# Componentes Principales - Laboratorio 2

*Mathias Bourel*

*1/5/2019*

```
rm(list=ls(all=TRUE) )
ls()
```

```
## character(0)
```

```
library(MASS)

if ( !require( HSAUR2 ) )
{
    chooseCRANmirror( graphics=F )
    install.packages( 'HSAUR2' )
    library( HSAUR2 )
}
```

```
## Loading required package: HSAUR2
```

```
## Loading required package: tools
```

Base de datos Heptathlon

```
data( 'heptathlon', package='HSAUR2' )
head( heptathlon )
```

```
##                      hurdles highjump  shot run200m longjump javelin
## Joyner-Kersee (USA)    12.69     1.86 15.80   22.56     7.27   45.66
## John (GDR)             12.85     1.80 16.23   23.65     6.71   42.56
## Behmer (GDR)           13.20     1.83 14.20   23.10     6.68   44.54
## Sablovskaite (URS)     13.61     1.80 15.23   23.92     6.25   42.78
## Choubenkova (URS)      13.51     1.74 14.76   23.93     6.32   47.46
## Schulz (GDR)           13.75     1.83 13.50   24.65     6.33   42.82
##                      run800m score
## Joyner-Kersee (USA)   128.51  7291
## John (GDR)            126.12  6897
## Behmer (GDR)          124.20  6858
## Sablovskaite (URS)    132.24  6540
## Choubenkova (URS)     127.90  6540
## Schulz (GDR)          125.79  6411
```

```
tail( heptathlon )
```

```
##                   hurdles highjump  shot run200m longjump javelin run800m
## Hautenauve (BEL)    14.04     1.77 11.81   25.61     5.99   35.68  133.90
## Kytola (FIN)        14.31     1.77 11.66   25.69     5.75   39.48  133.35
## Geremias (BRA)      14.23     1.71 12.95   25.50     5.50   39.64  144.02
## Hui-Ing (TAI)       14.85     1.68 10.00   25.23     5.47   39.14  137.30
## Jeong-Mi (KOR)      14.53     1.71 10.83   26.61     5.50   39.26  139.17
## Launa (PNG)         16.42     1.50 11.78   26.16     4.88   46.38  163.43
##                   score
## Hautenauve (BEL)   5734
## Kytola (FIN)       5686
## Geremias (BRA)     5508
```

```
## Hui-Ing (TAI)     5290
## Jeong-Mi (KOR)    5289
## Launa (PNG)       4566
```

```
dim(heptathlon)
```

```
## [1] 25  8
```

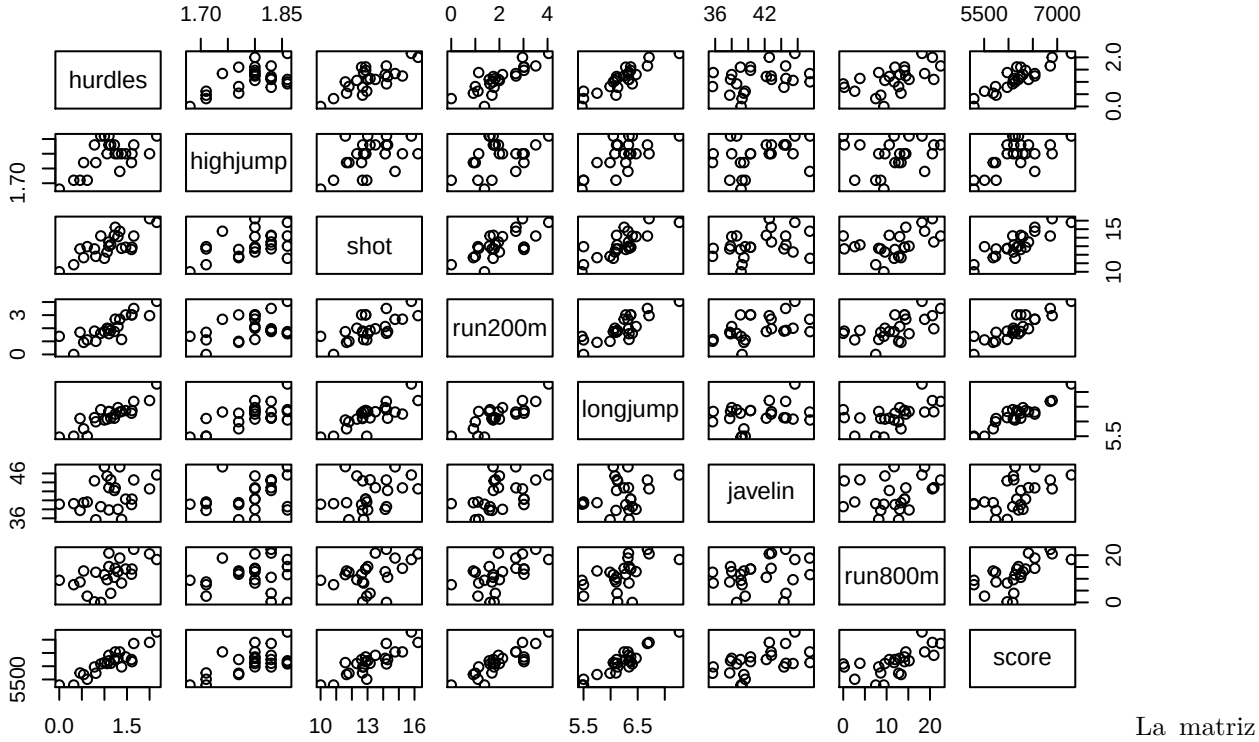
```
##sin el de Guinea
heptathlon=heptathlon[-25,]

##Para que los valores grandes sean buenos.
heptathlon$hurdles=max(heptathlon$hurdles)-heptathlon$hurdles
heptathlon$run200m=max(heptathlon$run200m)-heptathlon$run200m
heptathlon$run800m=max(heptathlon$run800m)-heptathlon$run800m
heptathlon
```

```
##                      hurdles highjump  shot run200m longjump javelin
## Joyner-Kersee (USA)    2.16     1.86 15.80    4.05     7.27   45.66
## John (GDR)             2.00     1.80 16.23    2.96     6.71   42.56
## Behmer (GDR)           1.65     1.83 14.20    3.51     6.68   44.54
## Sablovskaite (URS)     1.24     1.80 15.23    2.69     6.25   42.78
## Choubenkova (URS)      1.34     1.74 14.76    2.68     6.32   47.46
## Schulz (GDR)           1.10     1.83 13.50    1.96     6.33   42.82
## Fleming (AUS)          1.47     1.80 12.88    3.02     6.37   40.28
## Greiner (USA)          1.30     1.80 14.13    2.13     6.47   38.00
## Lajbnerova (CZE)       1.22     1.83 14.28    1.75     6.11   42.20
## Bouraga (URS)          1.60     1.77 12.62    3.02     6.28   39.06
## Wijnsma (HOL)          1.10     1.86 13.01    1.58     6.34   37.86
## Dimitrova (BUL)        1.61     1.80 12.88    3.02     6.37   40.28
## Scheider (SWI)         1.00     1.86 11.58    1.74     6.05   47.50
## Braun (FRG)            1.14     1.83 13.16    1.83     6.12   44.58
## Ruotsalainen (FIN)     1.06     1.80 12.32    2.00     6.08   45.44
## Yuping (CHN)           0.92     1.86 14.21    1.61     6.40   38.60
## Hagger (GB)            1.38     1.80 12.75    1.14     6.34   35.76
## Brown (USA)            0.78     1.83 12.69    1.78     6.13   44.34
## Mulliner (GB)          0.46     1.71 12.68    1.69     6.10   37.76
## Hautenauve (BEL)       0.81     1.77 11.81    1.00     5.99   35.68
## Kytola (FIN)           0.54     1.77 11.66    0.92     5.75   39.48
## Geremias (BRA)         0.62     1.71 12.95    1.11     5.50   39.64
## Hui-Ing (TAI)          0.00     1.68 10.00    1.38     5.47   39.14
## Jeong-Mi (KOR)         0.32     1.71 10.83    0.00     5.50   39.26
##                      run800m score
## Joyner-Kersee (USA)    18.16  7291
## John (GDR)             20.55  6897
## Behmer (GDR)           22.47  6858
## Sablovskaite (URS)     14.43  6540
## Choubenkova (URS)      18.77  6540
## Schulz (GDR)           20.88  6411
## Fleming (AUS)          14.13  6351
## Greiner (USA)          13.02  6297
## Lajbnerova (CZE)       10.62  6252
## Bouraga (URS)          11.93  6252
## Wijnsma (HOL)          15.18  6205
## Dimitrova (BUL)        14.13  6171
```

```
## Scheider (SWI)        11.74  6137
## Braun (FRG)            3.85  6109
## Ruotsalainen (FIN)     9.61  6101
## Yuping (CHN)           0.00  6087
## Hagger (GB)            8.19  5975
## Brown (USA)            0.24  5972
## Mulliner (GB)          8.65  5746
## Hautenauve (BEL)      12.77  5734
## Kytola (FIN)          13.32  5686
## Geremias (BRA)         2.65  5508
## Hui-Ing (TAI)          9.37  5290
## Jeong-Mi (KOR)         7.50  5289
```

Veamos el ScatterPlot

```
plot(heptathlon)
```




La matriz de correlación

```
round(cor(heptathlon[,-8]),2)
```

```
##          hurdles highjump shot run200m longjump javelin run800m
## hurdles     1.00     0.58 0.77    0.83     0.89    0.33    0.56
## highjump    0.58     1.00 0.46    0.39     0.66    0.35    0.15
## shot        0.77     0.46 1.00    0.67     0.78    0.34    0.41
## run200m     0.83     0.39 0.67    1.00     0.81    0.47    0.57
## longjump    0.89     0.66 0.78    0.81     1.00    0.29    0.52
## javelin     0.33     0.35 0.34    0.47     0.29    1.00    0.26
## run800m     0.56     0.15 0.41    0.57     0.52    0.26    1.00
```

# 1 - Análisis usando princomp de MASS

Con la función prcomp hago PCA sobre la matriz de var/cov (en principio no estandarizada)usando la descomposición SVD de una matriz

Recordamos la descomposición SVD de una matriz:

```
A=matrix(c(6,8,1,9),2,2)
svd(A)
```

```
## $d
## [1] 13.019889  3.533056
##
## $u
##            [,1]       [,2]
## [1,] -0.3951299 -0.9186252
## [2,] -0.9186252  0.3951299
##
## $v
##            [,1]       [,2]
## [1,] -0.7465334 -0.6653480
## [2,] -0.6653480  0.7465334
```

```
eigen(A)
```

```
## eigen() decomposition
## $values
## [1] 10.701562  4.298438
##
## $vectors
##            [,1]       [,2]
## [1,] -0.2080415 -0.5066742
## [2,] -0.9781200  0.8621376
```

Observar que para hacer la SVD la matriz no tiene porque ser cuadrada

```
B=matrix(c(6,8,1,9,4,2),2,3)
svd(B)
```

```
## $d
## [1] 13.483406  4.494193
##
## $u
##            [,1]       [,2]
## [1,] -0.4505313 -0.8927606
## [2,] -0.8927606  0.4505313
##
## $v
##            [,1]       [,2]
## [1,] -0.7301770 -0.3899060
## [2,] -0.6293200  0.7035793
## [3,] -0.2660787 -0.5940955
```

```
C=matrix(c(6,8,1,9,4,2,9,4,2,9,4,2),4,3)
svd(C)
```

```
## $d
## [1] 17.796123  7.657582  5.353451
```

```
##
## $u
##             [,1]         [,2]       [,3]
## [1,] -0.4052702 -0.008836769 -0.3726103
## [2,] -0.6331436  0.403238357  0.6606515
## [3,] -0.3993255 -0.892971661  0.1607017
## [4,] -0.5248089  0.199805787 -0.6315665
##
## $v
##             [,1]        [,2]       [,3]
## [1,] -0.7091081  0.53256584 -0.4621031
## [2,] -0.4821574 -0.84444307 -0.2333240
## [3,] -0.5144802  0.05735448  0.8555820
```

prcomp con matriz no normalizada

```
pca0=prcomp(heptathlon[, -8], scale = FALSE)
```

Es mejor usar la normalizaci?n:

```
pca=prcomp(heptathlon[, -8], scale = TRUE)
ls(pca)
```

```
## [1] "center"   "rotation" "scale"    "sdev"     "x"
```

```
objects(pca)
```

```
## [1] "center"   "rotation" "scale"    "sdev"     "x"
```

```
attributes( pca )
```

```
## $names
## [1] "sdev"     "rotation" "center"   "scale"    "x"
##
## $class
## [1] "prcomp"
```

Promedio en cada columna en la matriz de datos

```
pca$center
```

```
##   hurdles  highjump      shot   run200m  longjump   javelin   run800m
##  1.117500  1.793750 13.173333  2.023750  6.205417 41.278333 11.756667
```

```
colMeans(heptathlon)
```

```
##     hurdles    highjump        shot     run200m    longjump     javelin
##    1.117500    1.793750   13.173333    2.023750    6.205417   41.278333
##     run800m       score
##   11.756667 6154.125000
```

Desviación estandar de cada columna de la matriz de datos

```
pca$scale
```

```
##    hurdles   highjump       shot    run200m   longjump    javelin
## 0.51456398 0.05232112 1.49714995 0.93676972 0.40165938 3.46870690
##    run800m
## 6.14724800
```

```
apply(heptathlon,2,sd)
```

```
##      hurdles     highjump         shot      run200m     longjump
##   0.51456398   0.05232112   1.49714995   0.93676972   0.40165938
##      javelin      run800m        score
##   3.46870690   6.14724800 481.59755096
```

Los loadings, i.e los vectores de a's: es una matriz cuyas columnas contiene los vectores propios de la matriz de varianzas/covarianzas,

```
pca$rotation
```

```
##                 PC1         PC2        PC3         PC4         PC5
## hurdles  -0.4503876  0.05772161 -0.1739345  0.04840598 -0.19889364
## highjump -0.3145115 -0.65133162 -0.2088272 -0.55694554  0.07076358
## shot     -0.4024884 -0.02202088 -0.1534709  0.54826705  0.67166466
## run200m  -0.4270860  0.18502783  0.1301287  0.23095946 -0.61781764
## longjump -0.4509639 -0.02492486 -0.2697589 -0.01468275 -0.12151793
## javelin  -0.2423079 -0.32572229  0.8806995  0.06024757  0.07874396
## run800m  -0.3029068  0.65650503  0.1930020 -0.57418128  0.31880178
##                 PC6         PC7
## hurdles   0.84665086 -0.06961672
## highjump -0.09007544  0.33155910
## shot     -0.09886359  0.22904298
## run200m  -0.33279359  0.46971934
## longjump -0.38294411 -0.74940781
## javelin   0.07193437 -0.21108138
## run800m  -0.05217664  0.07718616
```

```
##coincide con
svd(cor(heptathlon[,-8]))$v
```

```
##            [,1]        [,2]       [,3]        [,4]        [,5]        [,6]
## [1,] -0.4503876  0.05772161 -0.1739345  0.04840598 -0.19889364 -0.84665086
## [2,] -0.3145115 -0.65133162 -0.2088272 -0.55694554  0.07076358  0.09007544
## [3,] -0.4024884 -0.02202088 -0.1534709  0.54826705  0.67166466  0.09886359
## [4,] -0.4270860  0.18502783  0.1301287  0.23095946 -0.61781764  0.33279359
## [5,] -0.4509639 -0.02492486 -0.2697589 -0.01468275 -0.12151793  0.38294411
## [6,] -0.2423079 -0.32572229  0.8806995  0.06024757  0.07874396 -0.07193437
## [7,] -0.3029068  0.65650503  0.1930020 -0.57418128  0.31880178  0.05217664
##             [,7]
## [1,] -0.06961672
## [2,]  0.33155910
## [3,]  0.22904298
## [4,]  0.46971934
## [5,] -0.74940781
## [6,] -0.21108138
## [7,]  0.07718616
```

y pca$rotation[,1] y pca rotation[,2] son entonces los dos primeros vectores propios de la matriz de varianzas/covarianzas.

Desviación estandar de cada a.

```
pca$sdev
```

```
## [1] 2.0793370 0.9481532 0.9109016 0.6831967 0.5461888 0.3374549 0.2620420
```

```
##Por lo tanto al cuadrado esto devuelve los valores propios de la matriz
##de correlacion:
(pca$sdev)^2
```

```
## [1] 4.32364217 0.89899445 0.82974172 0.46675769 0.29832218 0.11387578
## [7] 0.06866602
```

```
eigen(cor(heptathlon[,-8]))
```

```
## eigen() decomposition
## $values
## [1] 4.32364217 0.89899445 0.82974172 0.46675769 0.29832218 0.11387578
## [7] 0.06866602
##
## $vectors
##            [,1]        [,2]       [,3]        [,4]        [,5]        [,6]
## [1,] -0.4503876  0.05772161 -0.1739345 -0.04840598 -0.19889364  0.84665086
## [2,] -0.3145115 -0.65133162 -0.2088272  0.55694554  0.07076358 -0.09007544
## [3,] -0.4024884 -0.02202088 -0.1534709 -0.54826705  0.67166466 -0.09886359
## [4,] -0.4270860  0.18502783  0.1301287 -0.23095946 -0.61781764 -0.33279359
## [5,] -0.4509639 -0.02492486 -0.2697589  0.01468275 -0.12151793 -0.38294411
## [6,] -0.2423079 -0.32572229  0.8806995 -0.06024757  0.07874396  0.07193437
## [7,] -0.3029068  0.65650503  0.1930020  0.57418128  0.31880178 -0.05217664
##             [,7]
## [1,] -0.06961672
## [2,]  0.33155910
## [3,]  0.22904298
## [4,]  0.46971934
## [5,] -0.74940781
## [6,] -0.21108138
## [7,]  0.07718616
```

El porcentaje de varianza explicada por cada componente

```
(pca$sdev)^2/sum((pca$sdev)^2)*100
```

```
## [1] 61.7663167 12.8427778 11.8534531  6.6679670  4.2617454  1.6267969
## [7]  0.9809432
```

Proyección los individuos sobre las componentes

```
pca$x
```

```
##                              PC1         PC2          PC3          PC4
## Joyner-Kersee (USA) -4.757530189 -0.13986143 -0.006040526  0.293416339
## John (GDR)          -3.147943402  0.94859029 -0.243919842  0.549171385
## Behmer (GDR)        -2.926184760  0.69534239  0.622293440 -0.554744912
## Sablovskaite (URS)  -1.288135516  0.17900713  0.250632380  0.637174187
## Choubenkova (URS)   -1.503450994  0.96177329  1.780588549  0.784035325
## Schulz (GDR)        -0.958467101  0.35121643  0.413086366 -1.113546938
## Fleming (AUS)       -0.953445060  0.49982537 -0.265135015 -0.140202490
## Greiner (USA)       -0.633239267  0.37592917 -1.140338594  0.142558348
## Lajbnerova (CZE)    -0.381571974 -0.71213459 -0.068395353  0.087212735
## Bouraga (URS)       -0.522322004  0.77688861 -0.481071429  0.283745698
## Wijnsma (HOL)       -0.217701500 -0.23369645 -1.154221444 -1.260128609
## Dimitrova (BUL)     -1.075984276  0.51552998 -0.312458252 -0.127032432
## Scheider (SWI)       0.003014986 -1.44688825  1.582739069 -1.254415325
```

```
## Braun (FRG)          0.109183759 -1.63595645  0.469577294  0.362580442
## Ruotsalainen (FIN)   0.208868056 -0.68866173  1.152140223 -0.112914470
## Yuping (CHN)         0.232507119 -1.95999641 -1.541230813  0.598325122
## Hagger (GB)          0.659520046 -0.08775813 -1.796509771 -0.182375000
## Brown (USA)          0.756854602 -2.04292201  0.451506018  0.476926314
## Mulliner (GB)        1.880932819  0.91530324 -0.359311801  0.799619094
## Hautenauve (BEL)     1.828170404  0.72629699 -1.048640439 -0.711793161
## Kytola (FIN)         2.118203163  0.39921397  0.190158154 -0.788445056
## Geremias (BRA)       2.770706272  0.03463584  0.170274969  1.385562494
## Hui-Ing (TAI)        3.901166920  1.20175472  0.943677497 -0.002429122
## Jeong-Mi (KOR)       3.896847898  0.36656804  0.390599321 -0.152299968
##                             PC5         PC6         PC7
## Joyner-Kersee (USA) -0.36183307 -0.27050283 -0.47587527
## John (GDR)           0.75364464  0.37770017 -0.05172711
## Behmer (GDR)        -0.19035037 -0.25780287  0.11054960
## Sablovskaite (URS)   0.60362153 -0.21575716  0.53075152
## Choubenkova (URS)    0.58969949  0.08014332 -0.30081842
## Schulz (GDR)         0.71483887 -0.25436956  0.03838796
## Fleming (AUS)       -0.86581530  0.03691813  0.23005943
## Greiner (USA)        0.20807431 -0.14236240 -0.06374657
## Lajbnerova (CZE)     0.67727618  0.25014881  0.35555639
## Bouraga (URS)       -1.18784299  0.39881271  0.19712215
## Wijnsma (HOL)        0.37497195 -0.20267731  0.17459647
## Dimitrova (BUL)     -0.91992929  0.26727067  0.21111846
## Scheider (SWI)      -0.20526249  0.17597425 -0.03915701
## Braun (FRG)         -0.14712208  0.26134199 -0.01334416
## Ruotsalainen (FIN)  -0.31539746  0.18351622 -0.14127555
## Yuping (CHN)         0.17451428 -0.50175724  0.04999374
## Hagger (GB)         -0.05104049  0.55058471 -0.46388534
## Brown (USA)         -0.38154294 -0.26606429 -0.11099445
## Mulliner (GB)       -0.06942955 -0.73259727 -0.31281502
## Hautenauve (BEL)     0.14092347  0.06933542 -0.07548638
## Kytola (FIN)         0.41815113 -0.03363651  0.12143219
## Geremias (BRA)       0.28541366  0.38083979  0.34574480
## Hui-Ing (TAI)       -0.67080776 -0.52756760  0.09436975
## Jeong-Mi (KOR)       0.42524426  0.37250885 -0.41055719
```

```
#Proyección de los 5 primeros individuos sobre las 2 primeras componentes.
pca$x[1:5,1:2]
```

```
##                            PC1        PC2
## Joyner-Kersee (USA) -4.757530 -0.1398614
## John (GDR)          -3.147943  0.9485903
## Behmer (GDR)        -2.926185  0.6953424
## Sablovskaite (URS)  -1.288136  0.1790071
## Choubenkova (URS)   -1.503451  0.9617733
```

```
##Matriz X normalizada
scaleX=as.matrix(scale(heptathlon[,-8]))
##proyeccion individuos sobre componente 1
scaleX%*%pca$rotation[,1]
```

```
##                             [,1]
## Joyner-Kersee (USA) -4.757530189
## John (GDR)          -3.147943402
## Behmer (GDR)        -2.926184760
```

```
## Sablovskaite (URS)  -1.288135516
## Choubenkova (URS)   -1.503450994
## Schulz (GDR)        -0.958467101
## Fleming (AUS)       -0.953445060
## Greiner (USA)       -0.633239267
## Lajbnerova (CZE)    -0.381571974
## Bouraga (URS)       -0.522322004
## Wijnsma (HOL)       -0.217701500
## Dimitrova (BUL)     -1.075984276
## Scheider (SWI)       0.003014986
## Braun (FRG)          0.109183759
## Ruotsalainen (FIN)   0.208868056
## Yuping (CHN)         0.232507119
## Hagger (GB)          0.659520046
## Brown (USA)          0.756854602
## Mulliner (GB)        1.880932819
## Hautenauve (BEL)     1.828170404
## Kytola (FIN)         2.118203163
## Geremias (BRA)       2.770706272
## Hui-Ing (TAI)        3.901166920
## Jeong-Mi (KOR)       3.896847898
```

```
c1=pca$x[,1]

scaleX%*%pca$rotation
```

```
##                              PC1         PC2          PC3          PC4
## Joyner-Kersee (USA) -4.757530189 -0.13986143 -0.006040526  0.293416339
## John (GDR)          -3.147943402  0.94859029 -0.243919842  0.549171385
## Behmer (GDR)        -2.926184760  0.69534239  0.622293440 -0.554744912
## Sablovskaite (URS)  -1.288135516  0.17900713  0.250632380  0.637174187
## Choubenkova (URS)   -1.503450994  0.96177329  1.780588549  0.784035325
## Schulz (GDR)        -0.958467101  0.35121643  0.413086366 -1.113546938
## Fleming (AUS)       -0.953445060  0.49982537 -0.265135015 -0.140202490
## Greiner (USA)       -0.633239267  0.37592917 -1.140338594  0.142558348
## Lajbnerova (CZE)    -0.381571974 -0.71213459 -0.068395353  0.087212735
## Bouraga (URS)       -0.522322004  0.77688861 -0.481071429  0.283745698
## Wijnsma (HOL)       -0.217701500 -0.23369645 -1.154221444 -1.260128609
## Dimitrova (BUL)     -1.075984276  0.51552998 -0.312458252 -0.127032432
## Scheider (SWI)       0.003014986 -1.44688825  1.582739069 -1.254415325
## Braun (FRG)          0.109183759 -1.63595645  0.469577294  0.362580442
## Ruotsalainen (FIN)   0.208868056 -0.68866173  1.152140223 -0.112914470
## Yuping (CHN)         0.232507119 -1.95999641 -1.541230813  0.598325122
## Hagger (GB)          0.659520046 -0.08775813 -1.796509771 -0.182375000
## Brown (USA)          0.756854602 -2.04292201  0.451506018  0.476926314
## Mulliner (GB)        1.880932819  0.91530324 -0.359311801  0.799619094
## Hautenauve (BEL)     1.828170404  0.72629699 -1.048640439 -0.711793161
## Kytola (FIN)         2.118203163  0.39921397  0.190158154 -0.788445056
## Geremias (BRA)       2.770706272  0.03463584  0.170274969  1.385562494
## Hui-Ing (TAI)        3.901166920  1.20175472  0.943677497 -0.002429122
## Jeong-Mi (KOR)       3.896847898  0.36656804  0.390599321 -0.152299968
##                              PC5         PC6         PC7
## Joyner-Kersee (USA) -0.36183307 -0.27050283 -0.47587527
## John (GDR)           0.75364464  0.37770017 -0.05172711
## Behmer (GDR)        -0.19035037 -0.25780287  0.11054960
```

```
## Sablovskaite (URS)   0.60362153 -0.21575716  0.53075152
## Choubenkova (URS)    0.58969949  0.08014332 -0.30081842
## Schulz (GDR)         0.71483887 -0.25436956  0.03838796
## Fleming (AUS)       -0.86581530  0.03691813  0.23005943
## Greiner (USA)        0.20807431 -0.14236240 -0.06374657
## Lajbnerova (CZE)     0.67727618  0.25014881  0.35555639
## Bouraga (URS)       -1.18784299  0.39881271  0.19712215
## Wijnsma (HOL)        0.37497195 -0.20267731  0.17459647
## Dimitrova (BUL)     -0.91992929  0.26727067  0.21111846
## Scheider (SWI)      -0.20526249  0.17597425 -0.03915701
## Braun (FRG)         -0.14712208  0.26134199 -0.01334416
## Ruotsalainen (FIN)  -0.31539746  0.18351622 -0.14127555
## Yuping (CHN)         0.17451428 -0.50175724  0.04999374
## Hagger (GB)         -0.05104049  0.55058471 -0.46388534
## Brown (USA)         -0.38154294 -0.26606429 -0.11099445
## Mulliner (GB)       -0.06942955 -0.73259727 -0.31281502
## Hautenauve (BEL)     0.14092347  0.06933542 -0.07548638
## Kytola (FIN)         0.41815113 -0.03363651  0.12143219
## Geremias (BRA)       0.28541366  0.38083979  0.34574480
## Hui-Ing (TAI)       -0.67080776 -0.52756760  0.09436975
## Jeong-Mi (KOR)       0.42524426  0.37250885 -0.41055719
```

```
pca$x
```

```
##                             PC1         PC2          PC3          PC4
## Joyner-Kersee (USA) -4.757530189 -0.13986143 -0.006040526  0.293416339
## John (GDR)          -3.147943402  0.94859029 -0.243919842  0.549171385
## Behmer (GDR)        -2.926184760  0.69534239  0.622293440 -0.554744912
## Sablovskaite (URS)  -1.288135516  0.17900713  0.250632380  0.637174187
## Choubenkova (URS)   -1.503450994  0.96177329  1.780588549  0.784035325
## Schulz (GDR)        -0.958467101  0.35121643  0.413086366 -1.113546938
## Fleming (AUS)       -0.953445060  0.49982537 -0.265135015 -0.140202490
## Greiner (USA)       -0.633239267  0.37592917 -1.140338594  0.142558348
## Lajbnerova (CZE)    -0.381571974 -0.71213459 -0.068395353  0.087212735
## Bouraga (URS)       -0.522322004  0.77688861 -0.481071429  0.283745698
## Wijnsma (HOL)       -0.217701500 -0.23369645 -1.154221444 -1.260128609
## Dimitrova (BUL)     -1.075984276  0.51552998 -0.312458252 -0.127032432
## Scheider (SWI)       0.003014986 -1.44688825  1.582739069 -1.254415325
## Braun (FRG)          0.109183759 -1.63595645  0.469577294  0.362580442
## Ruotsalainen (FIN)   0.208868056 -0.68866173  1.152140223 -0.112914470
## Yuping (CHN)         0.232507119 -1.95999641 -1.541230813  0.598325122
## Hagger (GB)          0.659520046 -0.08775813 -1.796509771 -0.182375000
## Brown (USA)          0.756854602 -2.04292201  0.451506018  0.476926314
## Mulliner (GB)        1.880932819  0.91530324 -0.359311801  0.799619094
## Hautenauve (BEL)     1.828170404  0.72629699 -1.048640439 -0.711793161
## Kytola (FIN)         2.118203163  0.39921397  0.190158154 -0.788445056
## Geremias (BRA)       2.770706272  0.03463584  0.170274969  1.385562494
## Hui-Ing (TAI)        3.901166920  1.20175472  0.943677497 -0.002429122
## Jeong-Mi (KOR)       3.896847898  0.36656804  0.390599321 -0.152299968
##                             PC5         PC6         PC7
## Joyner-Kersee (USA) -0.36183307 -0.27050283 -0.47587527
## John (GDR)           0.75364464  0.37770017 -0.05172711
## Behmer (GDR)        -0.19035037 -0.25780287  0.11054960
## Sablovskaite (URS)   0.60362153 -0.21575716  0.53075152
## Choubenkova (URS)    0.58969949  0.08014332 -0.30081842
```

```
## Schulz (GDR)         0.71483887 -0.25436956  0.03838796
## Fleming (AUS)       -0.86581530  0.03691813  0.23005943
## Greiner (USA)        0.20807431 -0.14236240 -0.06374657
## Lajbnerova (CZE)     0.67727618  0.25014881  0.35555639
## Bouraga (URS)       -1.18784299  0.39881271  0.19712215
## Wijnsma (HOL)        0.37497195 -0.20267731  0.17459647
## Dimitrova (BUL)     -0.91992929  0.26727067  0.21111846
## Scheider (SWI)      -0.20526249  0.17597425 -0.03915701
## Braun (FRG)         -0.14712208  0.26134199 -0.01334416
## Ruotsalainen (FIN)  -0.31539746  0.18351622 -0.14127555
## Yuping (CHN)         0.17451428 -0.50175724  0.04999374
## Hagger (GB)         -0.05104049  0.55058471 -0.46388534
## Brown (USA)         -0.38154294 -0.26606429 -0.11099445
## Mulliner (GB)       -0.06942955 -0.73259727 -0.31281502
## Hautenauve (BEL)     0.14092347  0.06933542 -0.07548638
## Kytola (FIN)         0.41815113 -0.03363651  0.12143219
## Geremias (BRA)       0.28541366  0.38083979  0.34574480
## Hui-Ing (TAI)       -0.67080776 -0.52756760  0.09436975
## Jeong-Mi (KOR)       0.42524426  0.37250885 -0.41055719
```

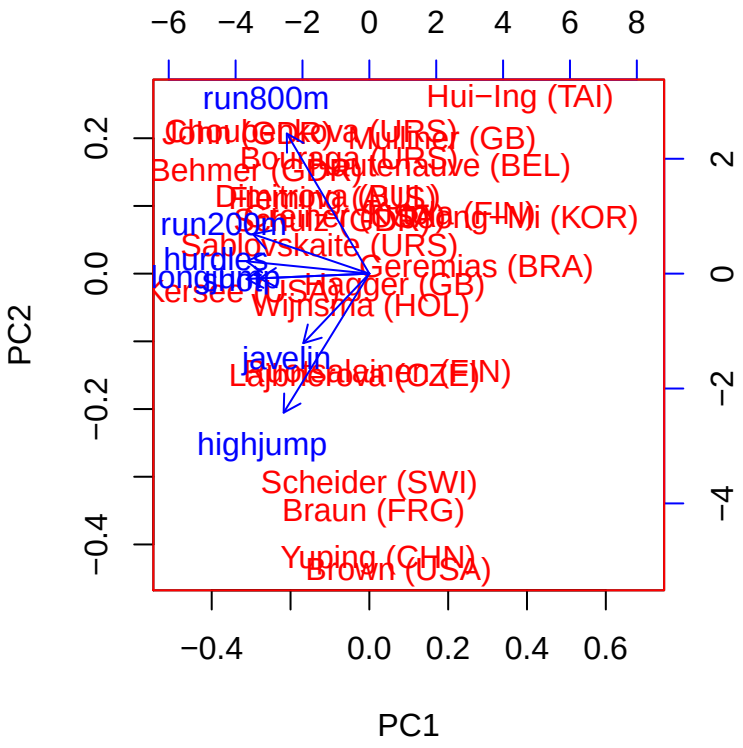
Plot

```
heptathlon
```

```
##                     hurdles highjump  shot run200m longjump javelin
## Joyner-Kersee (USA)    2.16     1.86 15.80    4.05     7.27   45.66
## John (GDR)             2.00     1.80 16.23    2.96     6.71   42.56
## Behmer (GDR)           1.65     1.83 14.20    3.51     6.68   44.54
## Sablovskaite (URS)     1.24     1.80 15.23    2.69     6.25   42.78
## Choubenkova (URS)      1.34     1.74 14.76    2.68     6.32   47.46
## Schulz (GDR)           1.10     1.83 13.50    1.96     6.33   42.82
## Fleming (AUS)          1.47     1.80 12.88    3.02     6.37   40.28
## Greiner (USA)          1.30     1.80 14.13    2.13     6.47   38.00
## Lajbnerova (CZE)       1.22     1.83 14.28    1.75     6.11   42.20
## Bouraga (URS)          1.60     1.77 12.62    3.02     6.28   39.06
## Wijnsma (HOL)          1.10     1.86 13.01    1.58     6.34   37.86
## Dimitrova (BUL)        1.61     1.80 12.88    3.02     6.37   40.28
## Scheider (SWI)         1.00     1.86 11.58    1.74     6.05   47.50
## Braun (FRG)            1.14     1.83 13.16    1.83     6.12   44.58
## Ruotsalainen (FIN)     1.06     1.80 12.32    2.00     6.08   45.44
## Yuping (CHN)           0.92     1.86 14.21    1.61     6.40   38.60
## Hagger (GB)            1.38     1.80 12.75    1.14     6.34   35.76
## Brown (USA)            0.78     1.83 12.69    1.78     6.13   44.34
## Mulliner (GB)          0.46     1.71 12.68    1.69     6.10   37.76
## Hautenauve (BEL)       0.81     1.77 11.81    1.00     5.99   35.68
## Kytola (FIN)           0.54     1.77 11.66    0.92     5.75   39.48
## Geremias (BRA)         0.62     1.71 12.95    1.11     5.50   39.64
## Hui-Ing (TAI)          0.00     1.68 10.00    1.38     5.47   39.14
## Jeong-Mi (KOR)         0.32     1.71 10.83    0.00     5.50   39.26
##                     run800m score
## Joyner-Kersee (USA)   18.16  7291
## John (GDR)            20.55  6897
## Behmer (GDR)          22.47  6858
## Sablovskaite (URS)    14.43  6540
## Choubenkova (URS)     18.77  6540
```

```
## Schulz (GDR)          20.88  6411
## Fleming (AUS)         14.13  6351
## Greiner (USA)         13.02  6297
## Lajbnerova (CZE)      10.62  6252
## Bouraga (URS)         11.93  6252
## Wijnsma (HOL)         15.18  6205
## Dimitrova (BUL)       14.13  6171
## Scheider (SWI)        11.74  6137
## Braun (FRG)            3.85  6109
## Ruotsalainen (FIN)     9.61  6101
## Yuping (CHN)           0.00  6087
## Hagger (GB)            8.19  5975
## Brown (USA)            0.24  5972
## Mulliner (GB)          8.65  5746
## Hautenauve (BEL)      12.77  5734
## Kytola (FIN)          13.32  5686
## Geremias (BRA)         2.65  5508
## Hui-Ing (TAI)          9.37  5290
## Jeong-Mi (KOR)         7.50  5289
```

```
datos=heptathlon[,-8]
biplot(pca, col = c('red', 'blue'), xlim =c(-0.5, 0.7))
```

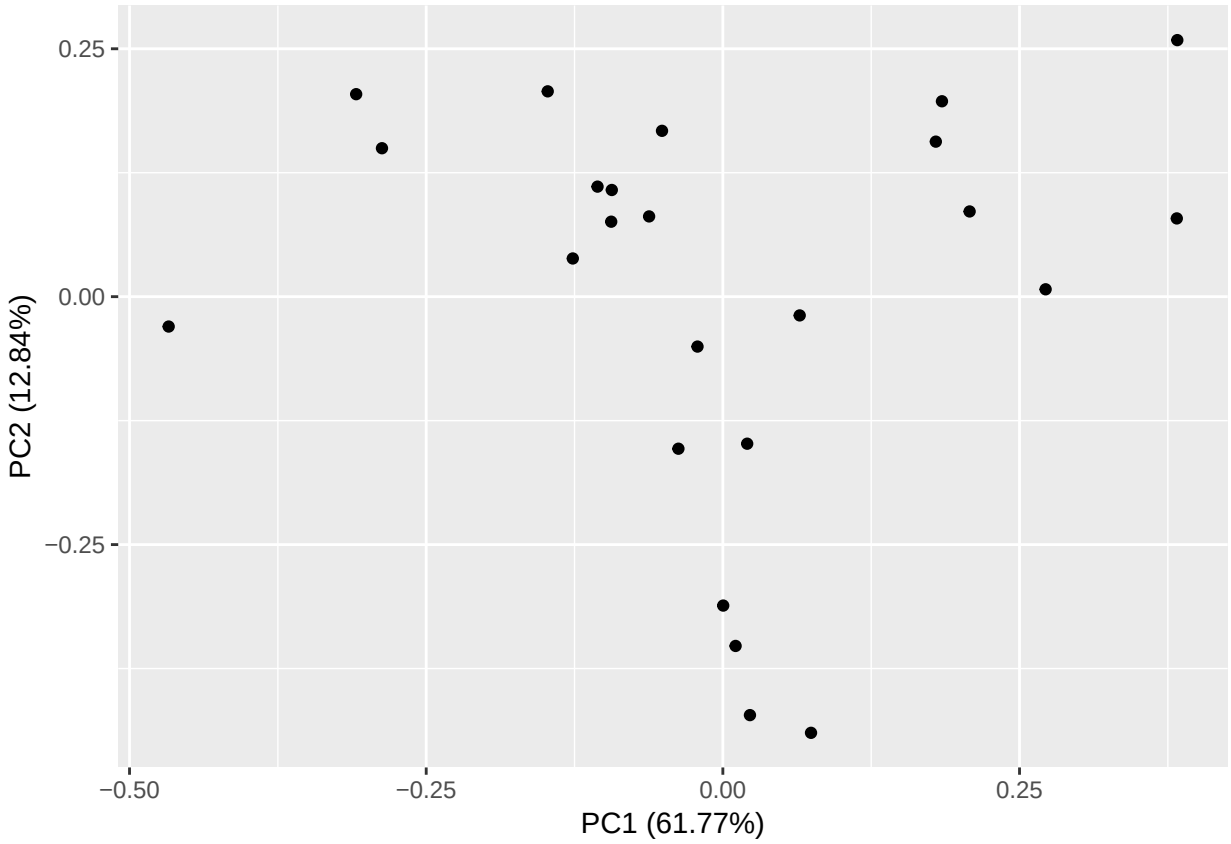



Podemos usar ggplot

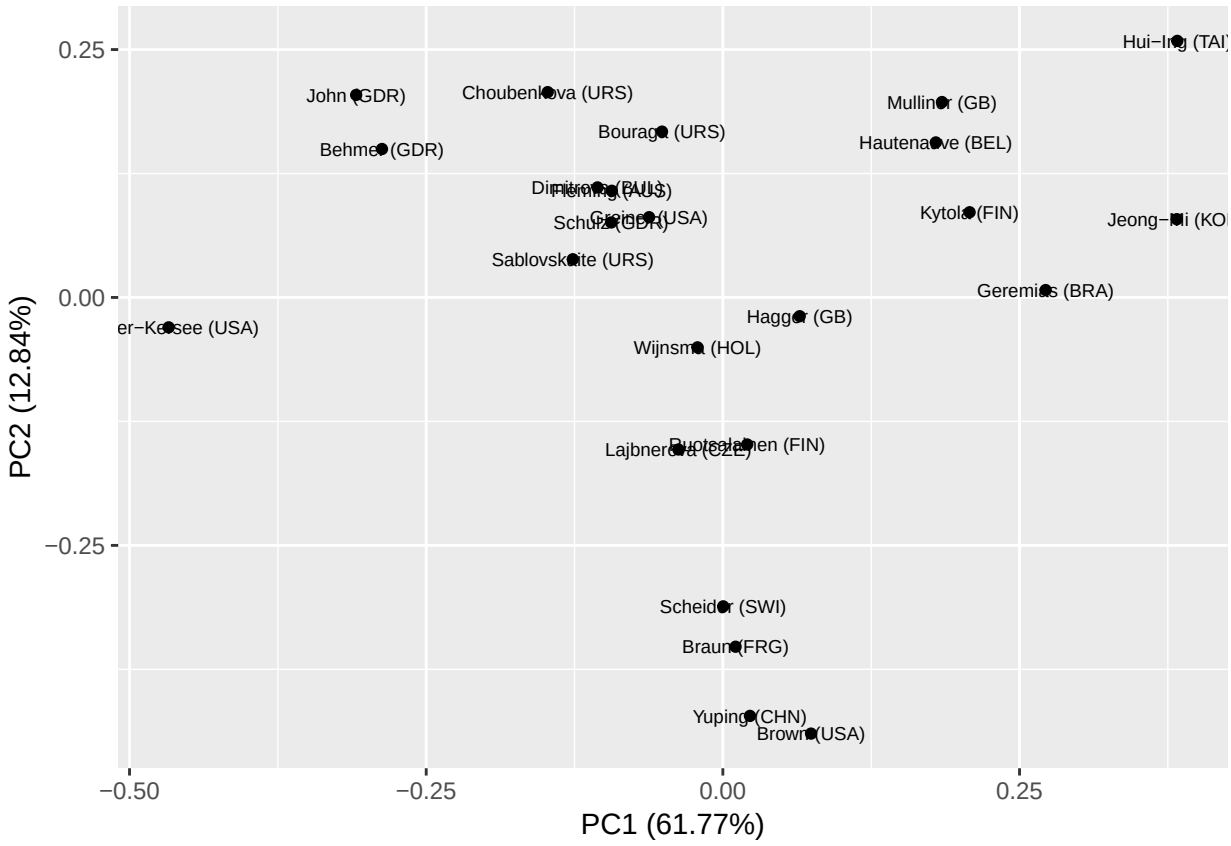
```
if ( !require( ggfortify ) )
{
    chooseCRANmirror( graphics=F )
    install.packages( 'ggfortify' )
    library( 'ggfortify' )
}
```

```
## Loading required package: ggfortify
## Loading required package: ggplot2
```

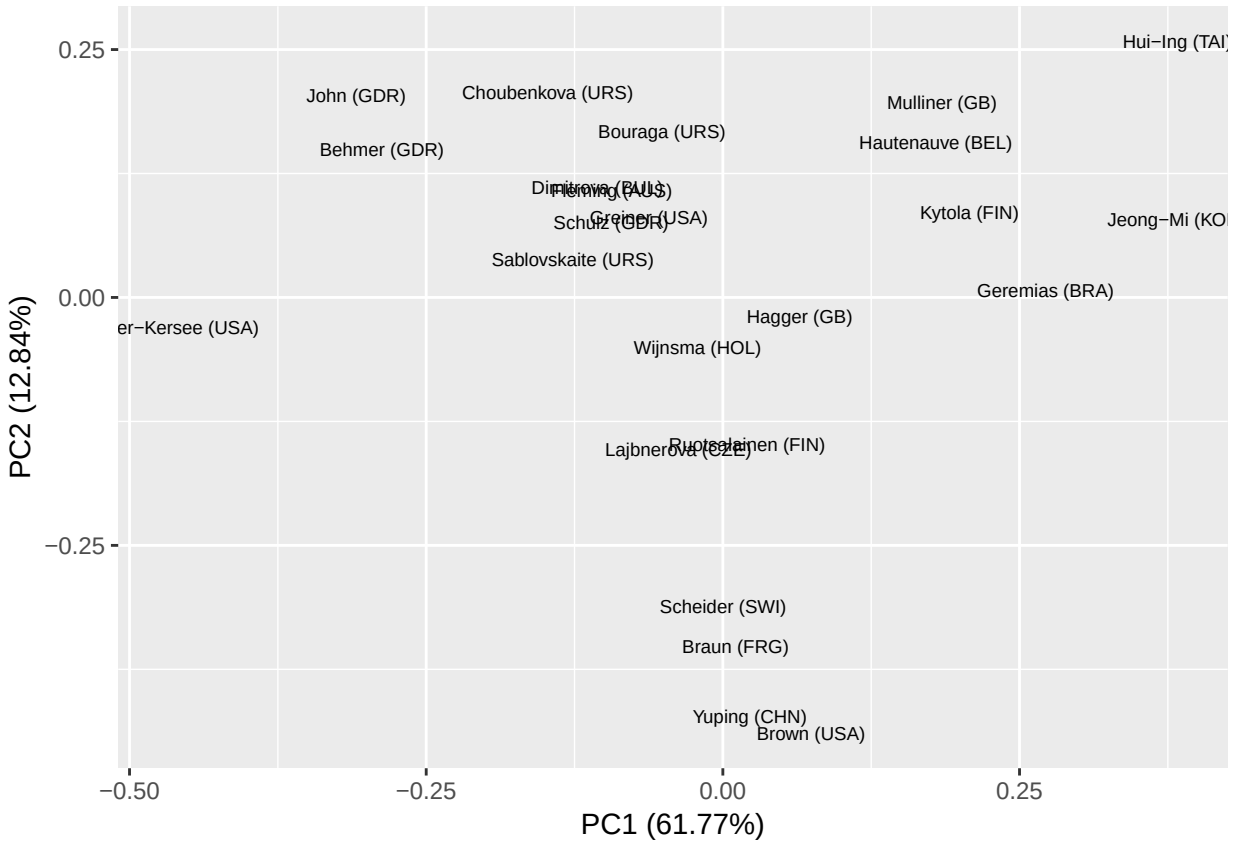
```
autoplot( pca )
```



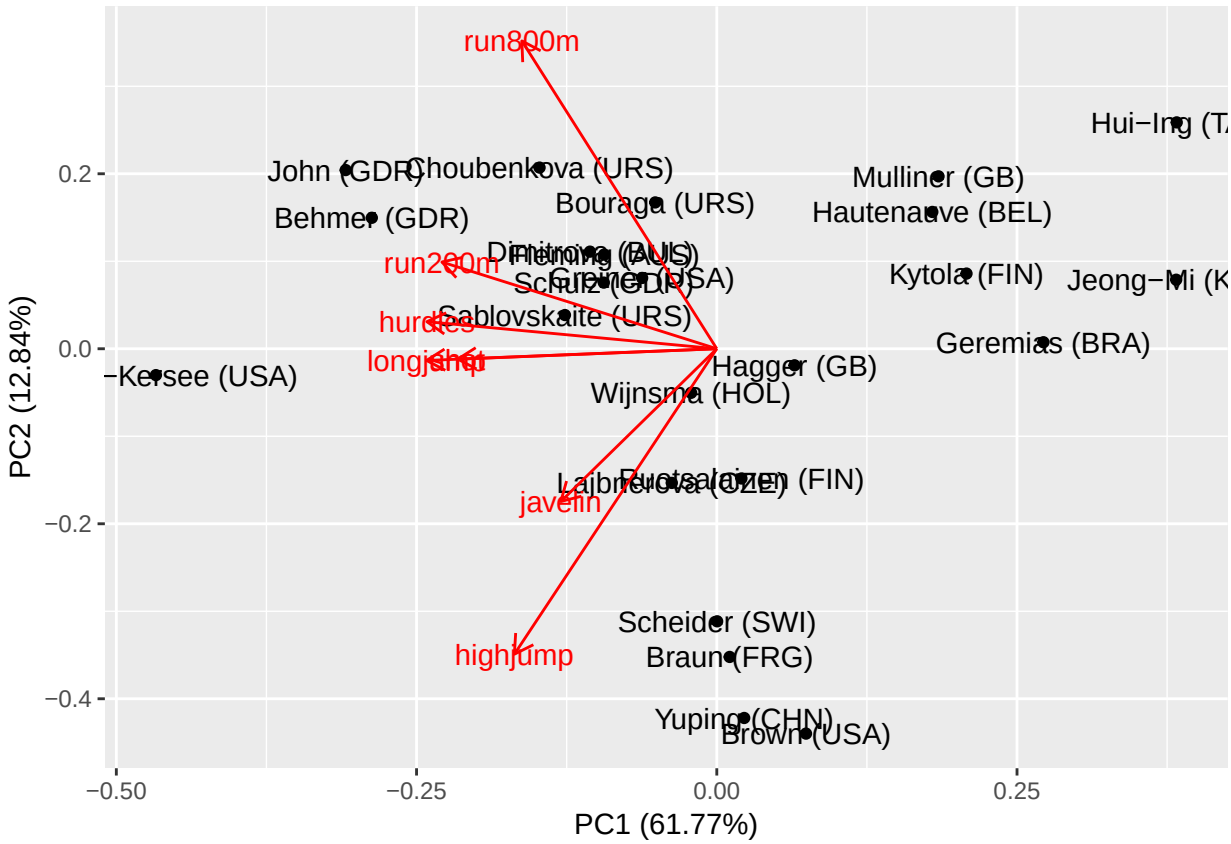

```
autoplot( pca, label=T, label.size=2.5 )
```




```
autoplot( pca, label=T, label.size=2.5, shape=F )
```




```
autoplot( pca, label=T, loadings.label = TRUE )
```




```
## Mas graficos en
## http://rpubs.com/sinhrks/plot_pca
## http://www.sthda.com/english/wiki/factominer-and-factoextra-principal-component-analysis-visualizati
```

## 2- Con FactoMineR, François Husson, 2012.

Cuando me va importar mucho el contenido gr?fico.

```
if ( !require( FactoMineR ) )
{
    chooseCRANmirror( graphics=F )
    install.packages( 'ggfortify' )
    library(FactoMineR)
}
```

```
## Loading required package: FactoMineR
```

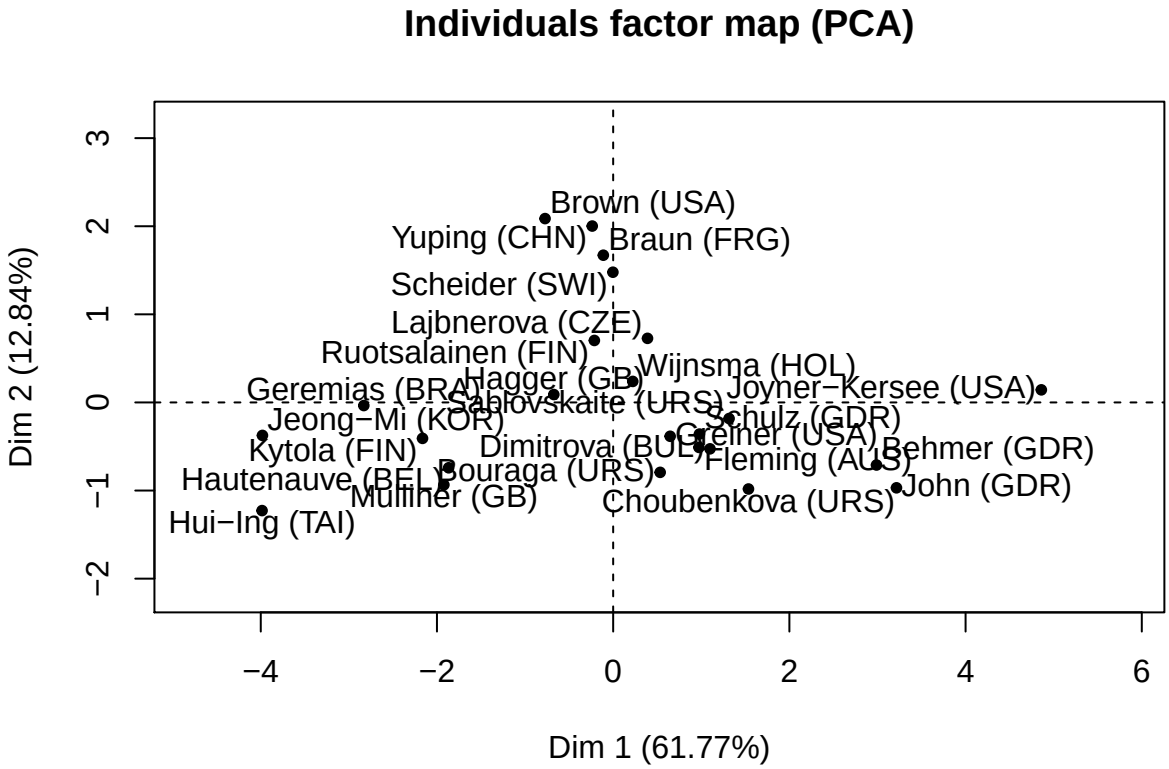
```
## La libreria factoextra grafica los objetos de FactoMineR en ggplot
if ( !require( factoextra ) )
{
    chooseCRANmirror( graphics=F )
    install.packages( 'factoextra' )
    library(factoextra)
}
```

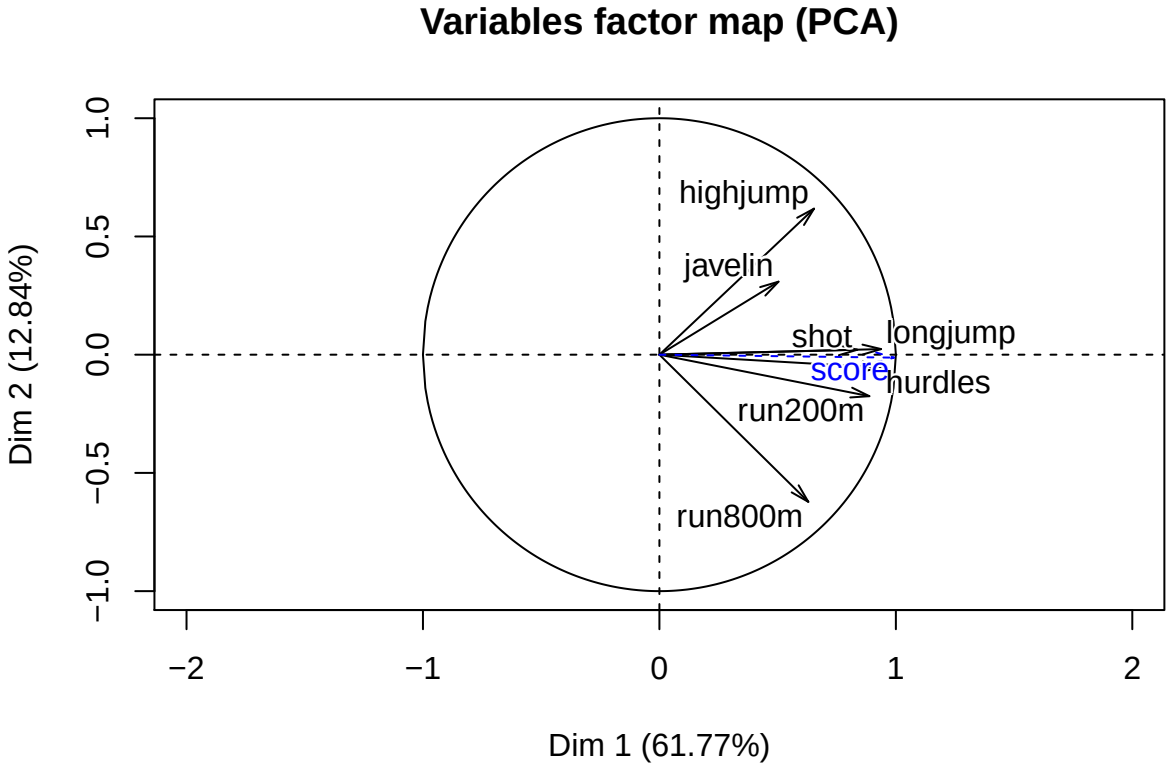
```
## Loading required package: factoextra
```

```
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

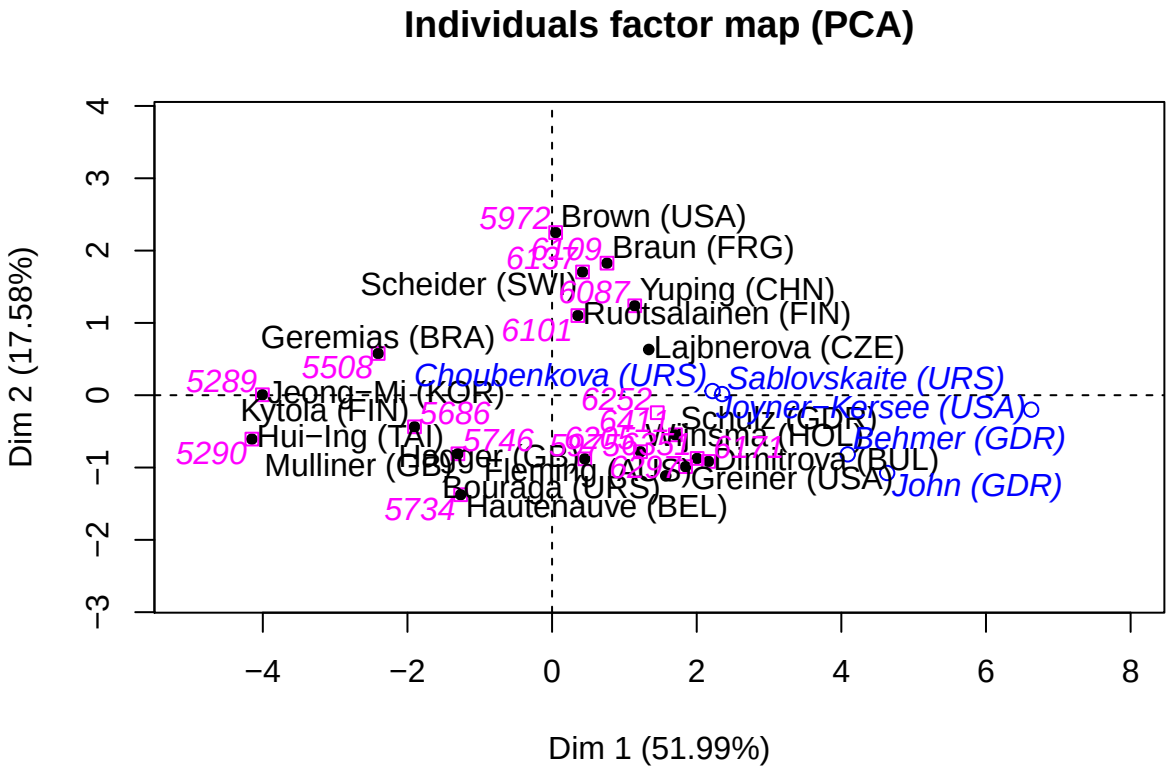
Por defecto FactomineR trabaja con la matriz de correlaciones (var/cov estandarizada).

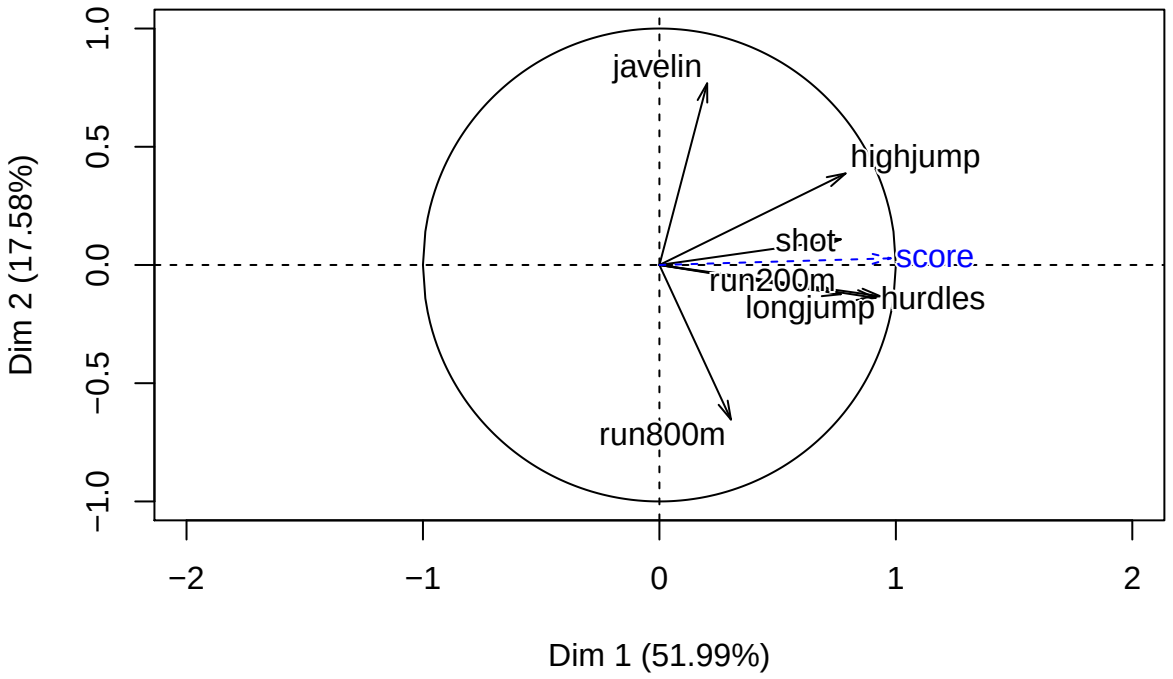
```
pca2=PCA(heptathlon,quanti.sup=8)
```



```
pca22=PCA(heptathlon,quali.sup=8,quanti.sup=8,ind.sup=1:5)
```



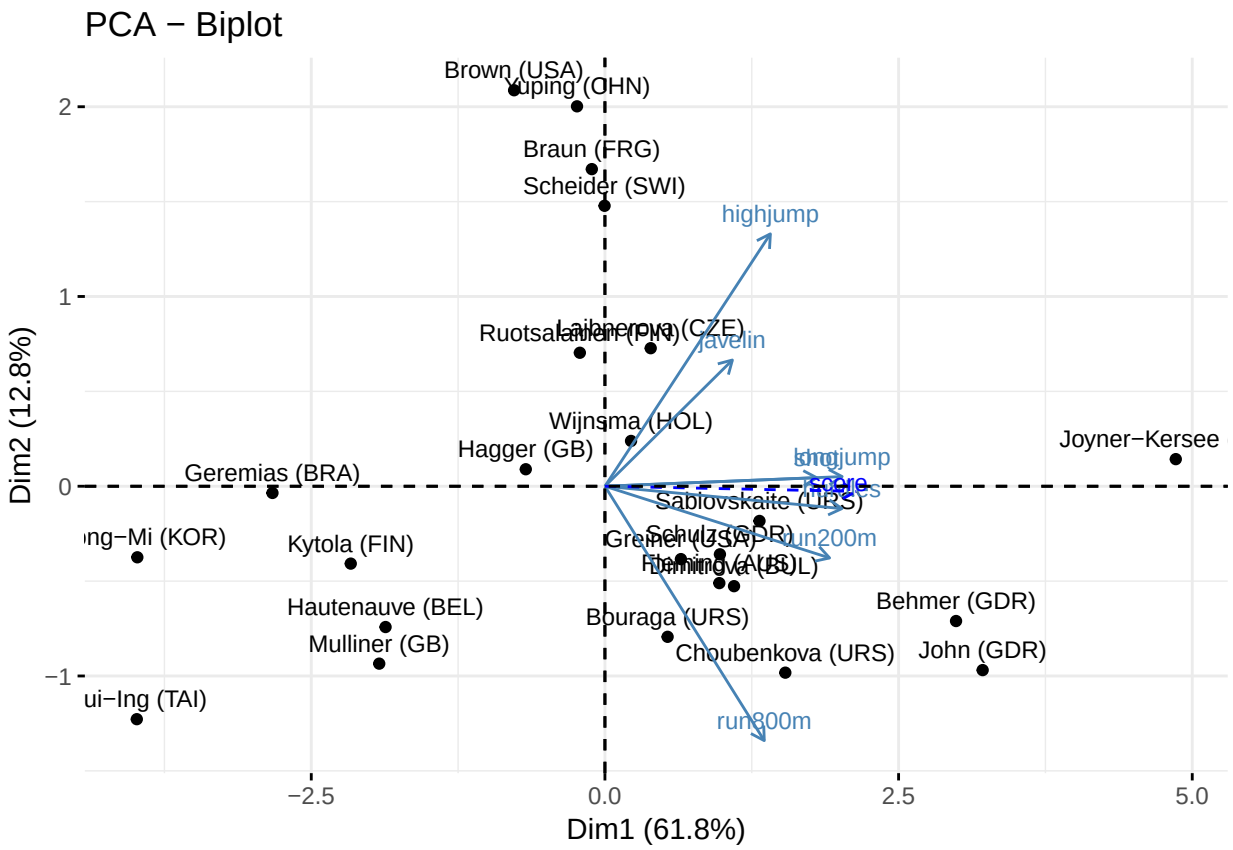

```
summary(pca2)
```

```
##
```

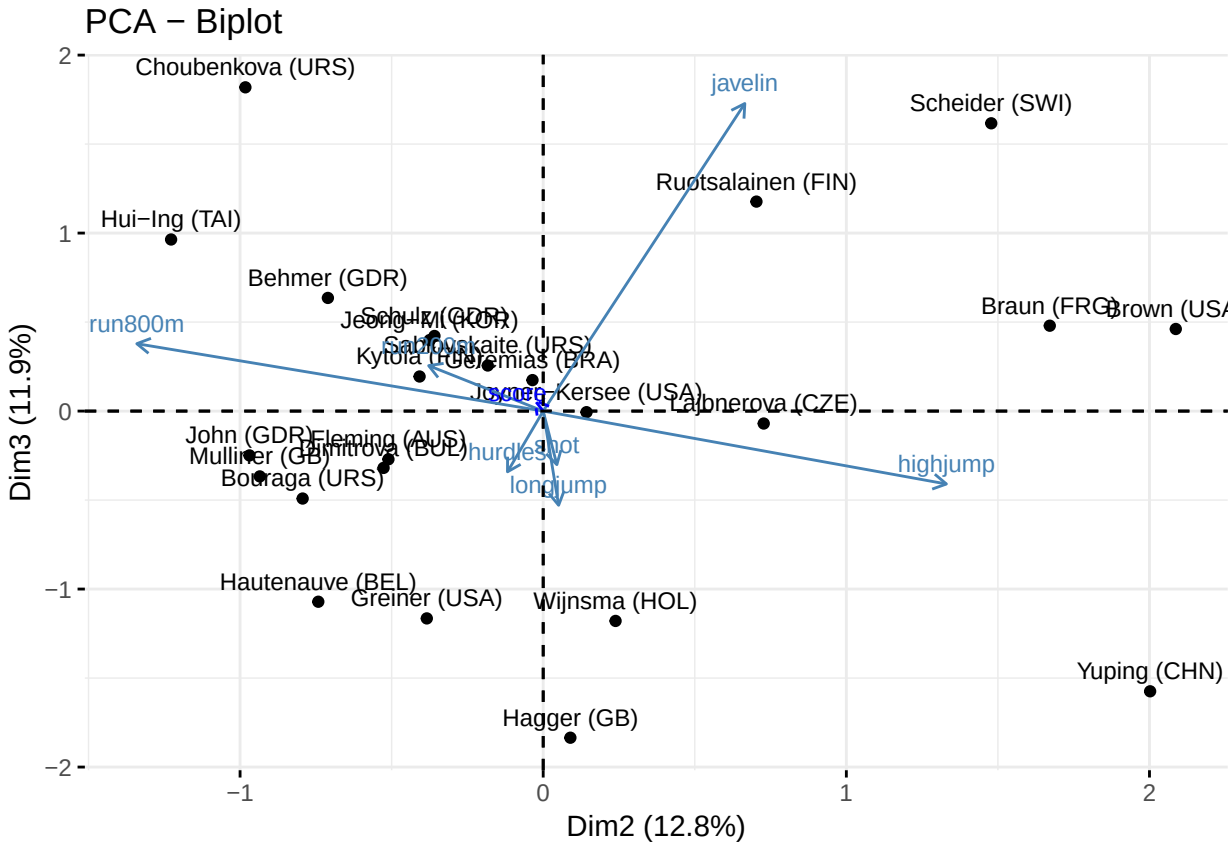
```
## Call:
## PCA(X = heptathlon, quanti.sup = 8)
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
## Variance               4.324   0.899   0.830   0.467   0.298   0.114
## % of var.             61.766  12.843  11.853   6.668   4.262   1.627
## Cumulative % of var.  61.766  74.609  86.463  93.131  97.392  99.019
##                        Dim.7
## Variance               0.069
## % of var.              0.981
## Cumulative % of var. 100.000
##
## Individuals (the 10 first)
##                         Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## Joyner-Kersee (USA) |  4.917 |  4.860 22.761  0.977 |  0.143  0.095  0.001
## John (GDR)          |  3.521 |  3.216  9.965  0.834 | -0.969  4.352  0.076
## Behmer (GDR)        |  3.207 |  2.989  8.610  0.869 | -0.710  2.338  0.049
## Sablovskaite (URS)  |  1.725 |  1.316  1.669  0.582 | -0.183  0.155  0.011
## Choubenkova (URS)   |  2.782 |  1.536  2.273  0.305 | -0.982  4.474  0.125
## Schulz (GDR)        |  1.778 |  0.979  0.924  0.303 | -0.359  0.597  0.041
## Fleming (AUS)       |  1.464 |  0.974  0.914  0.443 | -0.511  1.208  0.122
## Greiner (USA)       |  1.419 |  0.647  0.403  0.208 | -0.384  0.683  0.073
## Lajbnerova (CZE)    |  1.170 |  0.390  0.146  0.111 |  0.727  2.453  0.386
## Bouraga (URS)       |  1.708 |  0.534  0.274  0.098 | -0.794  2.919  0.216
##                        Dim.3    ctr   cos2
## Joyner-Kersee (USA) | -0.006  0.000  0.000 |
## John (GDR)          | -0.249  0.312  0.005 |
## Behmer (GDR)        |  0.636  2.029  0.039 |
## Sablovskaite (URS)  |  0.256  0.329  0.022 |
## Choubenkova (URS)   |  1.819 16.613  0.428 |
## Schulz (GDR)        |  0.422  0.894  0.056 |
## Fleming (AUS)       | -0.271  0.368  0.034 |
## Greiner (USA)       | -1.165  6.814  0.674 |
## Lajbnerova (CZE)    | -0.070  0.025  0.004 |
## Bouraga (URS)       | -0.491  1.213  0.083 |
##
## Variables
##                        Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3
## hurdles             |  0.937 20.285  0.877 | -0.055  0.333  0.003 | -0.158
## highjump            |  0.654  9.892  0.428 |  0.618 42.423  0.381 | -0.190
## shot                |  0.837 16.200  0.700 |  0.021  0.048  0.000 | -0.140
## run200m             |  0.888 18.240  0.789 | -0.175  3.424  0.031 |  0.119
## longjump            |  0.938 20.337  0.879 |  0.024  0.062  0.001 | -0.246
## javelin             |  0.504  5.871  0.254 |  0.309 10.610  0.095 |  0.802
## run800m             |  0.630  9.175  0.397 | -0.622 43.100  0.387 |  0.176
##                        ctr   cos2
## hurdles              3.025  0.025 |
## highjump             4.361  0.036 |
## shot                 2.355  0.020 |
## run200m              1.693  0.014 |
## longjump             7.277  0.060 |
## javelin             77.563  0.644 |
```

```
## run800m              3.725  0.031 |
##
## Supplementary continuous variable
##                        Dim.1   cos2    Dim.2   cos2    Dim.3   cos2
## score                | 0.993  0.986 | -0.013  0.000 |  0.027  0.001 |
```

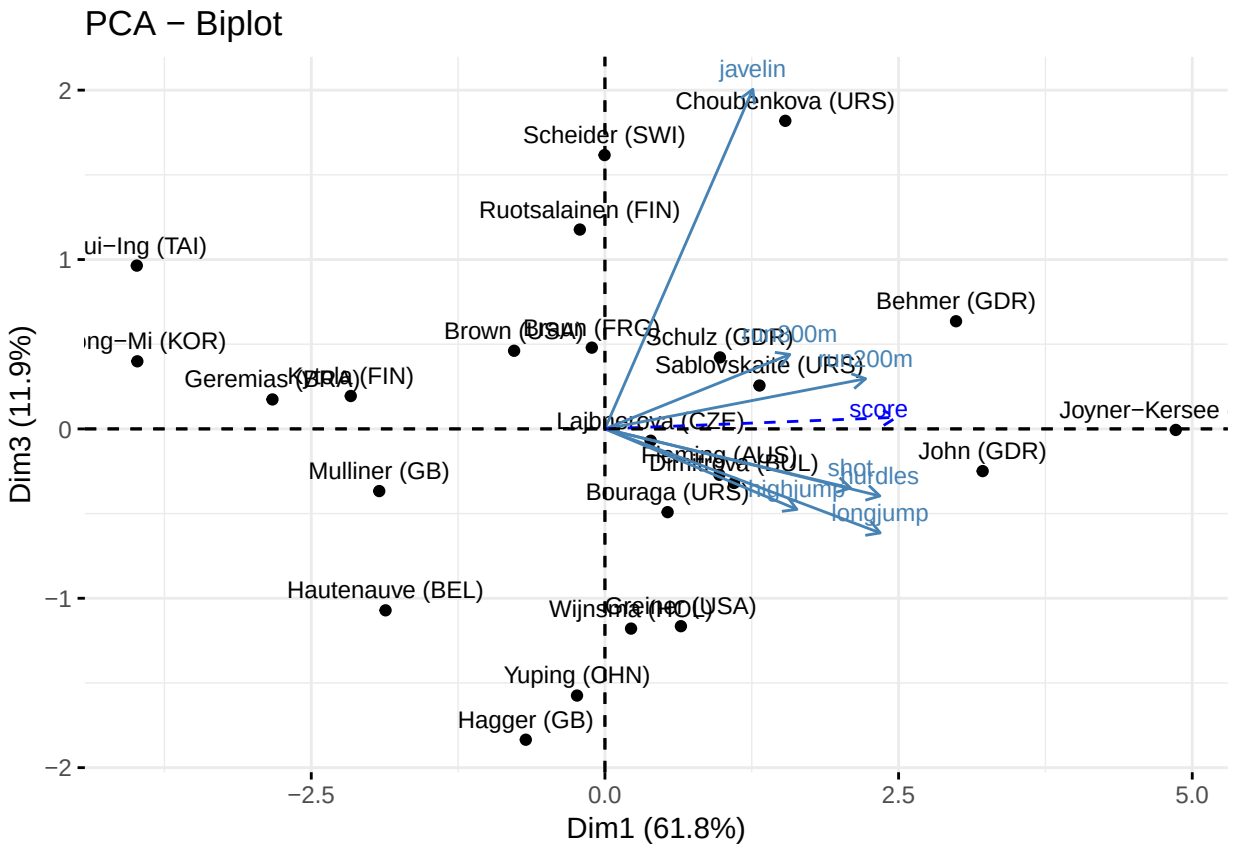
```
fviz_pca( pca2, labelsize=3 )
```



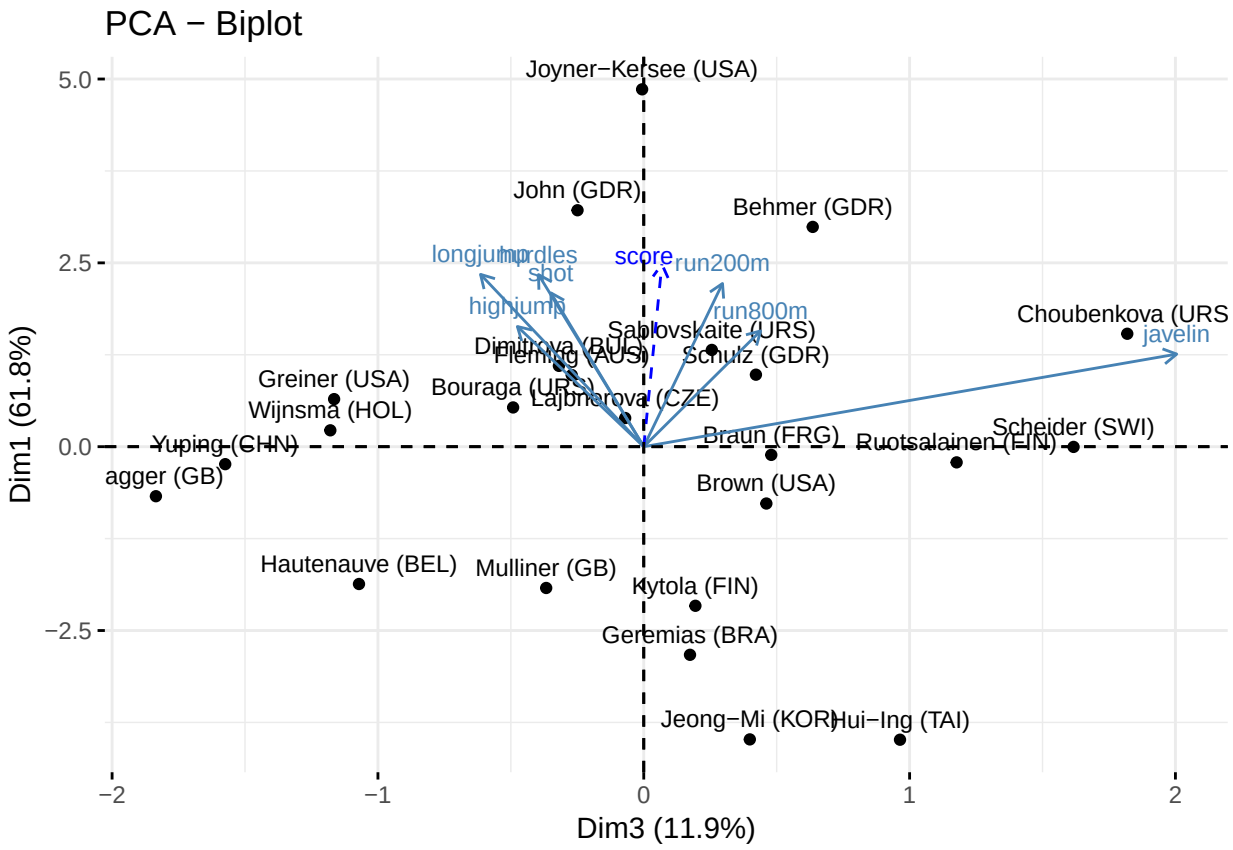

```
fviz_pca( pca2, axes=c(2, 3), labelsize=3 )
```



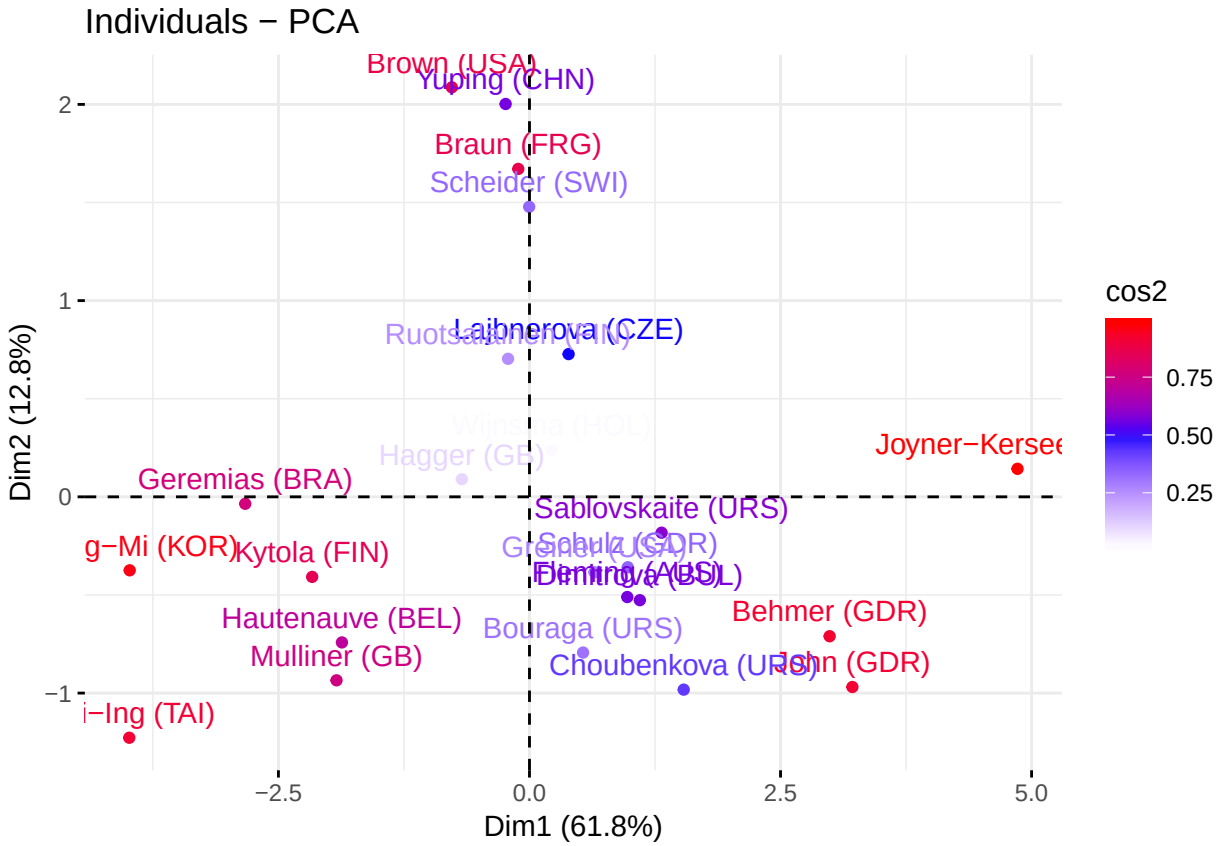

```
fviz_pca( pca2, axes=c(1, 3), labelsize=3 )
```




```
fviz_pca( pca2, axes=c(3, 1), labelsize=3 )
```




```
fviz_pca_ind( pca2, col.ind="cos2") +
scale_color_gradient2(low="white", mid="blue",
                      high="red", midpoint=0.50)
```

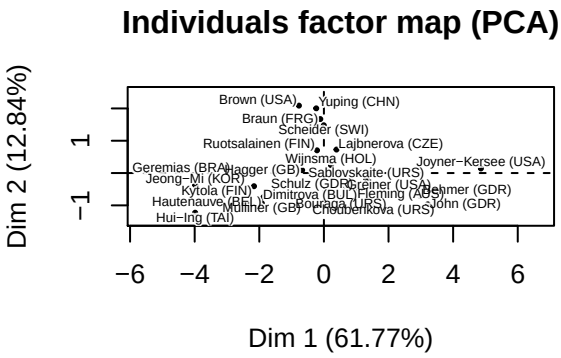


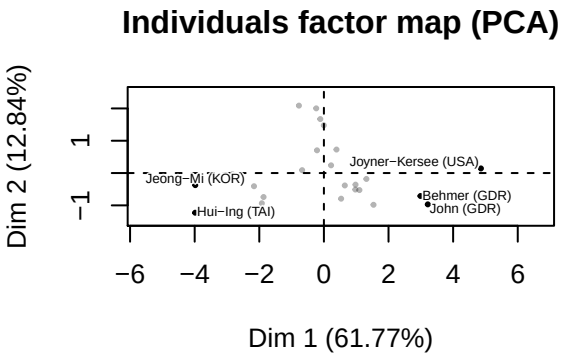

Aca obtengo los valores propios, el porcentage de variación. Comparo con valores propios encontrados anteriormente

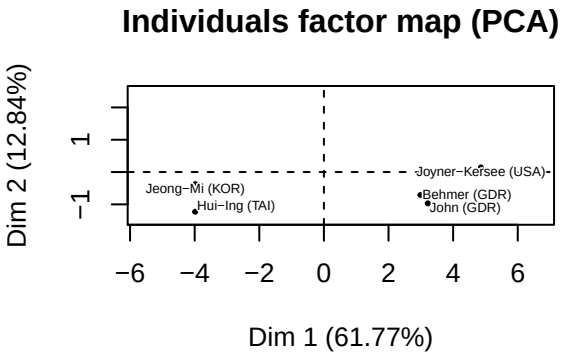
```
(pca$sdev)^2
```

```
## [1] 4.32364217 0.89899445 0.82974172 0.46675769 0.29832218 0.11387578
## [7] 0.06866602
```

```
summary(pca2,nbelements=Inf,file="prueba.txt")
par( mfrow=c(2, 2) )
plot(pca2,shadow=T,cex=0.5)
plot(pca2,shadow=T,cex=0.5,select="cos2 0.89")
plot(pca2,shadow=T,cex=0.5,select="cos2 0.89",unselect=1)
```

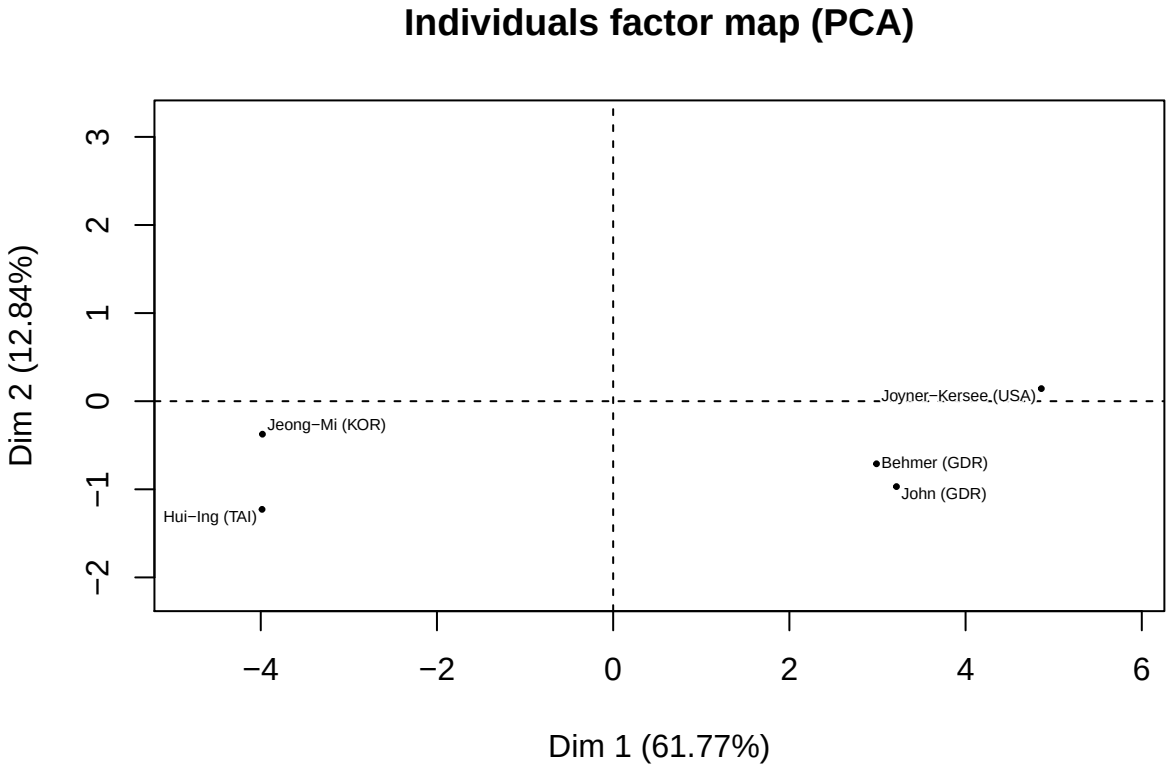


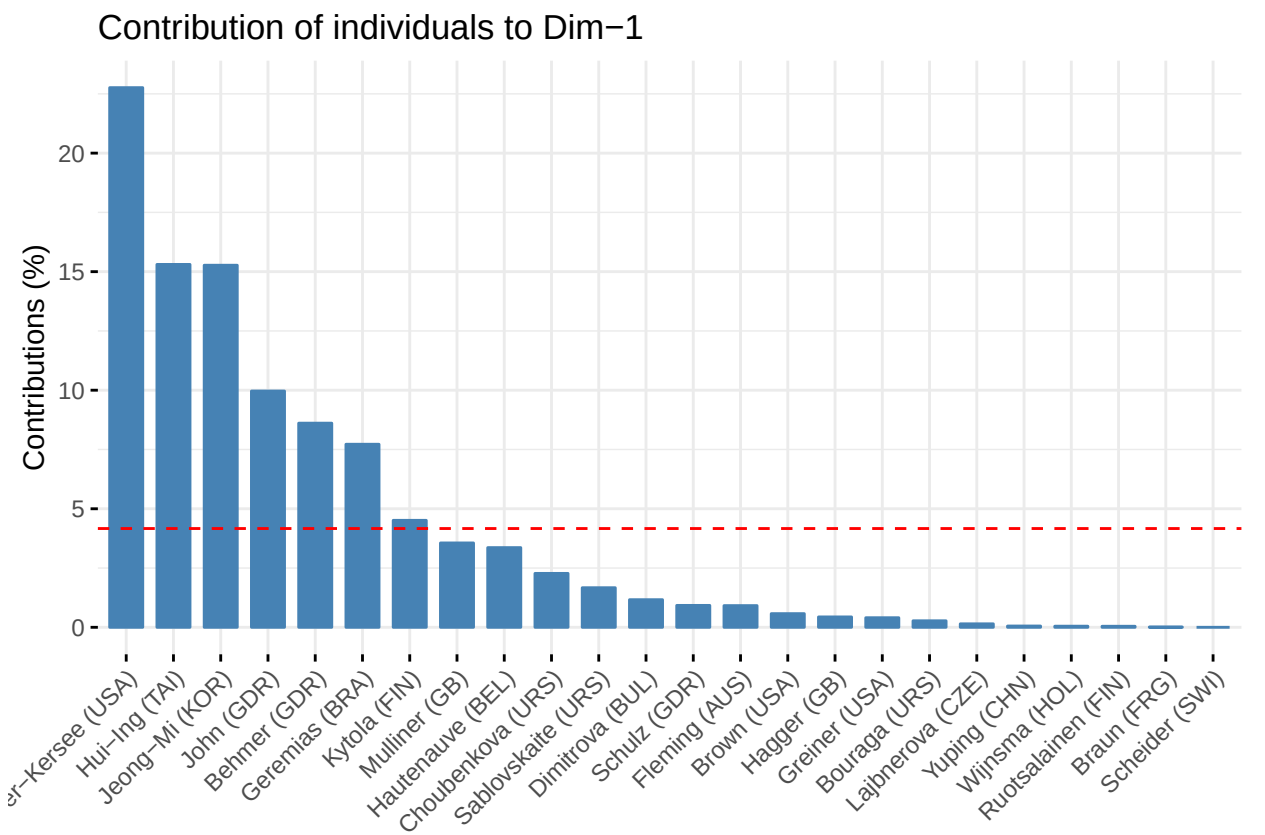




Los individuos que tienen mayor contribución

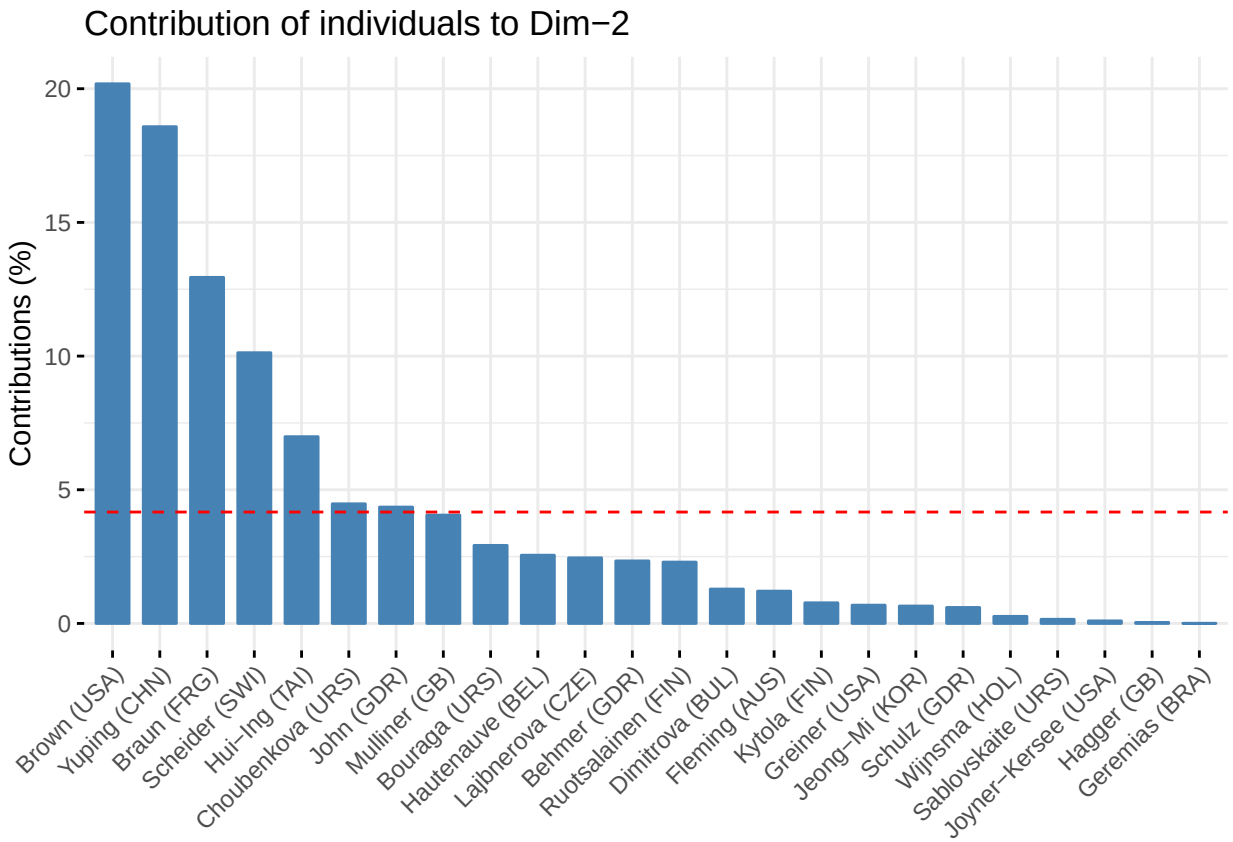
```
plot(pca2,shadow=T,cex=0.5,select="contrib 5",unselect=1)
```



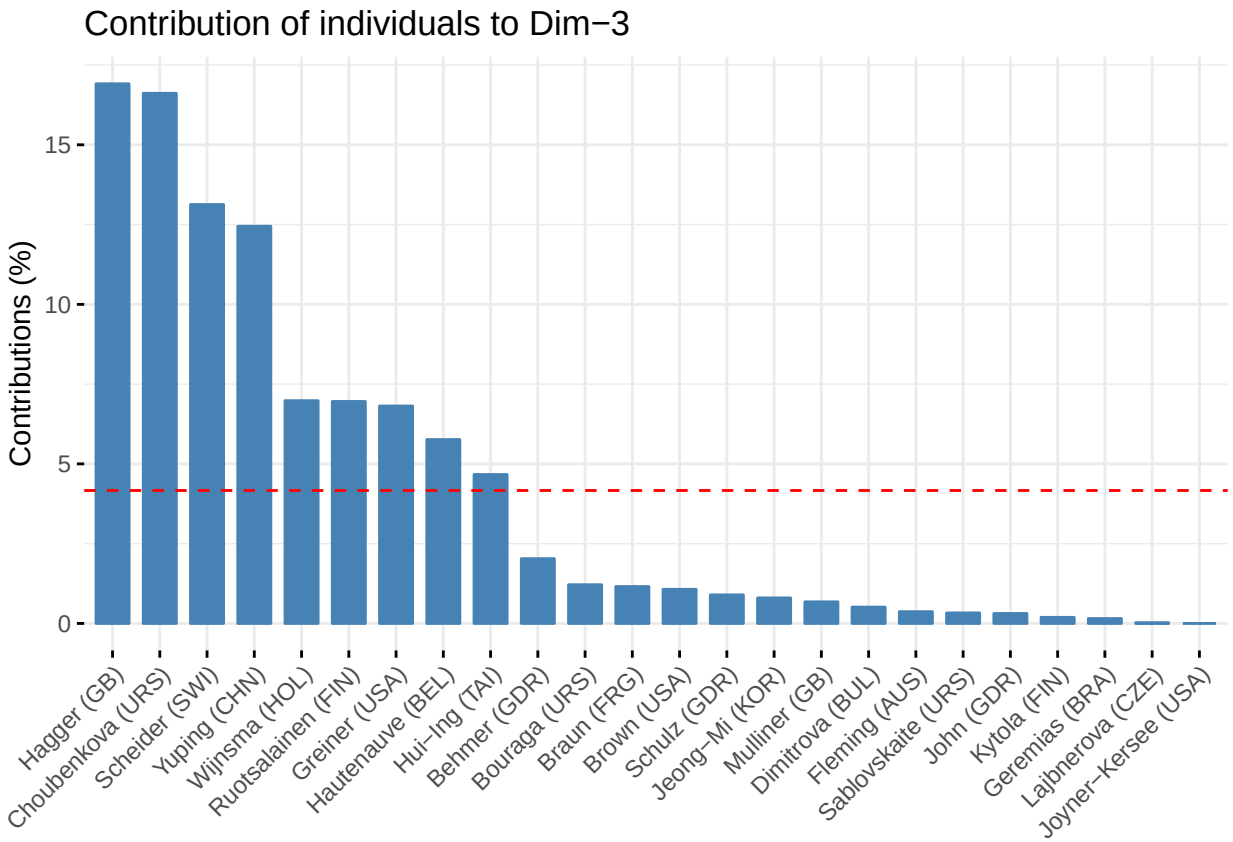
```
## Graficamente con ggfortify
fviz_contrib(pca2, choice="ind",axes=1)
```



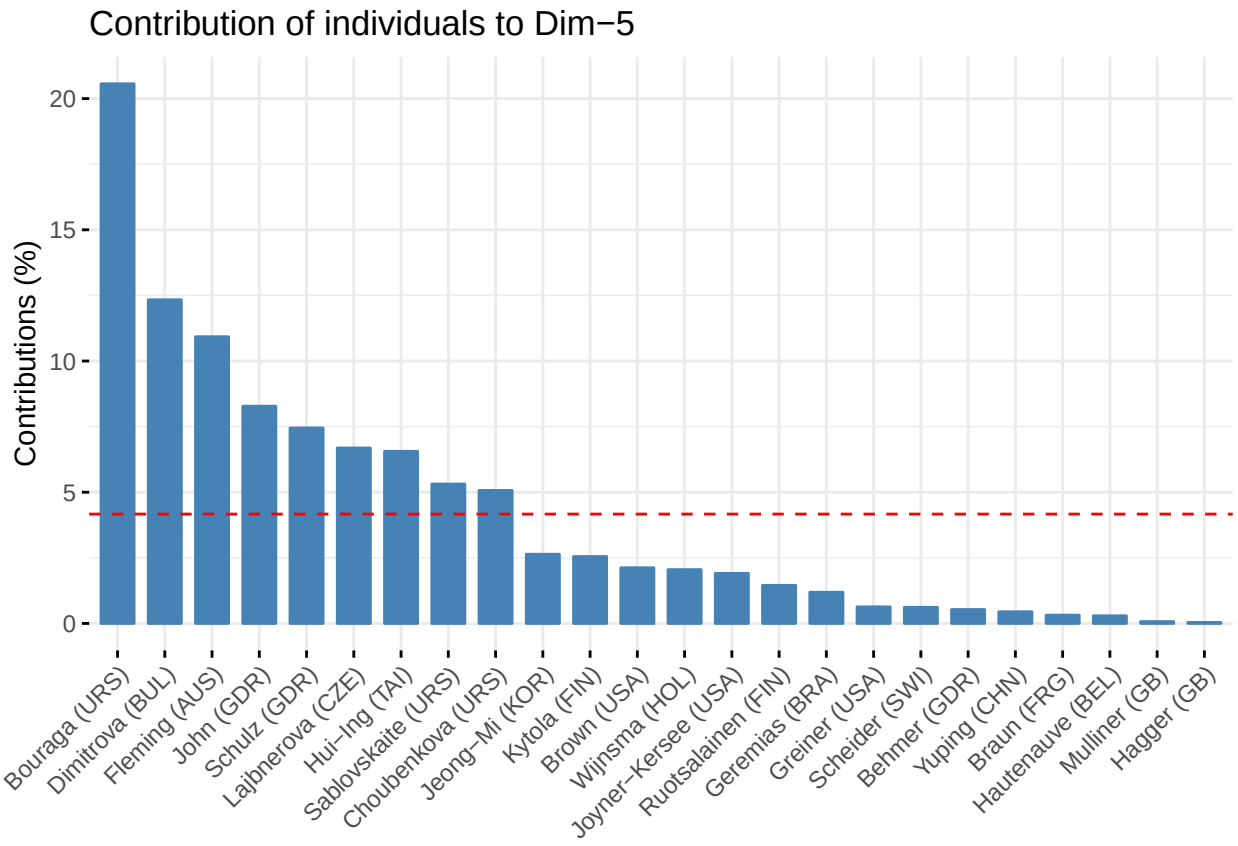

```
fviz_contrib(pca2, choice="ind", axes=2)
```




```
fviz_contrib(pca2, choice="ind", axes=3)
```

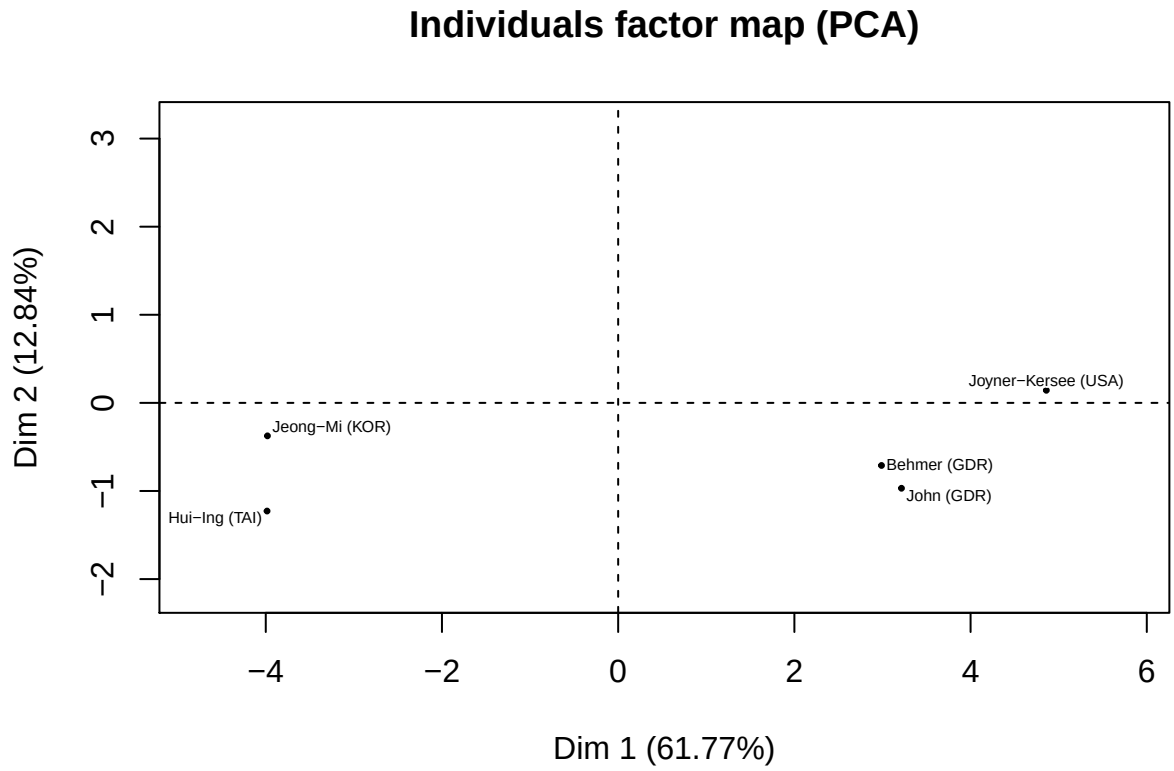



```
fviz_contrib(pca2, choice="ind", axes=5)
```

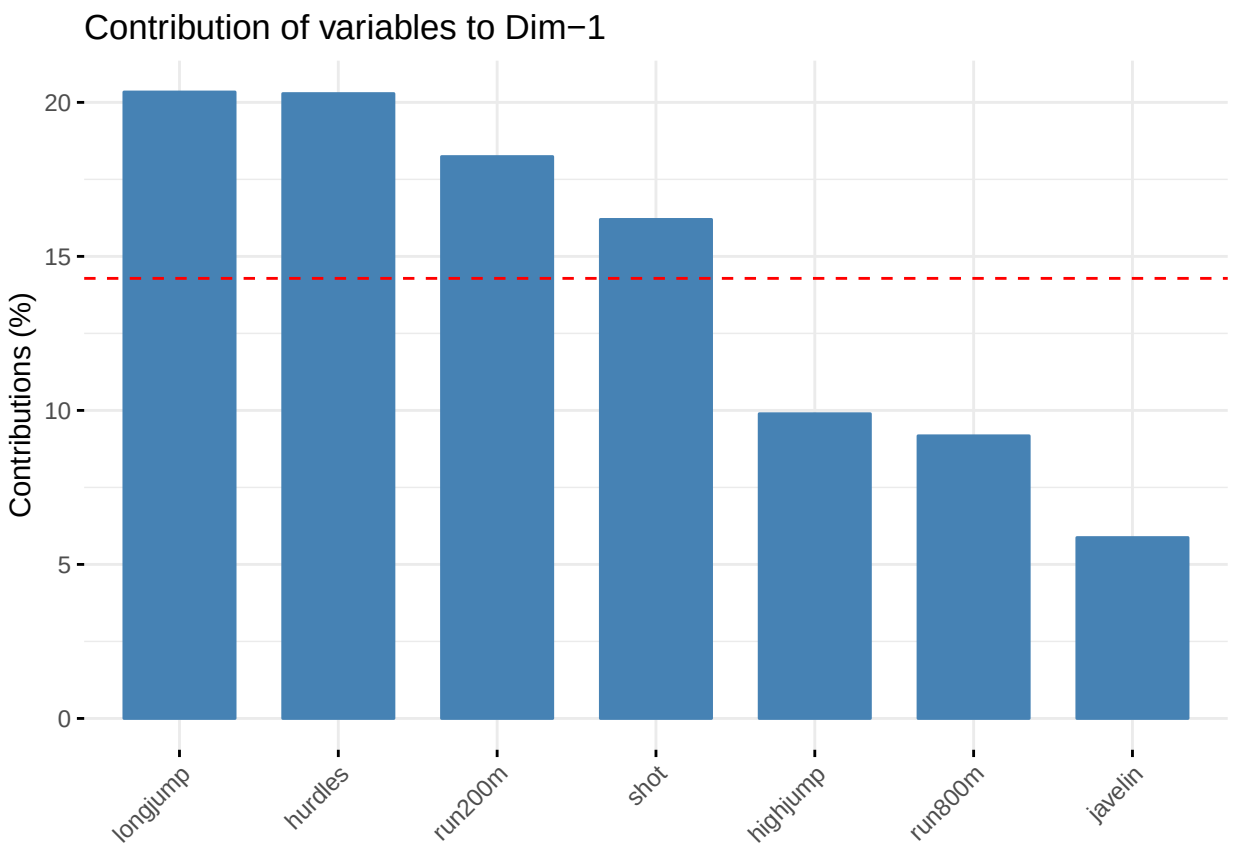



Las variables que tienen mayor contribución

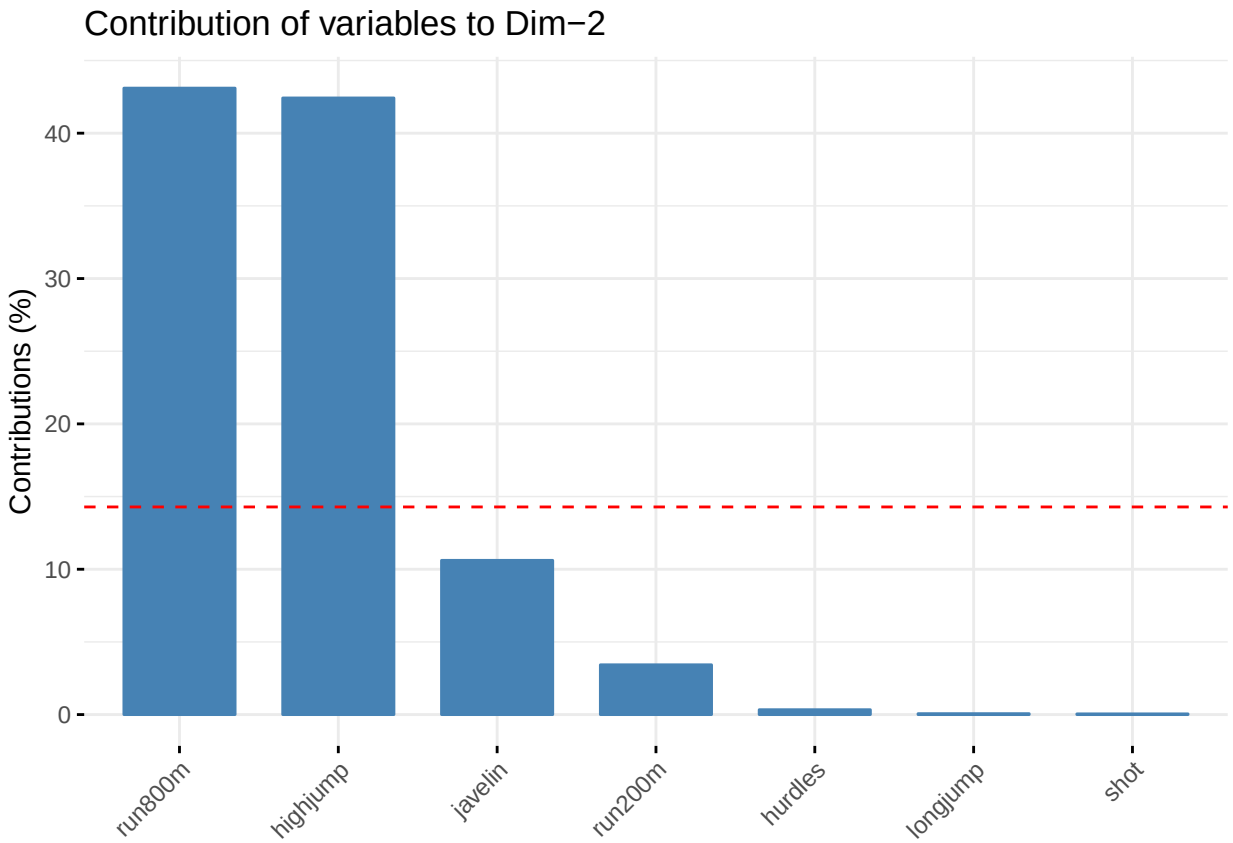
```
plot(pca2,shadow=T,cex=0.5,select="contrib 5",unselect=1)
```



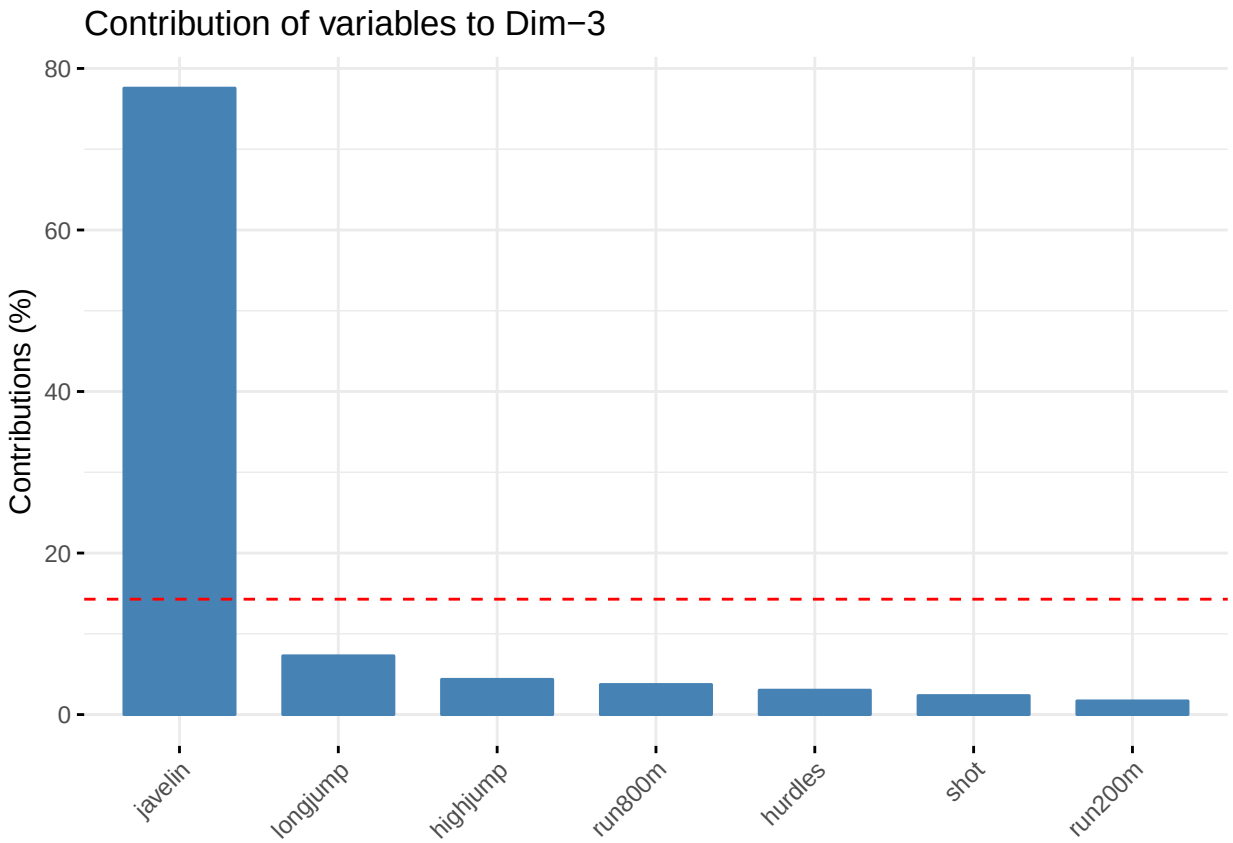
```
## Graficamente con ggfortify
fviz_contrib(pca2, choice="var",axes=1)
```




```
fviz_contrib(pca2, choice="var", axes=2)
```




```
fviz_contrib(pca2, choice="var", axes=3)
```




Los 5 primeros valores propios

```
round(pca2$eig,2)
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1       4.32                  61.77                             61.77
## comp 2       0.90                  12.84                             74.61
## comp 3       0.83                  11.85                             86.46
## comp 4       0.47                   6.67                             93.13
## comp 5       0.30                   4.26                             97.39
## comp 6       0.11                   1.63                             99.02
## comp 7       0.07                   0.98                            100.00
```

La distancia individuos a centro de gravedad

```
round(pca2$ind$dist,2)
```

```
## Joyner-Kersee (USA)          John (GDR)        Behmer (GDR)
##                4.92               3.52               3.21
##  Sablovskaite (URS)   Choubenkova (URS)        Schulz (GDR)
##                1.73               2.78               1.78
##       Fleming (AUS)       Greiner (USA)    Lajbnerova (CZE)
##                1.46               1.42               1.17
##       Bouraga (URS)       Wijnsma (HOL)     Dimitrova (BUL)
##                1.71               1.84               1.62
##      Scheider (SWI)         Braun (FRG)  Ruotsalainen (FIN)
##                2.55               1.81               1.45
##        Yuping (CHN)         Hagger (GB)         Brown (USA)
##                2.69               2.10               2.38
```

```
##       Mulliner (GB)    Hautenauve (BEL)        Kytola (FIN)
##                2.46                2.40                2.39
##      Geremias (BRA)       Hui-Ing (TAI)      Jeong-Mi (KOR)
##                3.23                4.37                4.08
```

Las contributiones de los individuos a las dimensiones

```
round(pca2$ind$contrib[ ,1:2],2)
```

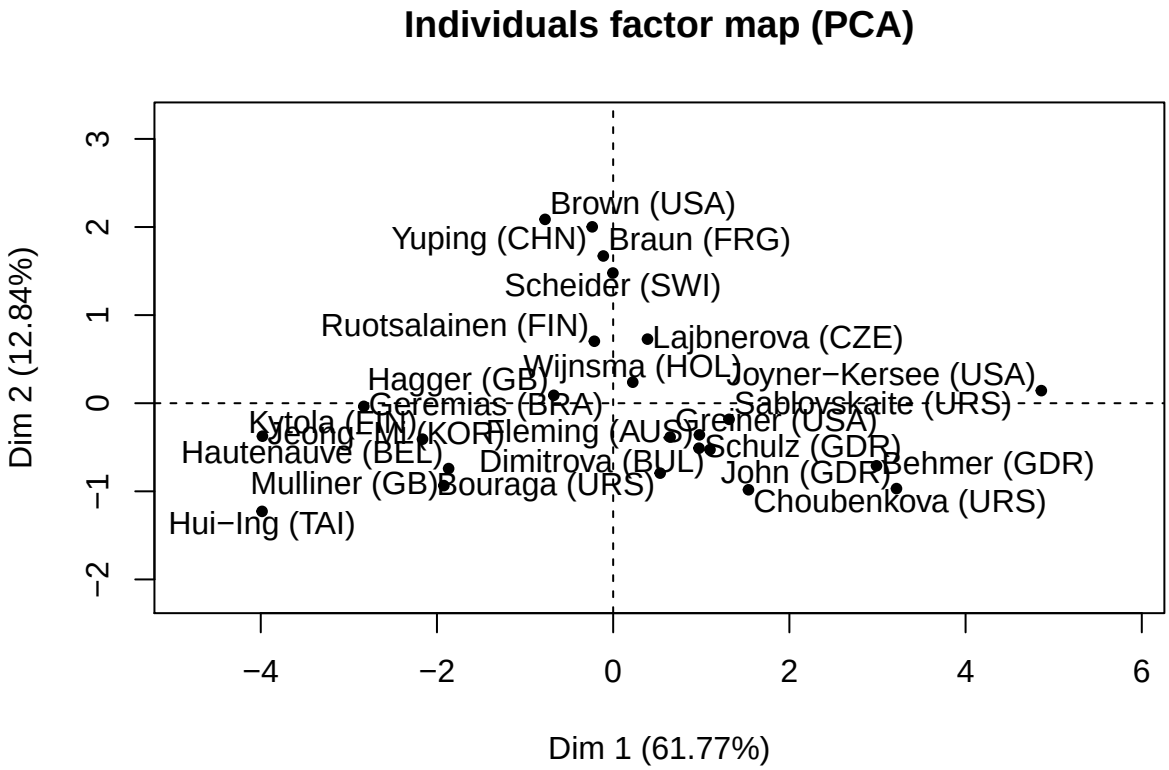
```
##                     Dim.1 Dim.2
## Joyner-Kersee (USA) 22.76  0.09
## John (GDR)           9.96  4.35
## Behmer (GDR)         8.61  2.34
## Sablovskaite (URS)   1.67  0.15
## Choubenkova (URS)    2.27  4.47
## Schulz (GDR)         0.92  0.60
## Fleming (AUS)        0.91  1.21
## Greiner (USA)        0.40  0.68
## Lajbnerova (CZE)     0.15  2.45
## Bouraga (URS)        0.27  2.92
## Wijnsma (HOL)        0.05  0.26
## Dimitrova (BUL)      1.16  1.29
## Scheider (SWI)       0.00 10.12
## Braun (FRG)          0.01 12.94
## Ruotsalainen (FIN)   0.04  2.29
## Yuping (CHN)         0.05 18.58
## Hagger (GB)          0.44  0.04
## Brown (USA)          0.58 20.18
## Mulliner (GB)        3.56  4.05
## Hautenauve (BEL)     3.36  2.55
## Kytola (FIN)         4.51  0.77
## Geremias (BRA)       7.72  0.01
## Hui-Ing (TAI)       15.30  6.98
## Jeong-Mi (KOR)      15.27  0.65
```

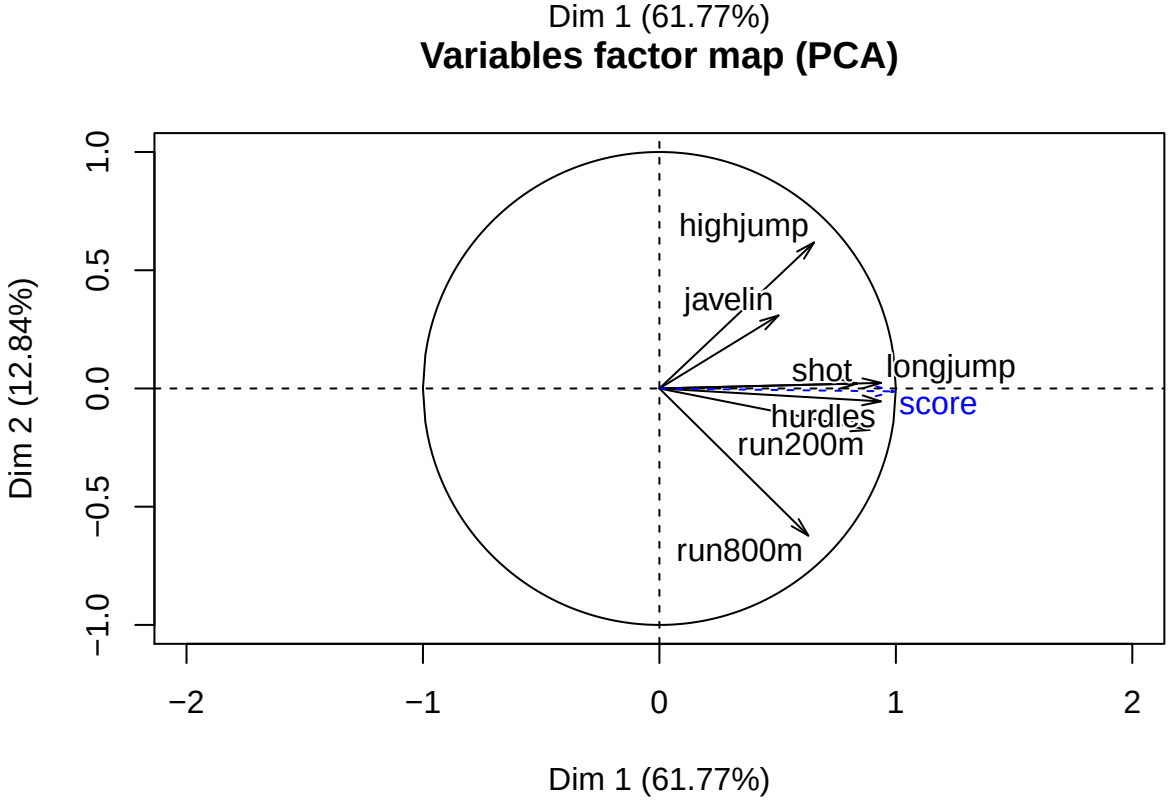
Las contribuciones de las variables a las dimensiones

```
round(pca2$var$contrib[ ,1:2],2)
```

```
##          Dim.1 Dim.2
## hurdles  20.28  0.33
## highjump  9.89 42.42
## shot     16.20  0.05
## run200m  18.24  3.42
## longjump 20.34  0.06
## javelin   5.87 10.61
## run800m   9.18 43.10
```

```
pca2=PCA(heptathlon,quanti.sup=8)
```


Individuals factor map (PCA)
Brown (USA)
Yuping (CHN)
Braun (FRG)
Scheider (SWI)
Ruotsalainen (FIN)
Lajbnerova (CZE)
Wijnsma (HOL)
Joyner−Kersee (USA)
Hagger (GB)
Geremias (BRA)
Sablovskaite (URS)
Kytola (FIN)
Jeong−Mi (KOR)
Fleming (AUS)
Greiner (USA)
Hautenauve (BEL)
Schulz (GDR)
Dimitrova (BUL)
Behmer (GDR)
Mulliner (GB)
Bouraga (URS)
John (GDR)
Choubenkova (URS)
Hui−Ing (TAI)
Dim 2 (12.84%)
Dim 1 (61.77%)


Variables factor map (PCA)
highjump
javelin
shot
longjump
score
hurdles
run200m
run800m
Dim 2 (12.84%)
Dim 1 (61.77%)