Analyse En Composantes Principales

Walid Keddad

Exercice 1 : Arrestations aux Etats-Unis

Le fichier de données USarrests, accessible en R via la commande data ("USArrests"), contient des statistiques collectées en 1973 sur les taux d'arrestation pour 100000 habitants pour agression, meurtre ou viol dans chacun des n = 50 états des USA. Une quatrième variable indique le pourcentage de résidents dans des zones urbaines pour chaque état

Chargement des données :

```
knitr::opts_chunk$set(warning=FALSE, message=FALSE)
data = USArrests
head(data)
```

```
##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58 21.2
## Alaska       10.0     263       48 44.5
## Arizona       8.1     294       80 31.0
## Arkansas      8.8     190       50 19.5
## California    9.0     276       91 40.6
## Colorado      7.9     204       78 38.7
```

```
dim(data)
```

```
## [1] 50  4
```

Faisant une analyse descriptive des variables :

```
summary(data)
```

```
##      Murder          Assault         UrbanPop          Rape      
##  Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30  
##  1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07  
##  Median : 7.250   Median :159.0   Median :66.00   Median :20.10  
##  Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23  
##  3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18  
##  Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00  
```

On va effectuer en premier temps une **ACP non normée** :

```
acp = prcomp(data,scale=FALSE)
summary(acp)
```

```
## Importance of components:
##                            PC1      PC2    PC3     PC4
## Standard deviation     83.7324 14.21240 6.4894 2.48279
## Proportion of Variance  0.9655  0.02782 0.0058 0.00085
## Cumulative Proportion   0.9655  0.99335 0.9991 1.00000
```

on va utiliser la biblithéque **factoextra** pour voir le résultat de l'acp :

```
library(factoextra)
var <- get_pca_var(acp)
var
```

```
## Principal Component Analysis Results for variables
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates for the variables"
## 2 "$cor"     "Correlations between variables and dimensions"
## 3 "$cos2"    "Cos2 for the variables"
## 4 "$contrib" "contributions of the variables"
```

Interpretation des axes

Variables :

```
var <- get_pca_var(acp)
var
```

```
## Principal Component Analysis Results for variables
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates for the variables"
## 2 "$cor"     "Correlations between variables and dimensions"
## 3 "$cos2"    "Cos2 for the variables"
## 4 "$contrib" "contributions of the variables"
```

```
var$contrib
```

```
##               Dim.1      Dim.2      Dim.3      Dim.4
## Murder    0.1739250  0.2008981  0.6382517 98.9869251
## Assault  99.0465399  0.3452741  0.4565669  0.1516191
## UrbanPop  0.2147001 95.4250536  4.0218813  0.3383649
## Rape      0.5648349  4.0287742 94.8833000  0.5230908
```

La variable qui contribue le plus à la formation de l'axe 1 est : **Assault** 99.04%

La variable qui contribue le plus à la formation de l'axe 2 est : **UrbanPop** 95.42%

Individues :

```
ind = get_pca_ind(acp)
ind
```

```
## Principal Component Analysis Results for individuals
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates for the individuals"
## 2 "$cos2"    "Cos2 for the individuals"
## 3 "$contrib" "contributions of the individuals"
```

```
ind$contrib
```

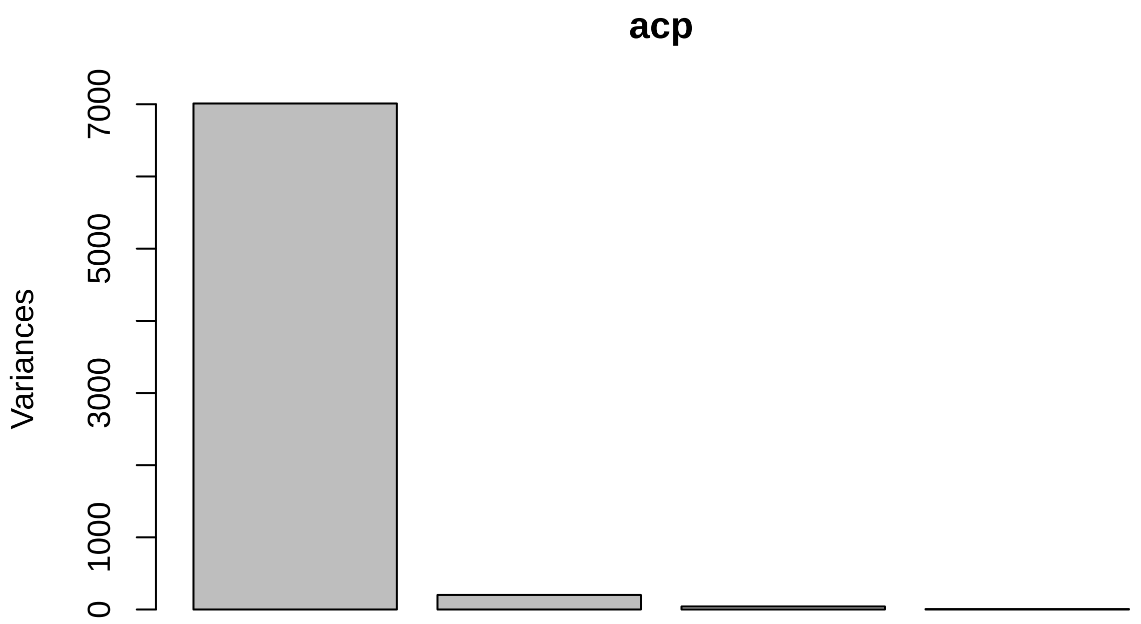
```
##                      Dim.1        Dim.2        Dim.3       Dim.4
## Alabama        1.197903759 1.2976418443 2.956209e-01  1.88116656
## Alaska         2.458078547 3.2019648471 1.923788e+01  5.43820626
## Arizona        4.391005602 0.7720689580 1.352317e-01  6.14984356
## Arkansas       0.095949619 2.7626851500 2.098161e-03  0.08806732
## California     3.291827649 5.0215119680 2.161194e+00  2.56523334
## Colorado       0.348965784 1.8637039554 7.160924e+00  0.96149219
## Connecticut    1.057538260 1.6560065045 3.367514e+00  0.15893694
## Delaware       1.270277225 0.0181469083 6.043790e+00  4.50917869
## Florida        7.789261053 0.3898338951 4.268363e-01  0.50507625
## Georgia        0.468713058 0.5262336818 6.187548e-01 17.49752661
## Hawaii         4.353421604 5.8424265482 6.587795e-01  3.91310880
## Idaho          0.765336041 0.8878115518 1.097339e-01  3.63643961
## Illinois       1.779959840 1.6469351944 1.643820e+00  0.04385280
## Indiana        0.944817808 0.0802131582 6.636411e-01  0.88270967
## Iowa           3.811178746 0.1105966181 2.031477e-02  0.24529302
## Kansas         0.887873052 0.0986980472 7.016220e-03  0.13826086
## Kentucky       1.110140486 1.1279507825 2.376752e-01  4.87490383
## Louisiana      1.747906772 0.1826436974 6.958757e-01  6.52222533
## Maine          2.272829412 1.3066847644 1.045703e+00  1.45299202
## Maryland       4.771362131 0.2482308085 2.616421e-01  1.20645227
## Massachusetts  0.129010803 3.7457798163 2.676501e+00  0.34743916
## Michigan       2.082967849 0.3451990828 1.984568e+00  0.08080430
## Minnesota      2.793294853 0.2687224148 2.052320e-06  0.17379944
## Mississippi    2.152019195 7.4489765851 1.188924e+00  4.88381523
## Missouri       0.018194198 0.2755786535 1.436926e+00  0.14976425
## Montana        1.113718649 0.8955749264 1.605005e-01  0.01962536
## Nebraska       1.361932458 0.0004416376 1.040274e-02  0.13986446
## Nevada         1.994327763 2.2582631479 1.198930e+01  0.03623708
## New Hampshire  3.757987534 0.2219494064 2.473979e-01  0.28419650
## New Jersey     0.033369846 5.3005612023 1.891027e+00  0.84354897
## New Mexico     3.763936281 0.0011208412 2.428411e-01  0.61900363
## New York       2.026929368 2.5107154612 1.058603e+00  0.25816572
## North Carolina 7.702876158 9.5746141114 6.496872e+00  1.44612484
## North Dakota   4.636959310 2.5777161839 8.172829e-02  1.71778828
## Ohio           0.715632181 1.4929455990 1.304475e-01  1.33586826
## Oklahoma       0.110636533 0.1124575478 9.751884e-03  0.01055266
## Oregon         0.035465927 0.1479905724 3.139037e+00  2.75506825
## Pennsylvania   1.193728864 0.7863233982 4.882828e-01  1.14057171
## Rhode Island   0.002678013 3.3427281949 1.449453e+01  1.72869888
## South Carolina 3.283137685 5.4848483787 1.962477e-01  0.50837317
## South Dakota   2.115032315 2.7277160686 8.204635e-02  0.50881281
## Tennessee      0.087423839 0.4191794654 1.767239e+00  4.99292104
## Texas          0.279309160 1.6694601113 7.350324e-03  5.83837758
## Utah           0.710685046 3.0839587031 1.518568e-01  1.13179220
## Vermont        4.436868928 7.3867198390 1.095475e+00  1.30428521
## Virginia       0.062631058 0.0304136187 5.190062e-02  0.44706271
## Washington     0.179371673 0.9838031609 1.085621e+00  2.34965359
## West Virginia  2.390610500 5.2163812660 7.674211e-03  0.17617380
## Wisconsin      3.983858420 0.3003423273 3.491236e-01  0.01363141
## Wyoming        0.031059144 0.3475293948 6.837768e-01  0.08701362
```

Les individus qui contribuent le plus à la formation de l'axe 1 sont : **North Carolina** et **Florida** 7%

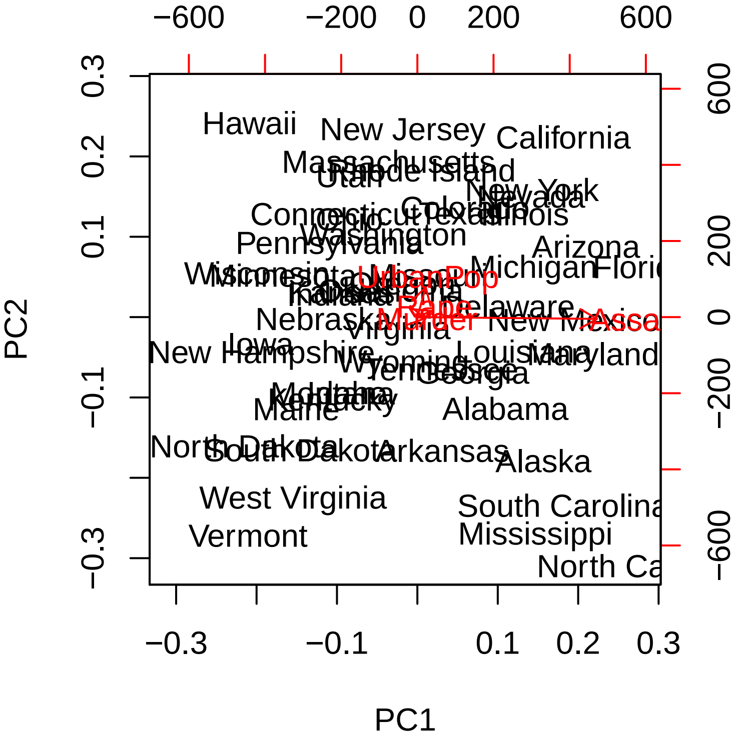
Les individus qui contribuent le plus à la formation de l'axe 2 sont : **North Carolina** 9% et **Mississippi , Verment** 7%

Une représention graphique donne le résultat suivant :

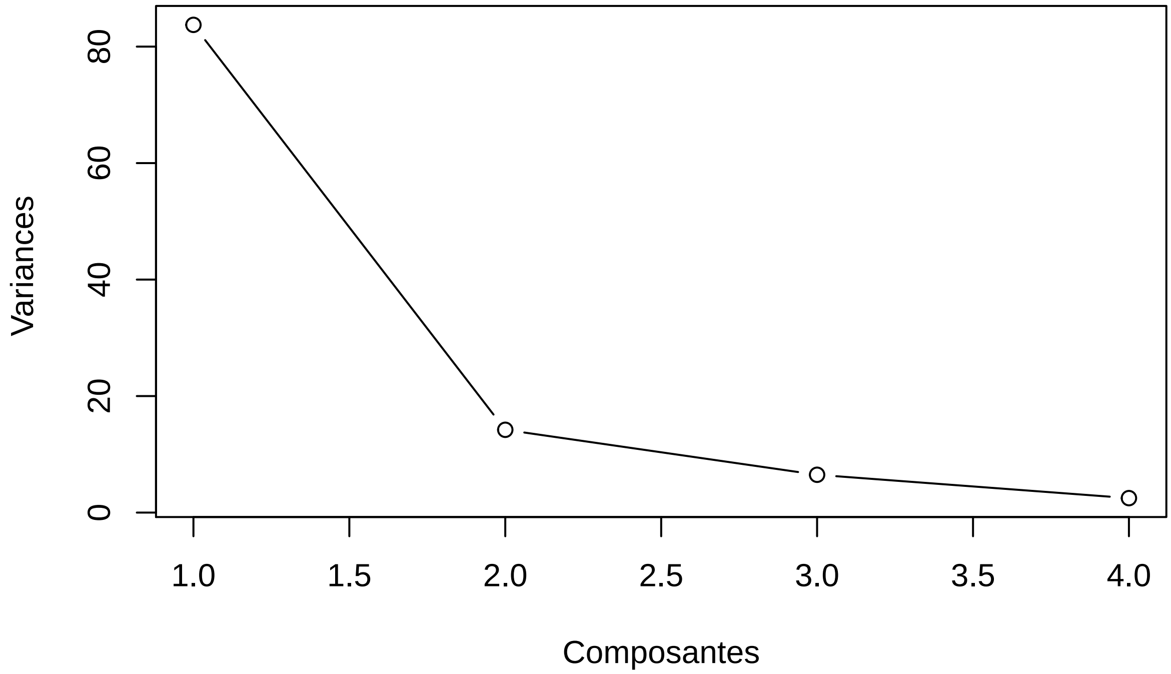
```
plot(acp)
```



```
biplot(acp , xlabs = row.names(USArrests))
```



```
plot(1:4,acp$sdev,ylab="Variances ",xlab="Composantes",type = "b")
```



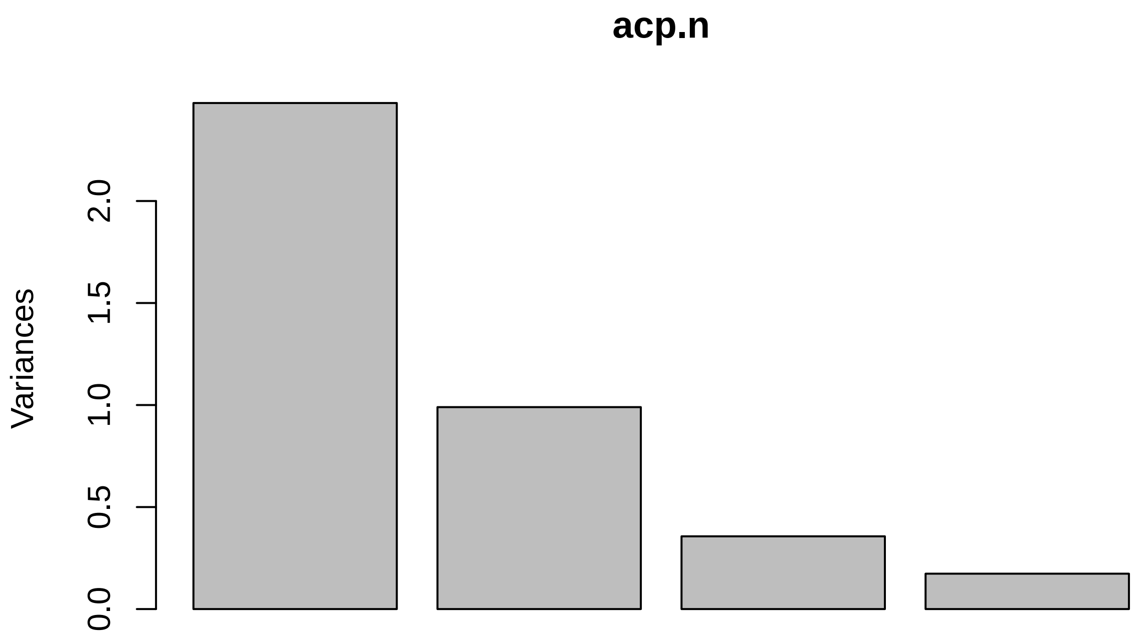
On va maintenant effectuer une **ACP normée** :

```
acp.n = prcomp(data,scale=TRUE)
summary(acp.n)
```

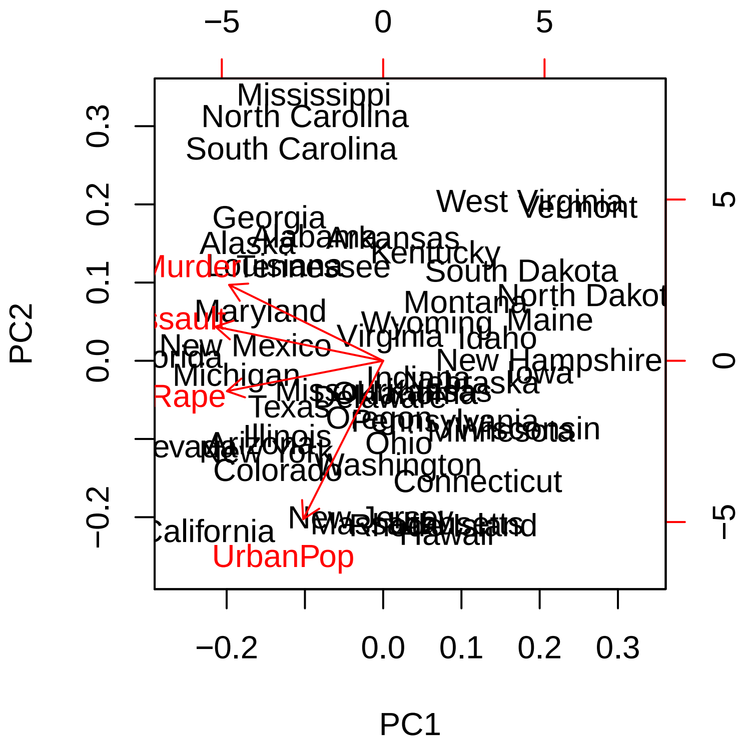
```
## Importance of components:
##                           PC1    PC2     PC3     PC4
## Standard deviation     1.5749 0.9949 0.59713 0.41645
## Proportion of Variance 0.6201 0.2474 0.08914 0.04336
## Cumulative Proportion  0.6201 0.8675 0.95664 1.00000
```

La représentation graphique de l'ACP normée donne le résultat suivant:

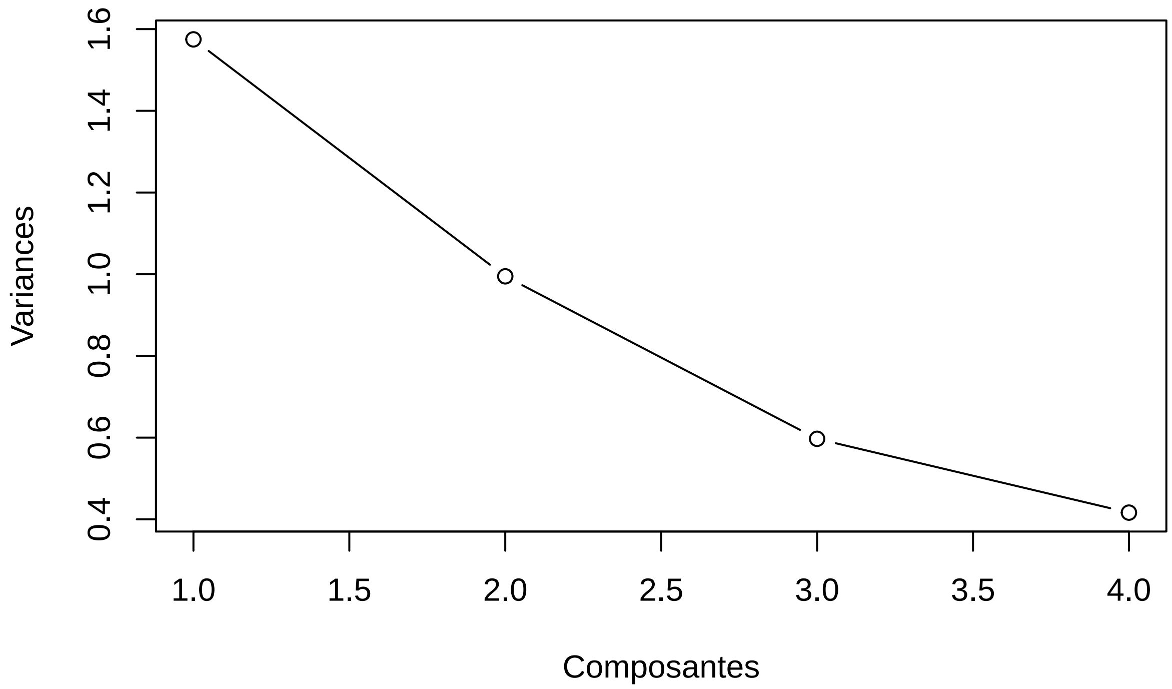
```
plot(acp.n)
```



```
biplot(acp.n , xlabs = row.names(USArrests))
```



```r
plot(1:4,acp.n$sdev,ylab="Variances",xlab="Composantes",type = "b")
```



On constate qu'après la **normalisation** le **2ème** composant participe plus que dans la version non normée

Exercice 2 : Analyse textuelle d'un corpus d'emails

Dans cette partie, il est proposé d'analyser un jeu de données textuelles dans le but de tenter de déterminer les caractéristiques de spams. Une telle analyse est classiquement basée sur la fréquence d'une sélection de mots dans un ensemble d'apprentissage constitué de courriels qui appartiennent à 2 catégories possibles : spam ou non-spam. Les données analysées dans ce TP sont publiques, et elles peuvent servir de "benchmark" pour la comparaison de méthodes d'apprentissage automatique (UCI Machine Learning Repository. Irvine, CA : University of California, School of Information and Computer Science. http://archive.ics.uci.edu/ml): Il a été constitué un ´échantillon de messages électroniques dans chacun desquels a été évalué le nombre d'occurrences d'une sélection de mots et caractères. Les variables considérées sont des ratios qui correspondent au nombre d'occurrences d'un mot spécifique sur le nombre total de mots, ou nombre d'occurrences d'un caractère sur le nombre de caractères du message. Il a également considéré trois variables prenant en compte la casse (majuscule / minuscule) des caractères et une dernière variable qualitative binaire indiquant le type de chaque message : spam ou Nsp.

Chargement des données :

```
setwd("C:/Users/W  7/Desktop/Master 2/AA2/TP2_ACP")
spam <- read.csv("Data\\data_spam.csv",header=FALSE,sep=";")
nom_spam <- read.csv("Data\\names_spam.csv",header=FALSE,sep=";")
names(spam) <- sapply((1:nrow(nom_spam)),function(i) toString(nom_spam[i,1]))
spam$y = as.factor(spam$y)

head(spam)
```

```
##   word_freq_make word_freq_address word_freq_all word_freq_3d
## 1           0.00              0.64          0.64            0
## 2           0.21              0.28          0.50            0
## 3           0.06              0.00          0.71            0
## 4           0.00              0.00          0.00            0
## 5           0.00              0.00          0.00            0
## 6           0.00              0.00          0.00            0
##   word_freq_our word_freq_over word_freq_remove word_freq_internet
## 1          0.32           0.00             0.00               0.00
## 2          0.14           0.28             0.21               0.07
## 3          1.23           0.19             0.19               0.12
## 4          0.63           0.00             0.31               0.63
## 5          0.63           0.00             0.31               0.63
## 6          1.85           0.00             0.00               1.85
##   word_freq_order word_freq_mail word_freq_receive word_freq_will
## 1            0.00           0.00              0.00           0.64
## 2            0.00           0.94              0.21           0.79
## 3            0.64           0.25              0.38           0.45
## 4            0.31           0.63              0.31           0.31
## 5            0.31           0.63              0.31           0.31
## 6            0.00           0.00              0.00           0.00
##   word_freq_people word_freq_report word_freq_addresses word_freq_free
## 1             0.00             0.00                0.00           0.32
## 2             0.65             0.21                0.14           0.14
## 3             0.12             0.00                1.75           0.06
## 4             0.31             0.00                0.00           0.31
## 5             0.31             0.00                0.00           0.31
## 6             0.00             0.00                0.00           0.00
##   word_freq_business word_freq_email word_freq_you word_freq_credit
## 1               0.00            1.29          1.93             0.00
## 2               0.07            0.28          3.47             0.00
## 3               0.06            1.03          1.36             0.32
## 4               0.00            0.00          3.18             0.00
## 5               0.00            0.00          3.18             0.00
## 6               0.00            0.00          0.00             0.00
##   word_freq_your word_freq_font word_freq_000 word_freq_money word_freq_hp
## 1           0.96              0          0.00            0.00            0
## 2           1.59              0          0.43            0.43            0
## 3           0.51              0          1.16            0.06            0
## 4           0.31              0          0.00            0.00            0
## 5           0.31              0          0.00            0.00            0
## 6           0.00              0          0.00            0.00            0
##   word_freq_hpl word_freq_george word_freq_650 word_freq_lab
## 1             0                0             0             0
## 2             0                0             0             0
## 3             0                0             0             0
## 4             0                0             0             0
## 5             0                0             0             0
## 6             0                0             0             0
##   word_freq_labs word_freq_telnet word_freq_857 word_freq_data
## 1              0                0             0              0
## 2              0                0             0              0
## 3              0                0             0              0
## 4              0                0             0              0
## 5              0                0             0              0
## 6              0                0             0              0
##   word_freq_415 word_freq_85 word_freq_technology word_freq_1999
## 1             0            0                    0           0.00
## 2             0            0                    0           0.07
## 3             0            0                    0           0.00
## 4             0            0                    0           0.00
## 5             0            0                    0           0.00
## 6             0            0                    0           0.00
##   word_freq_parts word_freq_pm word_freq_direct word_freq_cs
## 1               0            0             0.00            0
## 2               0            0             0.00            0
## 3               0            0             0.06            0
## 4               0            0             0.00            0
## 5               0            0             0.00            0
## 6               0            0             0.00            0
##   word_freq_meeting word_freq_original word_freq_project word_freq_re
## 1                 0               0.00                 0         0.00
## 2                 0               0.00                 0         0.00
## 3                 0               0.12                 0         0.06
## 4                 0               0.00                 0         0.00
## 5                 0               0.00                 0         0.00
## 6                 0               0.00                 0         0.00
##   word_freq_edu word_freq_table word_freq_conference char_freq_;
## 1          0.00               0                    0        0.00
## 2          0.00               0                    0        0.00
## 3          0.06               0                    0        0.01
## 4          0.00               0                    0        0.00
## 5          0.00               0                    0        0.00
## 6          0.00               0                    0        0.00
##   char_freq_( char_freq_[ char_freq_! char_freq_$ char_freq_#
## 1       0.000           0       0.778       0.000       0.000
## 2       0.132           0       0.372       0.180       0.048
## 3       0.143           0       0.276       0.184       0.010
## 4       0.137           0       0.137       0.000       0.000
## 5       0.135           0       0.135       0.000       0.000
## 6       0.223           0       0.000       0.000       0.000
##   capital_run_length_average capital_run_length_longest
## 1                      3.756                         61
## 2                      5.114                        101
## 3                      9.821                        485
## 4                      3.537                         40
## 5                      3.537                         40
## 6                      3.000                         15
##   capital_run_length_total y
## 1                      278 1
## 2                     1028 1
## 3                     2259 1
## 4                      191 1
## 5                      191 1
## 6                       54 1
```

```r
dim(spam)
```

```
## [1] 4601   58
```

L’analyse descriptive des variables donne le résultat suivant :

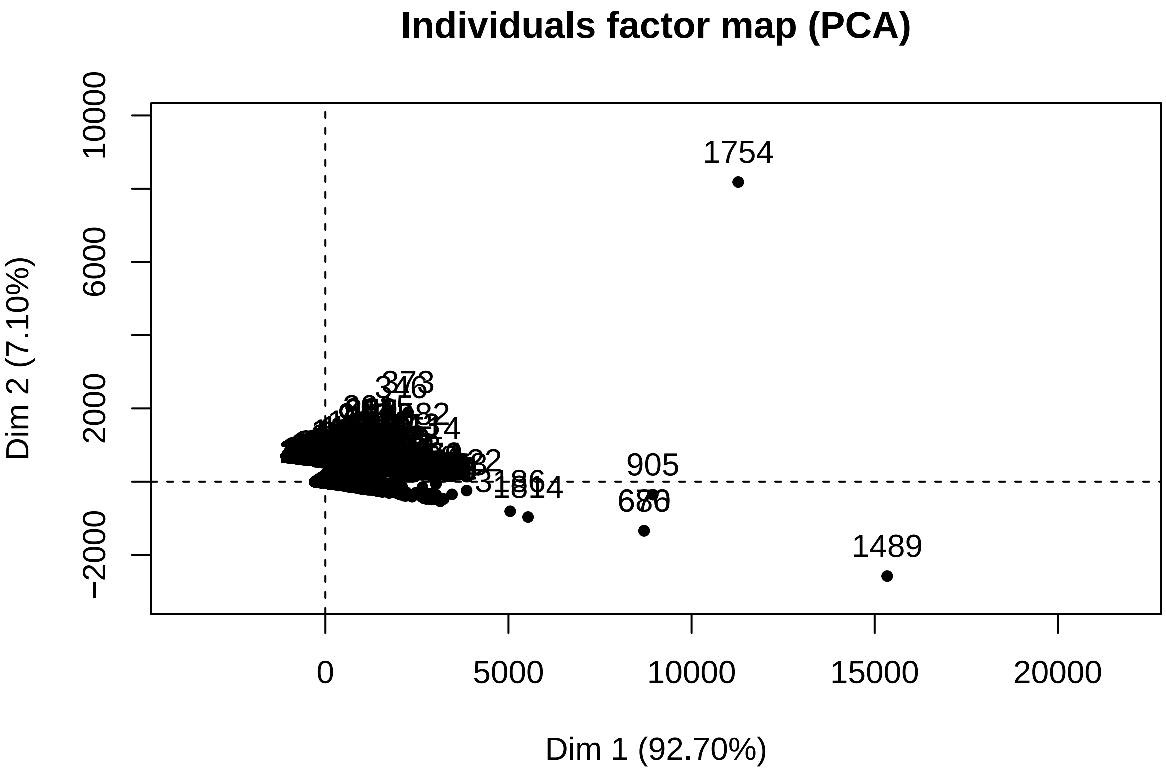
```r
data = spam[,1:57]
head(summary(data))
```

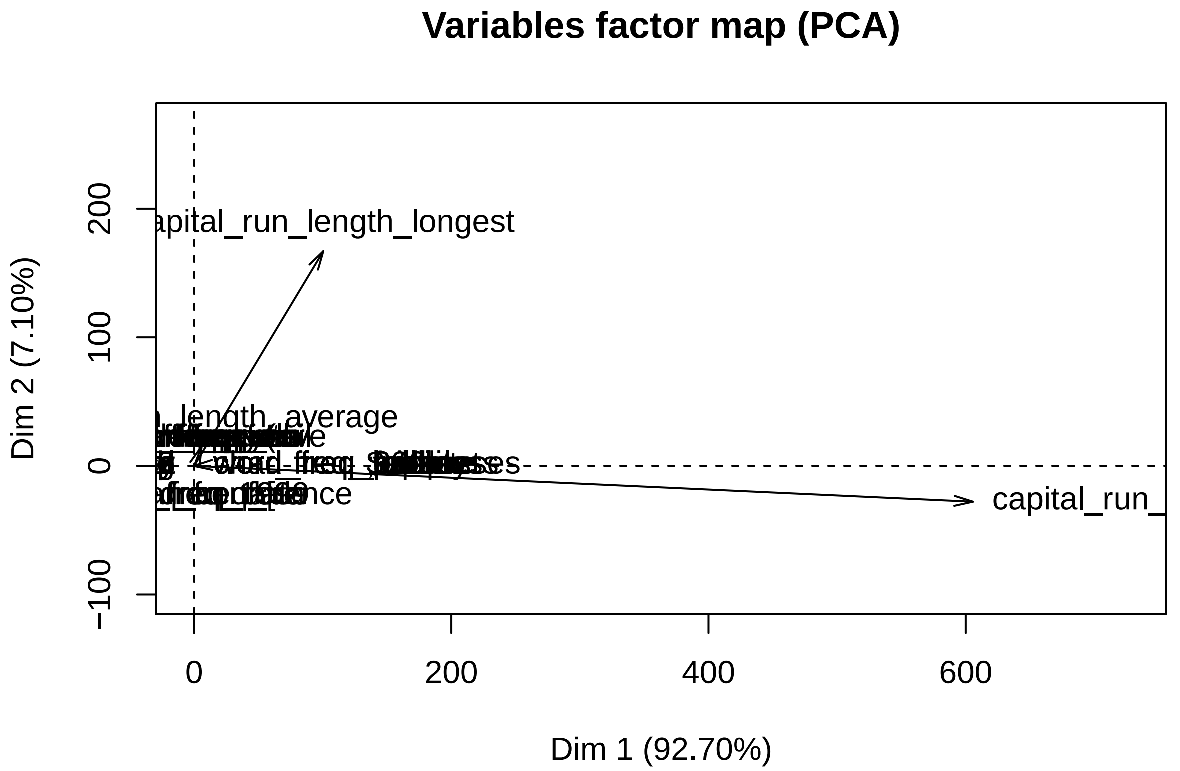
```
##  word_freq_make     word_freq_address  word_freq_all
##  "Min.   :0.0000  " "Min.   : 0.000  " "Min.   :0.0000  "
##  "1st Qu.:0.0000  " "1st Qu.: 0.000  " "1st Qu.:0.0000  "
##  "Median :0.0000  " "Median : 0.000  " "Median :0.0000  "
##  "Mean   :0.1046  " "Mean   : 0.213  " "Mean   :0.2807  "
##  "3rd Qu.:0.0000  " "3rd Qu.: 0.000  " "3rd Qu.:0.4200  "
##  "Max.   :4.5400  " "Max.   :14.280  " "Max.   :5.1000  "
##   word_freq_3d        word_freq_our       word_freq_over
##  "Min.   : 0.00000  " "Min.   : 0.0000  " "Min.   :0.0000  "
##  "1st Qu.: 0.00000  " "1st Qu.: 0.0000  " "1st Qu.:0.0000  "
##  "Median : 0.00000  " "Median : 0.0000  " "Median :0.0000  "
##  "Mean   : 0.06542  " "Mean   : 0.3122  " "Mean   :0.0959  "
##  "3rd Qu.: 0.00000  " "3rd Qu.: 0.3800  " "3rd Qu.:0.0000  "
##  "Max.   :42.81000  " "Max.   :10.0000  " "Max.   :5.8800  "
##  word_freq_remove   word_freq_internet  word_freq_order
##  "Min.   :0.0000  " "Min.   : 0.0000  " "Min.   :0.00000  "
##  "1st Qu.:0.0000  " "1st Qu.: 0.0000  " "1st Qu.:0.00000  "
##  "Median :0.0000  " "Median : 0.0000  " "Median :0.00000  "
##  "Mean   :0.1142  " "Mean   : 0.1053  " "Mean   :0.09007  "
##  "3rd Qu.:0.0000  " "3rd Qu.: 0.0000  " "3rd Qu.:0.00000  "
##  "Max.   :7.2700  " "Max.   :11.1100  " "Max.   :5.26000  "
##  word_freq_mail      word_freq_receive   word_freq_will
##  "Min.   : 0.0000  " "Min.   :0.00000  " "Min.   :0.0000  "
##  "1st Qu.: 0.0000  " "1st Qu.:0.00000  " "1st Qu.:0.0000  "
##  "Median : 0.0000  " "Median :0.00000  " "Median :0.1000  "
##  "Mean   : 0.2394  " "Mean   :0.05982  " "Mean   :0.5417  "
##  "3rd Qu.: 0.1600  " "3rd Qu.:0.00000  " "3rd Qu.:0.8000  "
##  "Max.   :18.1800  " "Max.   :2.61000  " "Max.   :9.6700  "
##  word_freq_people    word_freq_report     word_freq_addresses
##  "Min.   :0.00000  " "Min.   : 0.00000  " "Min.   :0.0000  "
##  "1st Qu.:0.00000  " "1st Qu.: 0.00000  " "1st Qu.:0.0000  "
##  "Median :0.00000  " "Median : 0.00000  " "Median :0.0000  "
##  "Mean   :0.09393  " "Mean   : 0.05863  " "Mean   :0.0492  "
##  "3rd Qu.:0.00000  " "3rd Qu.: 0.00000  " "3rd Qu.:0.0000  "
##  "Max.   :5.55000  " "Max.   :10.00000  " "Max.   :4.4100  "
##  word_freq_free      word_freq_business word_freq_email
##  "Min.   : 0.0000  " "Min.   :0.0000  " "Min.   :0.0000  "
##  "1st Qu.: 0.0000  " "1st Qu.:0.0000  " "1st Qu.:0.0000  "
##  "Median : 0.0000  " "Median :0.0000  " "Median :0.0000  "
##  "Mean   : 0.2488  " "Mean   :0.1426  " "Mean   :0.1847  "
##  "3rd Qu.: 0.1000  " "3rd Qu.:0.0000  " "3rd Qu.:0.0000  "
##  "Max.   :20.0000  " "Max.   :7.1400  " "Max.   :9.0900  "
##  word_freq_you      word_freq_credit     word_freq_your
##  "Min.   : 0.000  " "Min.   : 0.00000  " "Min.   : 0.0000  "
##  "1st Qu.: 0.000  " "1st Qu.: 0.00000  " "1st Qu.: 0.0000  "
##  "Median : 1.310  " "Median : 0.00000  " "Median : 0.2200  "
##  "Mean   : 1.662  " "Mean   : 0.08558  " "Mean   : 0.8098  "
##  "3rd Qu.: 2.640  " "3rd Qu.: 0.00000  " "3rd Qu.: 1.2700  "
##  "Max.   :18.750  " "Max.   :18.18000  " "Max.   :11.1100  "
##  word_freq_font     word_freq_000      word_freq_money
##  "Min.   : 0.0000  " "Min.   :0.0000  " "Min.   : 0.00000  "
##  "1st Qu.: 0.0000  " "1st Qu.:0.0000  " "1st Qu.: 0.00000  "
##  "Median : 0.0000  " "Median :0.0000  " "Median : 0.00000  "
##  "Mean   : 0.1212  " "Mean   :0.1016  " "Mean   : 0.09427  "
##  "3rd Qu.: 0.0000  " "3rd Qu.:0.0000  " "3rd Qu.: 0.00000  "
##  "Max.   :17.1000  " "Max.   :5.4500  " "Max.   :12.50000  "
##   word_freq_hp      word_freq_hpl      word_freq_george
##  "Min.   : 0.0000  " "Min.   : 0.0000  " "Min.   : 0.0000  "
##  "1st Qu.: 0.0000  " "1st Qu.: 0.0000  " "1st Qu.: 0.0000  "
##  "Median : 0.0000  " "Median : 0.0000  " "Median : 0.0000  "
##  "Mean   : 0.5495  " "Mean   : 0.2654  " "Mean   : 0.7673  "
##  "3rd Qu.: 0.0000  " "3rd Qu.: 0.0000  " "3rd Qu.: 0.0000  "
##  "Max.   :20.8300  " "Max.   :16.6600  " "Max.   :33.3300  "
##  word_freq_650     word_freq_lab       word_freq_labs
##  "Min.   :0.0000  " "Min.   : 0.00000  " "Min.   :0.0000  "
##  "1st Qu.:0.0000  " "1st Qu.: 0.00000  " "1st Qu.:0.0000  "
##  "Median :0.0000  " "Median : 0.00000  " "Median :0.0000  "
##  "Mean   :0.1248  " "Mean   : 0.09892  " "Mean   :0.1029  "
##  "3rd Qu.:0.0000  " "3rd Qu.: 0.00000  " "3rd Qu.:0.0000  "
##  "Max.   :9.0900  " "Max.   :14.28000  " "Max.   :5.8800  "
##  word_freq_telnet     word_freq_857       word_freq_data
##  "Min.   : 0.00000  " "Min.   :0.00000  " "Min.   : 0.00000  "
##  "1st Qu.: 0.00000  " "1st Qu.:0.00000  " "1st Qu.: 0.00000  "
##  "Median : 0.00000  " "Median :0.00000  " "Median : 0.00000  "
##  "Mean   : 0.06475  " "Mean   :0.04705  " "Mean   : 0.09723  "
##  "3rd Qu.: 0.00000  " "3rd Qu.:0.00000  " "3rd Qu.: 0.00000  "
##  "Max.   :12.50000  " "Max.   :4.76000  " "Max.   :18.18000  "
##  word_freq_415        word_freq_85       word_freq_technology
##  "Min.   :0.00000  " "Min.   : 0.0000  " "Min.   :0.00000  "
##  "1st Qu.:0.00000  " "1st Qu.: 0.0000  " "1st Qu.:0.00000  "
##  "Median :0.00000  " "Median : 0.0000  " "Median :0.00000  "
##  "Mean   :0.04784  " "Mean   : 0.1054  " "Mean   :0.09748  "
##  "3rd Qu.:0.00000  " "3rd Qu.: 0.0000  " "3rd Qu.:0.00000  "
##  "Max.   :4.76000  " "Max.   :20.0000  " "Max.   :7.69000  "
##  word_freq_1999    word_freq_parts     word_freq_pm
##  "Min.   :0.000  " "Min.   :0.0000  " "Min.   : 0.00000  "
##  "1st Qu.:0.000  " "1st Qu.:0.0000  " "1st Qu.: 0.00000  "
##  "Median :0.000  " "Median :0.0000  " "Median : 0.00000  "
##  "Mean   :0.137  " "Mean   :0.0132  " "Mean   : 0.07863  "
##  "3rd Qu.:0.000  " "3rd Qu.:0.0000  " "3rd Qu.: 0.00000  "
##  "Max.   :6.890  " "Max.   :8.3300  " "Max.   :11.11000  "
##  word_freq_direct     word_freq_cs       word_freq_meeting
##  "Min.   :0.00000  " "Min.   :0.00000  " "Min.   : 0.0000  "
##  "1st Qu.:0.00000  " "1st Qu.:0.00000  " "1st Qu.: 0.0000  "
##  "Median :0.00000  " "Median :0.00000  " "Median : 0.0000  "
##  "Mean   :0.06483  " "Mean   :0.04367  " "Mean   : 0.1323  "
##  "3rd Qu.:0.00000  " "3rd Qu.:0.00000  " "3rd Qu.: 0.0000  "
##  "Max.   :4.76000  " "Max.   :7.14000  " "Max.   :14.2800  "
##  word_freq_original word_freq_project    word_freq_re
##  "Min.   :0.0000  " "Min.   : 0.0000  " "Min.   : 0.0000  "
##  "1st Qu.:0.0000  " "1st Qu.: 0.0000  " "1st Qu.: 0.0000  "
##  "Median :0.0000  " "Median : 0.0000  " "Median : 0.0000  "
##  "Mean   :0.0461  " "Mean   : 0.0792  " "Mean   : 0.3012  "
##  "3rd Qu.:0.0000  " "3rd Qu.: 0.0000  " "3rd Qu.: 0.1100  "
##  "Max.   :3.5700  " "Max.   :20.0000  " "Max.   :21.4200  "
##  word_freq_edu       word_freq_table      word_freq_conference
##  "Min.   : 0.0000  " "Min.   :0.000000  " "Min.   : 0.00000  "
##  "1st Qu.: 0.0000  " "1st Qu.:0.000000  " "1st Qu.: 0.00000  "
##  "Median : 0.0000  " "Median :0.000000  " "Median : 0.00000  "
##  "Mean   : 0.1798  " "Mean   :0.005444  " "Mean   : 0.03187  "
##  "3rd Qu.: 0.0000  " "3rd Qu.:0.000000  " "3rd Qu.: 0.00000  "
##  "Max.   :22.0500  " "Max.   :2.170000  " "Max.   :10.00000  "
##   char_freq_;         char_freq_(       char_freq_[
##  "Min.   :0.00000  " "Min.   :0.000  " "Min.   :0.00000  "
##  "1st Qu.:0.00000  " "1st Qu.:0.000  " "1st Qu.:0.00000  "
##  "Median :0.00000  " "Median :0.065  " "Median :0.00000  "
##  "Mean   :0.03857  " "Mean   :0.139  " "Mean   :0.01698  "
##  "3rd Qu.:0.00000  " "3rd Qu.:0.188  " "3rd Qu.:0.00000  "
##  "Max.   :4.38500  " "Max.   :9.752  " "Max.   :4.08100  "
##   char_freq_!         char_freq_$         char_freq_#
##  "Min.   : 0.0000  " "Min.   :0.00000  " "Min.   : 0.00000  "
##  "1st Qu.: 0.0000  " "1st Qu.:0.00000  " "1st Qu.: 0.00000  "
##  "Median : 0.0000  " "Median :0.00000  " "Median : 0.00000  "
##  "Mean   : 0.2691  " "Mean   :0.07581  " "Mean   : 0.04424  "
##  "3rd Qu.: 0.3150  " "3rd Qu.:0.05200  " "3rd Qu.: 0.00000  "
##  "Max.   :32.4780  " "Max.   :6.00300  " "Max.   :19.82900  "
##  capital_run_length_average capital_run_length_longest
##  "Min.   :   1.000  "       "Min.   :   1.00  "
##  "1st Qu.:   1.588  "       "1st Qu.:   6.00  "
##  "Median :   2.276  "       "Median :  15.00  "
##  "Mean   :   5.191  "       "Mean   :  52.17  "
##  "3rd Qu.:   3.706  "       "3rd Qu.:  43.00  "
##  "Max.   :1102.500  "       "Max.   :9989.00  "
##  capital_run_length_total
##  "Min.   :    1.0  "
##  "1st Qu.:   35.0  "
##  "Median :   95.0  "
##  "Mean   :  283.3  "
##  "3rd Qu.:  266.0  "
##  "Max.   :15841.0  "
```

On va utiliser la bibliothèque **FactoMineR** pour effectuer l'ACP :

Une ACP non normée :

```
library(FactoMineR)
pca = PCA(data , scale.unit = F)
```


Individuals factor map (PCA)
Dim 2 (7.10%)
10000
6000
2000
−2000
Dim 1 (92.70%)
0
5000
10000
15000
20000
1754
905
1489
373
346



```r
head(summary(pca))
```

```
##
## Call:
## PCA(X = data, scale.unit = F)
##
##
## Eigenvalues
##                             Dim.1      Dim.2      Dim.3      Dim.4
## Variance               376923.652  28885.649    749.647     11.408
## % of var.                  92.703      7.104      0.184      0.003
## Cumulative % of var.       92.703     99.807     99.991     99.994
##                             Dim.5      Dim.6      Dim.7      Dim.8
## Variance                    4.134      2.626      1.946      1.637
## % of var.                   0.001      0.001      0.000      0.000
## Cumulative % of var.       99.995     99.996     99.996     99.997
##                             Dim.9     Dim.10     Dim.11     Dim.12
## Variance                    1.334      1.064      1.002      0.856
## % of var.                   0.000      0.000      0.000      0.000
## Cumulative % of var.       99.997     99.997     99.998     99.998
##                            Dim.13     Dim.14     Dim.15     Dim.16
## Variance                    0.829      0.760      0.680      0.613
## % of var.                   0.000      0.000      0.000      0.000
## Cumulative % of var.       99.998     99.998     99.998     99.998
##                            Dim.17     Dim.18     Dim.19     Dim.20
## Variance                  0.568      0.474      0.422      0.402
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.     99.999     99.999     99.999     99.999
##                          Dim.21     Dim.22     Dim.23     Dim.24
## Variance                  0.377      0.306      0.289      0.269
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.     99.999     99.999     99.999     99.999
##                          Dim.25     Dim.26     Dim.27     Dim.28
## Variance                  0.227      0.220      0.194      0.188
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.     99.999     99.999     99.999     99.999
##                          Dim.29     Dim.30     Dim.31     Dim.32
## Variance                  0.179      0.175      0.146      0.139
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.     99.999    100.000    100.000    100.000
##                          Dim.33     Dim.34     Dim.35     Dim.36
## Variance                  0.132      0.128      0.125      0.117
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.    100.000    100.000    100.000    100.000
##                          Dim.37     Dim.38     Dim.39     Dim.40
## Variance                  0.111      0.108      0.085      0.082
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.    100.000    100.000    100.000    100.000
##                          Dim.41     Dim.42     Dim.43     Dim.44
## Variance                  0.081      0.078      0.070      0.064
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.    100.000    100.000    100.000    100.000
##                          Dim.45     Dim.46     Dim.47     Dim.48
## Variance                  0.055      0.052      0.050      0.046
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.    100.000    100.000    100.000    100.000
##                          Dim.49     Dim.50     Dim.51     Dim.52
## Variance                  0.045      0.042      0.040      0.038
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.    100.000    100.000    100.000    100.000
##                          Dim.53     Dim.54     Dim.55     Dim.56
## Variance                  0.033      0.022      0.011      0.006
## % of var.                 0.000      0.000      0.000      0.000
## Cumulative % of var.    100.000    100.000    100.000    100.000
##                          Dim.57
## Variance                  0.000
## % of var.                 0.000
## Cumulative % of var.    100.000
##
## Individuals (the 10 first)
##                                  Dist      Dim.1      ctr     cos2
## 1                          |   10.550 |   -3.787    0.000    0.129 |
## 2                          |  746.314 |  742.632    0.032    0.990 |
## 3                          | 2022.573 | 2019.854    0.235    0.997 |
## 4                          |   93.129 |  -93.048    0.000    0.998 |
## 5                          |   93.129 |  -93.048    0.000    0.998 |
## 6                          |  232.317 | -232.291    0.003    1.000 |
## 7                          |  178.000 | -176.887    0.002    0.988 |
## 8                          |  237.918 | -237.883    0.003    1.000 |
## 9                          | 1049.982 | 1024.854    0.061    0.953 |
## 10                         |  465.816 |  457.883    0.012    0.966 |
##                               Dim.2      ctr     cos2      Dim.3      ctr
## 1                             9.412    0.000    0.796 |   -2.247    0.000
## 2                           -74.001    0.004    0.010 |   -0.568    0.000
## 3                           102.077    0.008    0.003 |  -23.777    0.016
## 4                             3.009    0.000    0.001 |   -1.044    0.000
## 5                             3.009    0.000    0.001 |   -1.044    0.000
## 6                             0.836    0.000    0.000 |   -0.049    0.000
## 7                           -19.587    0.000    0.012 |   -0.099    0.000
## 8                            -2.324    0.000    0.000 |   -0.266    0.000
## 9                           226.877    0.039    0.047 |  -25.491    0.019
## 10                          -85.593    0.006    0.034 |   -0.204    0.000
##                                cos2
## 1                             0.045 |
## 2                             0.000 |
## 3                             0.000 |
## 4                             0.000 |
## 5                             0.000 |
## 6                             0.000 |
## 7                             0.000 |
## 8                             0.000 |
## 9                             0.001 |
## 10                            0.000 |
##
## Variables (the 10 first)
##                               Dim.1    ctr   cos2    Dim.2    ctr   cos2
## word_freq_make             |  0.028  0.000  0.008 |  0.005  0.000  0.000 |
## word_freq_address          | -0.028  0.000  0.000 |  0.018  0.000  0.000 |
## word_freq_all              |  0.037  0.000  0.005 |  0.041  0.000  0.007 |
## word_freq_3d               |  0.031  0.000  0.000 |  0.017  0.000  0.000 |
## word_freq_our              |  0.003  0.000  0.000 |  0.039  0.000  0.003 |
## word_freq_over             |  0.023  0.000  0.007 |  0.015  0.000  0.003 |
## word_freq_remove           | -0.002  0.000  0.000 |  0.028  0.000  0.005 |
## word_freq_internet         |  0.017  0.000  0.002 |  0.008  0.000  0.000 |
## word_freq_order            |  0.070  0.000  0.064 |  0.019  0.000  0.005 |
## word_freq_mail             |  0.058  0.000  0.008 |  0.043  0.000  0.004 |
##                             Dim.3    ctr   cos2
## word_freq_make               0.007  0.000  0.001 |
## word_freq_address            0.000  0.000  0.000 |
## word_freq_all                0.026  0.000  0.003 |
## word_freq_3d                -0.008  0.000  0.000 |
## word_freq_our                0.019  0.000  0.001 |
## word_freq_over              -0.016  0.000  0.004 |
## word_freq_remove             0.004  0.000  0.000 |
## word_freq_internet          -0.003  0.000  0.000 |
## word_freq_order              0.010  0.000  0.001 |
## word_freq_mail               0.019  0.000  0.001 |
```

```
## NULL
```

Interpretation :

Variables :

```
df = as.data.frame(pca$var$contrib)
head(df[order(df$Dim.1 , decreasing = T),])
```

```
##                                  Dim.1        Dim.2        Dim.3
## capital_run_length_total   9.731493e+01 2.682320e+00 2.709562e-03
## capital_run_length_longest 2.675963e+00 9.652560e+01 7.982684e-01
## capital_run_length_average 9.059934e-03 7.919517e-01 9.919725e+01
## word_freq_george           2.823689e-05 1.102488e-06 5.507539e-05
## word_freq_font             2.944941e-06 3.081760e-06 5.152856e-05
## word_freq_re               2.485624e-06 3.412451e-08 1.152040e-05
##                                  Dim.4        Dim.5
## capital_run_length_total   2.563382e-05 2.712709e-08
## capital_run_length_longest 5.075434e-06 4.469053e-05
## capital_run_length_average 1.954353e-05 3.360502e-04
## word_freq_george           9.784176e+01 1.132735e+00
## word_freq_font             2.302156e-03 4.632671e-04
## word_freq_re               4.360961e-03 1.840655e-02
```

La variable qui contribue le plus à la formation de l'axe 1 est : **capital_run_length_total** 97.3% , **capital_run_length_longest** 2.67%

```
head(df[order(df$Dim.2 , decreasing = T),])
```

```
##                                  Dim.1        Dim.2        Dim.3
## capital_run_length_longest 2.675963e+00 9.652560e+01 7.982684e-01
## capital_run_length_total   9.731493e+01 2.682320e+00 2.709562e-03
## capital_run_length_average 9.059934e-03 7.919517e-01 9.919725e+01
## char_freq_(                3.205798e-07 3.150584e-05 3.275013e-04
## word_freq_your             1.153215e-06 2.209680e-05 1.677025e-07
## word_freq_hp               1.488300e-06 1.033634e-05 2.198292e-05
##                                  Dim.4        Dim.5
## capital_run_length_longest 5.075434e-06 4.469053e-05
## capital_run_length_total   2.563382e-05 2.712709e-08
## capital_run_length_average 1.954353e-05 3.360502e-04
## char_freq_(                1.834514e-07 1.062297e-01
## word_freq_your             3.358894e-01 5.927688e+00
## word_freq_hp               2.810181e-03 3.948854e+01
```

La variable qui contribue le plus à la formation de l'axe 2 est : **capital_run_length_longest** 96.5% , **capital_run_length_total** 2.68%

Individus :

```
df = as.data.frame(pca$ind$contrib)
head(df[order(df$Dim.1 , decreasing = T),])
```

```
##          Dim.1       Dim.2       Dim.3      Dim.4        Dim.5
## 1489 13.572976  5.00414575  0.19009575 0.09942892 0.0005325704
## 1754  7.329447 50.37152944 10.55915067 0.08762297 0.2962729524
## 905   4.611469  0.09452977  0.06301744 0.02977187 0.0012335745
## 680   4.368939  1.35366149  0.06707356 0.03031312 0.0055186449
## 676   4.366959  1.35299019  0.06720291 0.03029929 0.0055203258
## 1814  1.766585  0.70318042  0.02925296 0.01066288 0.0053578463
```

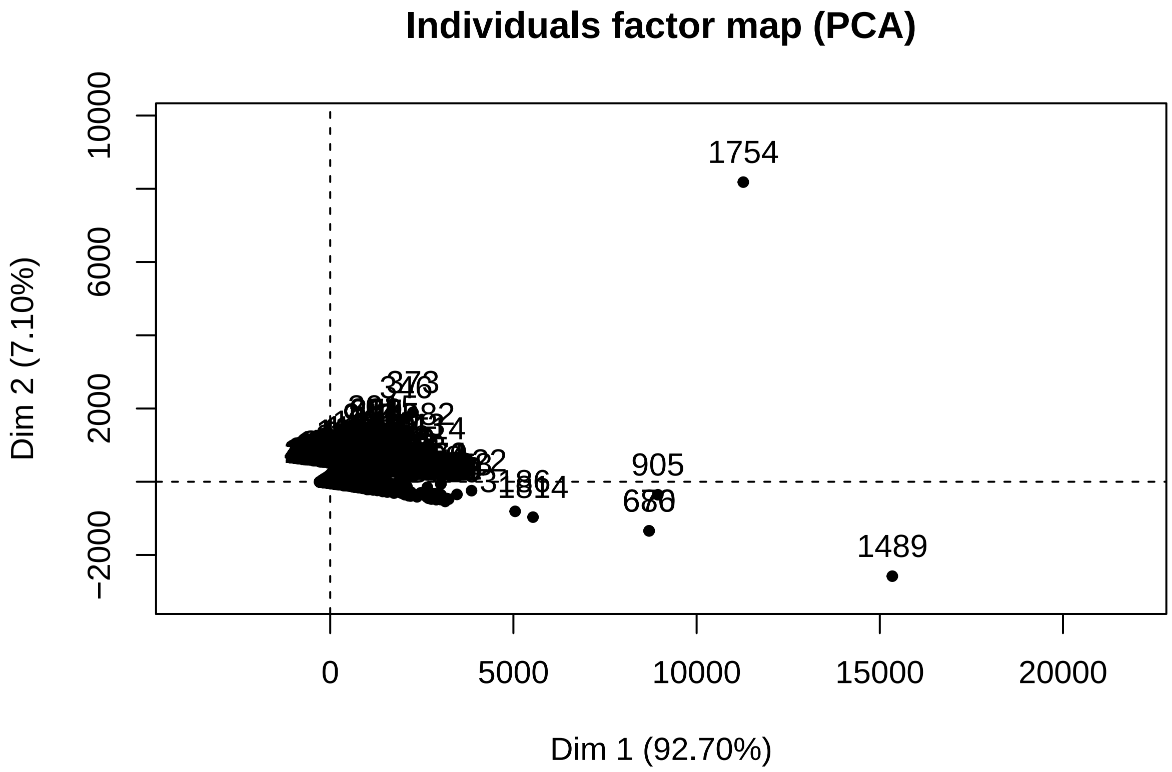
Les individus qui contribuent le plus à la formation de l'axe 1 sont : **1489** *13%* et **1754** *7.32%* et **905 , 680 , 676** 4%

```
df = as.data.frame(pca$ind$contrib)
head(df[order(df$Dim.2 , decreasing = T),])
```

```
##           Dim.1     Dim.2       Dim.3       Dim.4        Dim.5
## 1754  7.3294474 50.371529 10.55915067 0.087622968 2.962730e-01
## 1489 13.5729764  5.004146  0.19009575 0.099428921 5.325704e-04
## 373   0.2940446  2.707792 24.04310664 0.003644853 5.519886e-05
## 346   0.2473419  2.323303 20.63327151 0.002862750 3.148129e-04
## 680   4.3689388  1.353661  0.06707356 0.030313124 5.518645e-03
## 676   4.3669590  1.352990  0.06720291 0.030299291 5.520326e-03
```

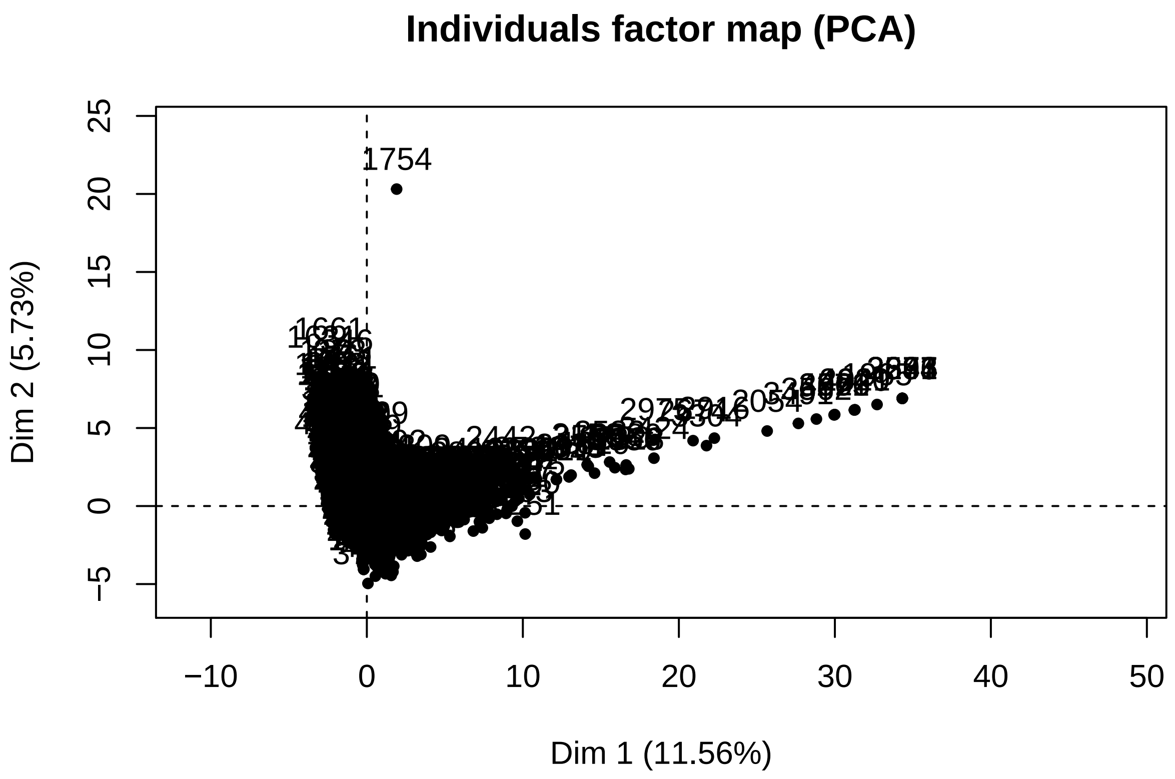
Les individus qui contribuent le plus à la formation de l'axe 2 sont : **1754** *50.37%* et **1489** *5%*

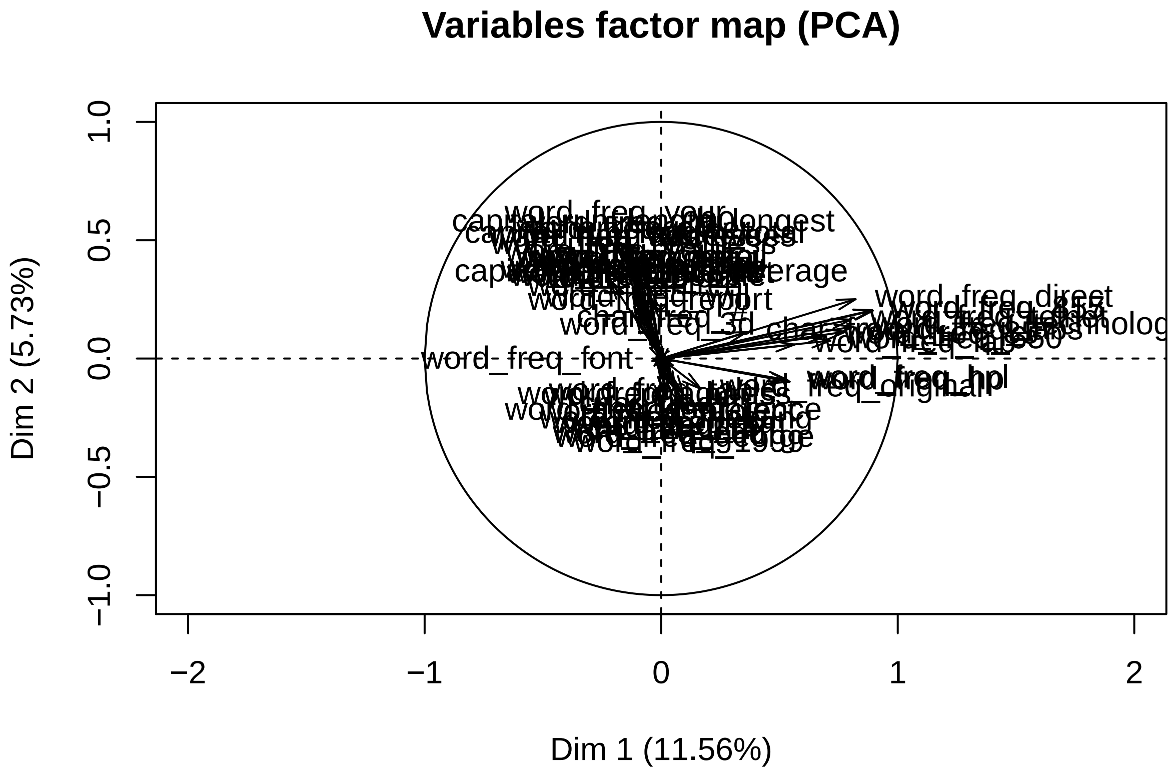
Une représentation graphique de **l'ACP non normée :**



Une ACP normée :

```r
acp.n = PCA(data,scale.unit =TRUE)
```



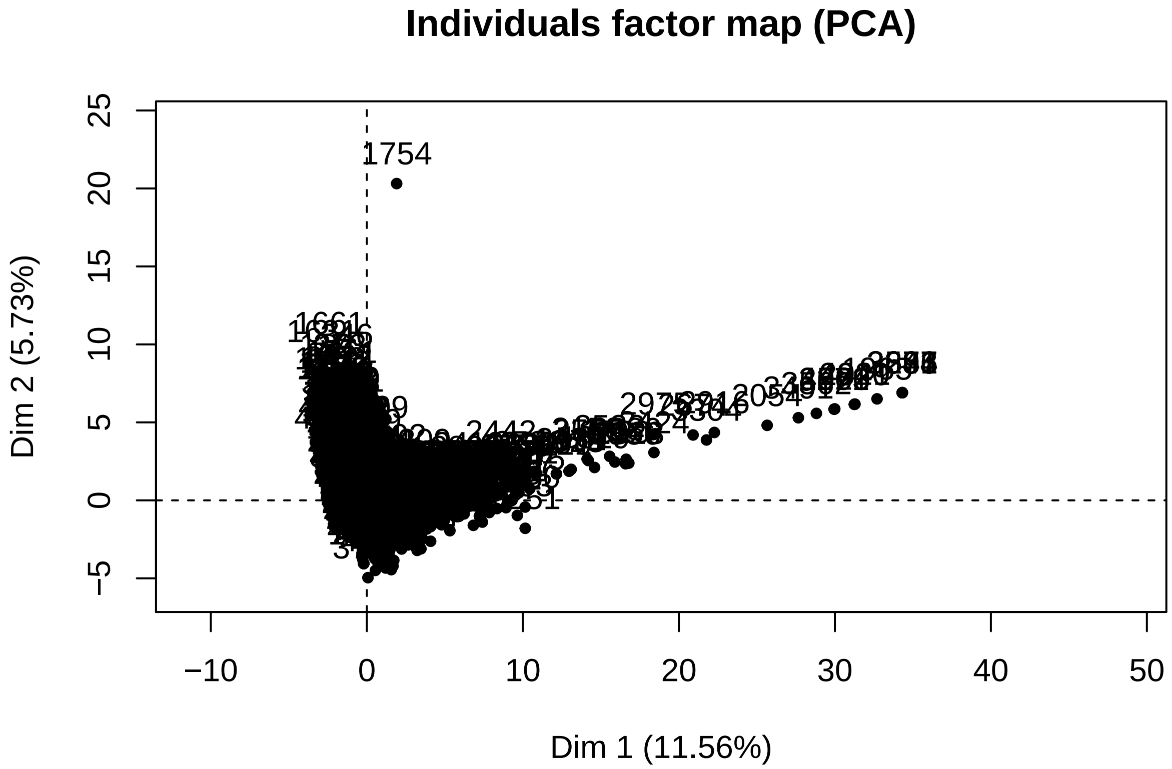


```
summary(acp.n)
```

```
##
## Call:
## PCA(X = data, scale.unit = TRUE)
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
## Variance               6.592   3.267   2.003   1.613   1.546   1.463
## % of var.             11.565   5.732   3.514   2.830   2.713   2.566
## Cumulative % of var.  11.565  17.297  20.811  23.642  26.354  28.920
##                        Dim.7   Dim.8   Dim.9  Dim.10  Dim.11  Dim.12
## Variance               1.414   1.375   1.295   1.277   1.217   1.130
## % of var.              2.481   2.412   2.272   2.240   2.135   1.983
## Cumulative % of var.  31.401  33.813  36.085  38.326  40.460  42.443
##                       Dim.13  Dim.14  Dim.15  Dim.16  Dim.17  Dim.18
## Variance               1.112   1.095   1.087   1.063   1.049   1.023
## % of var.              1.950   1.921   1.907   1.866   1.840   1.795
## Cumulative % of var.  44.394  46.315  48.222  50.088  51.927  53.723
##                       Dim.19  Dim.20  Dim.21  Dim.22  Dim.23  Dim.24
## Variance               1.013   1.003   0.996   0.978   0.965   0.941
## % of var.              1.777   1.759   1.747   1.716   1.692   1.651
## Cumulative % of var.  55.499  57.259  59.006  60.722  62.414  64.066
##                       Dim.25  Dim.26  Dim.27  Dim.28  Dim.29  Dim.30
## Variance               0.937   0.924   0.915   0.905   0.873   0.866
## % of var.              1.643   1.622   1.606   1.587   1.532   1.519
## Cumulative % of var.  65.709  67.331  68.936  70.524  72.056  73.575
##                       Dim.31  Dim.32  Dim.33  Dim.34  Dim.35  Dim.36
## Variance               0.836   0.827   0.798   0.782   0.777   0.756
## % of var.              1.468   1.450   1.399   1.371   1.362   1.325
## Cumulative % of var.  75.043  76.493  77.893  79.264  80.626  81.952
##                       Dim.37  Dim.38  Dim.39  Dim.40  Dim.41  Dim.42
## Variance               0.734   0.723   0.705   0.690   0.675   0.666
## % of var.              1.288   1.269   1.236   1.210   1.184   1.168
## Cumulative % of var.  83.240  84.508  85.744  86.955  88.138  89.307
##                       Dim.43  Dim.44  Dim.45  Dim.46  Dim.47  Dim.48
## Variance               0.619   0.608   0.582   0.577   0.524   0.489
## % of var.              1.086   1.067   1.021   1.012   0.920   0.857
## Cumulative % of var.  90.393  91.460  92.481  93.494  94.414  95.271
##                       Dim.49  Dim.50  Dim.51  Dim.52  Dim.53  Dim.54
## Variance               0.450   0.409   0.376   0.366   0.335   0.305
## % of var.              0.790   0.718   0.659   0.642   0.587   0.536
## Cumulative % of var.  96.061  96.779  97.438  98.079  98.667  99.203
##                       Dim.55  Dim.56  Dim.57
## Variance               0.260   0.190   0.004
## % of var.              0.457   0.334   0.007
## Cumulative % of var.  99.659  99.993 100.000
##
## Individuals (the 10 first)
##                                 Dist    Dim.1    ctr   cos2    Dim.2    ctr
## 1                          |  2.826 | -0.732  0.002  0.067 | -0.043  0.000
## 2                          |  3.512 | -1.185  0.005  0.114 |  2.068  0.028
## 3                          |  9.018 | -1.468  0.007  0.026 |  5.024  0.168
## 4                          |  2.910 | -0.805  0.002  0.077 |  0.428  0.001
## 5                          |  2.910 | -0.806  0.002  0.077 |  0.427  0.001
## 6                          |  5.392 | -0.493  0.001  0.008 | -0.490  0.002
## 7                          |  5.635 | -1.026  0.003  0.033 |  1.021  0.007
## 8                          |  5.470 | -0.504  0.001  0.009 | -0.488  0.002
## 9                          |  8.901 | -1.266  0.005  0.020 |  3.641  0.088
## 10                         |  2.227 | -0.841  0.002  0.143 |  0.401  0.001
##                               cos2    Dim.3    ctr   cos2
## 1                            0.000 | -0.581  0.004  0.042 |
## 2                            0.347 |  0.036  0.000  0.000 |
## 3                            0.310 |  3.278  0.117  0.132 |
## 4                            0.022 | -0.583  0.004  0.040 |
## 5                            0.022 | -0.585  0.004  0.040 |
## 6                            0.008 | -0.373  0.002  0.005 |
## 7                            0.033 | -1.738  0.033  0.095 |
## 8                            0.008 | -0.419  0.002  0.006 |
## 9                            0.167 |  1.492  0.024  0.028 |
## 10                           0.032 |  0.030  0.000  0.000 |
##
## Variables (the 10 first)
##                                Dim.1    ctr   cos2    Dim.2    ctr   cos2
## word_freq_make             | -0.112  0.191  0.013 |  0.307  2.877  0.094 |
## word_freq_address          | -0.029  0.012  0.001 | -0.030  0.028  0.001 |
## word_freq_all              | -0.121  0.222  0.015 |  0.299  2.729  0.089 |
## word_freq_3d               | -0.016  0.004  0.000 |  0.020  0.012  0.000 |
## word_freq_our             | -0.094  0.135  0.009 |  0.220  1.480  0.048 |
## word_freq_over            | -0.118  0.210  0.014 |  0.303  2.815  0.092 |
## word_freq_remove          | -0.119  0.214  0.014 |  0.261  2.084  0.068 |
## word_freq_internet        | -0.087  0.115  0.008 |  0.239  1.749  0.057 |
## word_freq_order           | -0.117  0.207  0.014 |  0.425  5.523  0.180 |
## word_freq_mail            | -0.052  0.041  0.003 |  0.278  2.359  0.077 |
##                            Dim.3    ctr   cos2  
## word_freq_make            -0.090  0.405  0.008 |
## word_freq_address         -0.014  0.009  0.000 |
## word_freq_all             -0.029  0.043  0.001 |
## word_freq_3d               0.018  0.016  0.000 |
## word_freq_our             -0.194  1.873  0.038 |
## word_freq_over             0.010  0.005  0.000 |
## word_freq_remove          -0.183  1.670  0.033 |
## word_freq_internet        -0.067  0.222  0.004 |
## word_freq_order            0.186  1.731  0.035 |
## word_freq_mail             0.086  0.365  0.007 |
```

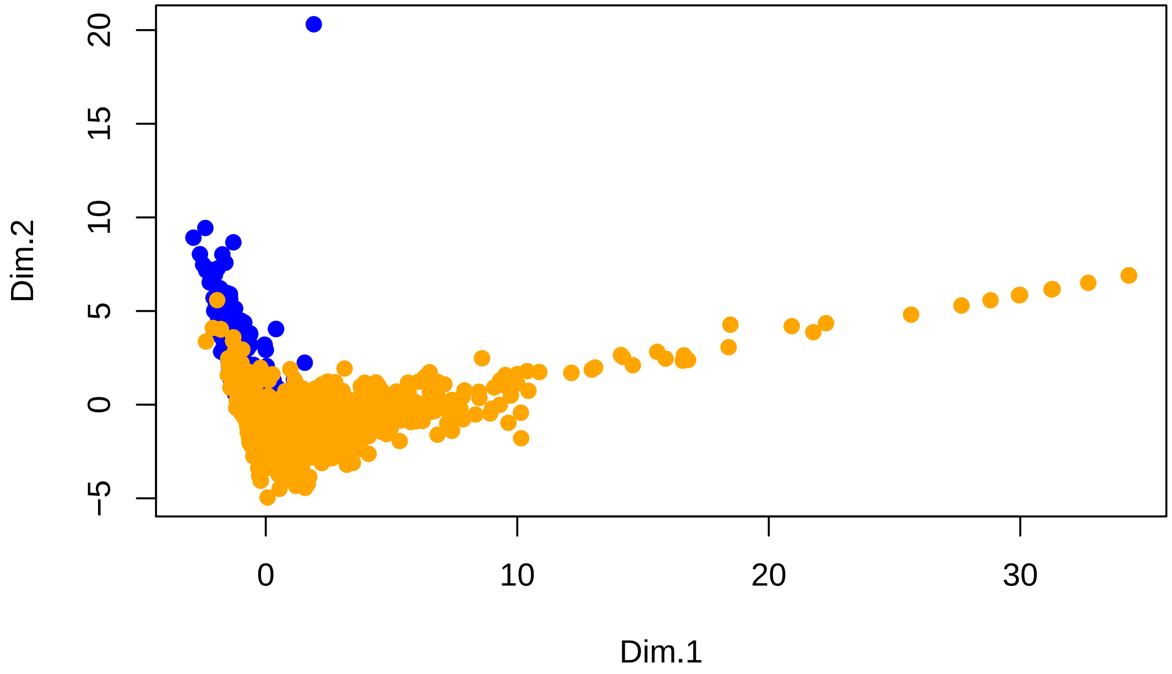
Une représentation graphique de **l'ACP normée :**

```
plot(acp.n)
```



Représentation graphique avec séparation des données :

```
coord <- acp.n$ind$coord
col = rep("blue",nrow(data))
col[spam$y == 0] = "orange"
plot(coord[,1:2],col=col , pch = 19)
```



recodage en forme de facteurs :

```
make=factor(spam[,"word_freq_make"] > 0, c(TRUE, FALSE),labels=c("make", "Nmk"))
table(make)
```

```
## make
## make  Nmk
## 1053 3548
```

```
CapLMq=cut(spam[,"capital_run_length_total"],breaks=quantile(spam[,"capital_run_length_total"], probs =
labels = c("Mm1","Mm2","Mm3"),include.lowest = TRUE)
table(CapLMq)
```

```
## CapLMq
##  Mm1  Mm2  Mm3
## 1537 1530 1534
```

```
table(make, CapLMq)
```

```
##       CapLMq
## make    Mm1  Mm2  Mm3
##   make   75  264  714
##   Nmk  1462 1266  820
```

```
data = cbind(as.numeric(make), as.numeric(CapLMq))
names(data) = c("make","CapLMq")
```

Analayse descriptive :

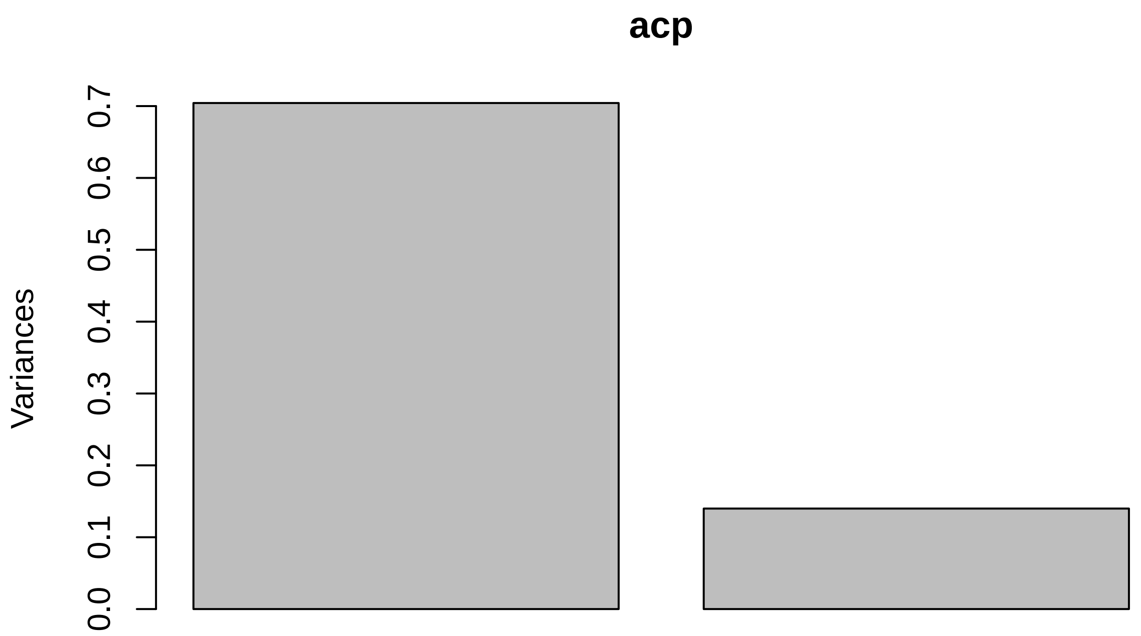
```
summary(data)
```

```
##        V1              V2
##  Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:1.000
##  Median :2.000   Median :2.000
##  Mean   :1.771   Mean   :1.999
##  3rd Qu.:2.000   3rd Qu.:3.000
##  Max.   :2.000   Max.   :3.000
```

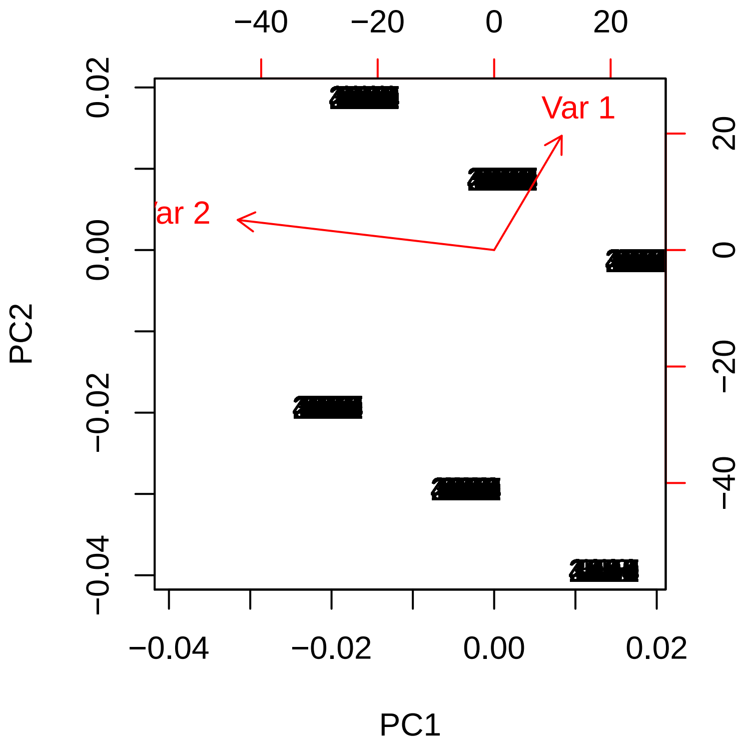
```
acp = prcomp(data)

summary(acp)
```

```
## Importance of components:
##                           PC1    PC2
## Standard deviation     0.8392 0.3740
## Proportion of Variance 0.8343 0.1657
## Cumulative Proportion  0.8343 1.0000
```

```
plot(acp)
```



```
biplot(acp)
```



Générateur Aléatoire de visages

```
setwd("C:\\Users\\W  7\\Desktop\\Master 2\\AA2\\TP2_ACP\\Images\\")
n = 10;
img = list()
names = c('img1.dat','img2.dat','img3.dat','img4.dat','img5.dat','img6.dat','img7.dat','img8.dat','img9

for (i in 1:n) {
    aux = t(as.matrix(read.table(names[i],sep=",")))
    img[[i]] = aux[,112:1]
    }

N1 = dim(img[[1]])[1]
N2 = dim(img[[1]])[2]

for (i in 1:n) {
    image(img[[i]], col=gray(0:256/256) , xlab="", ylab="",axes=FALSE)
    }
```

Image moyenne :

```
# Image moyenne
moy = matrix(0,N1,N2)

for (i in 1:n) {
    moy = moy + img[[i]]
    }
moy = moy/n

image(moy, col=gray(0:256/256) , xlab="", ylab="",axes=FALSE)
```

```
# Matrice de donnees
X = matrix(0,n,N1*N2)
for (i in 1:n) {
    X[i,] = as.vector(img[[i]]-moy)
}
```

Géneration Aléatoire :

On multiplie la matrice des composantes principales par la transposé de la matrice de corrélation, on obtient une matrice qui represente les images à l'aide des 2 premieres composantes

```
acp = prcomp(X)
mat = acp$x[,1:10]%*%t(acp$rotation[,1:10])
mat = scale(mat, center = -1*moy, scale=FALSE)
for (i in 1:5){
  img <- matrix((mat[i,]), 92, 112)
  image(1:92, 1:112,img,col=gray((0:256)/256) , xlab = "" , ylab = "" , axes =F)
}
```