interprétation des composantes retenues

En observant le tableau précédent, nous voyons quelles variables ont le plus contribué à chaque composante, et ainsi nous pouvons interpréter.

- **Composante 1** : elle est très corrélée positivement avec EVH, EVF, PIBH, et TACT, et très corrélée négativement avec TCHOM et TCHOMLD. Nous pouvons donc l'interpréter comme une composante du "Bien-être global" de la population des pays.  
En effet, plus l'espérance de vie et la richesse par habitant (PIBH) sont élevés, plus le "bien-être" des habitants est élevé. À l'inverse, plus le taux de chômage est élevé, plus le "bien-être" des habitants est faible ;
- **Composante 2** : elle est très corrélée positivement avec POP et TEL. Nous pouvons donc l'interpréter comme une composante de la "Démographie" des pays ;
- **Composante 3** : elle est très corrélée positivement avec TEMP, et très corrélée négativement avec TINF. Nous pouvons donc l'interpréter comme une composante de l' "Activité Économique" des pays ;
- **Composante 4** : elle est très corrélée positivement avec MARIAG. Nous pouvons donc l'interpréter comme une composante des "Mariages" des pays.

### 3.4 Graphiques des individus sur les composantes retenues

Nous allons à présent interpréter les graphiques des individus sur les composantes retenues. Pour cela, nous avons besoin de regarder les  $\cos^2$  des pays sur chaque composante (c'est à dire la qualité de représentation de chaque pays sur les composantes). Pour chaque composante, nous ne commenterons que les pays bien représentés.

#### 3.4.1 Composante 1 et Composante 2

Le tableau ci-dessous nous montre les individus correctement représentés sur la composante 1 (seuil :  $\cos^2 > 0.5$ ).

Estonie	Lettonie	Lituanie	Pays-Bas	Autriche	Pologne	Slovaquie	Suède
0.63521	0.72848	0.68499	0.64328	0.6874	0.72635	0.7727	0.66803

TABLE 11 – Pays bien représentés sur la composante 1

Nous ne commenterons donc que ces pays pour la première composante ("Bien-être global de la population").

Le tableau ci-dessous nous montre les individus correctement représentés sur la composante 2 (seuil :  $\cos^2 > 0.25$ ).

Danemark	Allemagne	Espagne	France	Irlande	Italie	Portugal
0.40718	0.67949	0.71467	0.76029	0.25572	0.90716	0.45126

TABLE 12 – Pays bien représentés sur la composante 2

Nous ne commenterons donc que ces pays pour la deuxième composante ("Démographie").

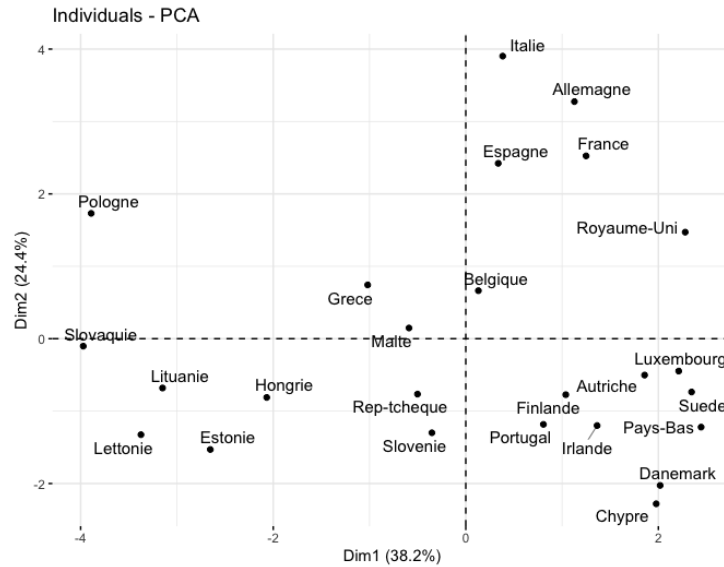


FIGURE 15 – Graphique des individus sur les composantes 1 et 2

Parmi les individus bien représentés sur la composante 1, nous pouvons ainsi voir à l'aide du graphique que la population de la Suède, des Pays-bas et de l'Autriche a un "bien-être global" assez élevé, tandis que la population de la Pologne, la Slovaquie, et les Pays Baltes (Lituanie, Lettonie, Estonie), qui sont tous des pays de l'est de l'Europe, a un "bien-être global" assez faible.

Concernant la "Démographie", qui correspond plus ou moins au nombre d'habitants (car le coefficient de corrélation linéaire entre population et nombre d'abonnés au service téléphonique, c'est-à-dire les 2 variables le plus corrélées à la composante 2, est presque de 1) nous pouvons voir que la France, le Danemark, l'Italie et l'Espagne, sont des pays ayant une population nombreuse, alors que le Portugal et l'Irlande ont une population moindre.

### 3.4.2 Composante 3

Le tableau ci-dessous nous montre les individus correctement représentés sur la composante 3 (seuil :  $\cos^2 > 0.15$ ).

Belgique	Rep-tcheque	Danemark	Allemagne	Grece	Irlande	Lituanie	Hongrie	Malte	Slovenie	Royaume-Uni
0.22482	0.362	0.15909	0.15029	0.54243	0.24554	0.18253	0.30794	0.42705	0.40324	0.26149

TABLE 13 – Pays bien représentés sur la composante 3

Nous ne commenterons donc que ces pays pour la troisième composante ("Activité Économique").

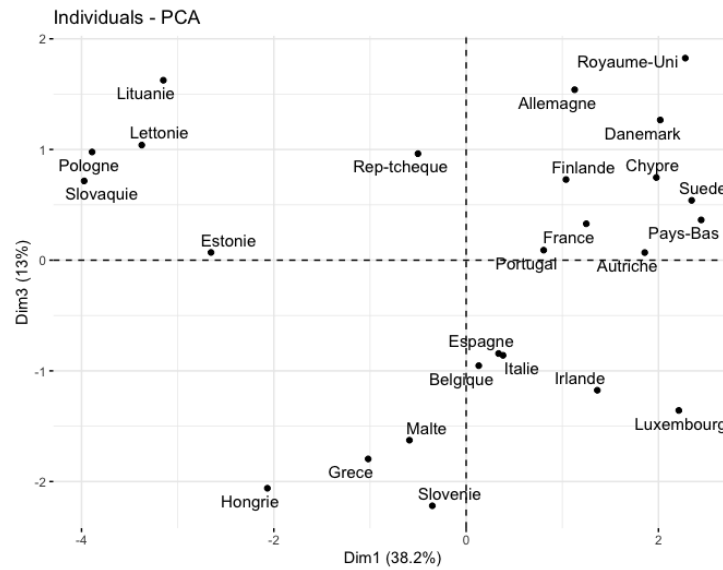


FIGURE 16 – Graphique des individus sur les composantes 1 et 3

Nous pouvons observer que la Lituanie, le Royaume-Uni, le Danemark, l'Allemagne et la République tchèque ont une "Activité économique" élevée (c'est-à-dire un taux d'emploi élevé et une inflation faible) tandis que la Slove, la Grèce, la Hongrie et Malte ont une "Activité économique" faible. L'Irlande et la Belgique ont une "Activité économique" assez faible, mais un peu plus importante que les 4 pays précédemment cités.

### 3.4.3 Composante 4

Le tableau ci-dessous nous montre les individus correctement représentés sur la composante 4 (seuil :  $\cos^2 > 0.1$ ).

Grece	Chypre	Malte
0.1154	0.60083	0.2006

TABLE 14 – Pays bien représentés sur la composante 4

Nous ne commenterons donc que ces pays pour la quatrième composante ("Mariages").

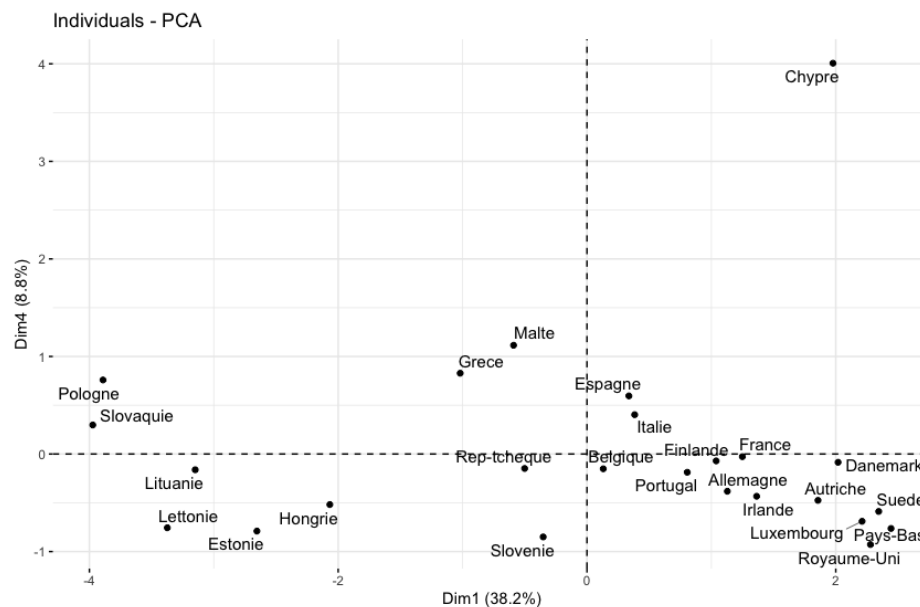


FIGURE 17 – Graphique des individus sur les composantes 1 et 4



Nous remarquons immédiatement Chypre qui a un nombre très élevé de mariages pour 1000 habitants comparé au reste des pays qui ont tous un nombre bien plus faible de mariages. La Grèce et Malte sortent un peu du lot avec légèrement plus de mariages. Nous pouvons noter que les 3 pays cités sont des îles méditerranéennes ou des pays méditerranéens.

### 3.5 Pays ayant fortement contribué à la première composante principale

Le tableau suivant donne les contributions à la première composante de chaque pays.

Nous avons 25 pays, ce qui nous donne le seuil de contribution pour considérer qu'un pays a fortement contribué à  $\frac{100}{25} = 4$ .

Pays	Contribution
Belgique	0.01631
Rep-tcheque	0.23931
Danemark	3.88106
Allemagne	1.21239
Estonie	<b>6.7065</b>
Grèce	0.98948
Espagne	0.10808
France	1.48641
Irlande	1.77162
Italie	0.14
Chypre	3.72491
Lettonie	<b>10.84139</b>
Lituanie	<b>9.44715</b>
Luxembourg	<b>4.65864</b>
Hongrie	<b>4.07073</b>
Malte	0.33153
Pays-Bas	<b>5.68992</b>
Autriche	3.28479
Pologne	<b>14.41767</b>
Portugal	0.61888
Slovenie	0.11803
Slovaquie	<b>15.03314</b>
Finlande	1.02712
Suede	<b>5.23846</b>
Royaume-Uni	<b>4.94648</b>

TABLE 15 – Contribution de chaque pays à la composante 1

Nous trouvons ainsi 10 pays ayant fortement contribué à la première composante principale : l'Estonie, la Lettonie, la Lituanie, le Luxembourg, la Hongrie, les Pays-Bas, la Pologne, la Slovaquie, la Suède et le Royaume-Uni.

## 4 Conclusion

Grâce à la méthode d'Analyse en Composante Principale, nous avons réussi à réduire le nombre de variables du tableau de données, passant de 11 variables initiales à 4 nouvelles variables, en conservant le maximum d'information : la première résume le "bien-être global" de la population de chaque pays, la deuxième la "démographie" de chaque pays, la troisième l' "activité économique" de chaque pays, et la quatrième les mariages. Le choix de conserver cette quatrième composante principale, bien qu'elle ne soit corrélée qu'à une seule des variables initiales, est discutable, mais nous avons jugé que ne pas la conserver nous aurait fait perdre trop d'information sur la variable "Mariages".

Finalement, grâce à l'ACP, nous avons pu observer plus facilement et plus rapidement des données intéressantes sur les 25 pays de l'Union européenne dont nous disposions.