Analyse en Composantes Principales (ACP)

Analyse R

07 mars 2025

2 Données fictives

2.1 Création des données

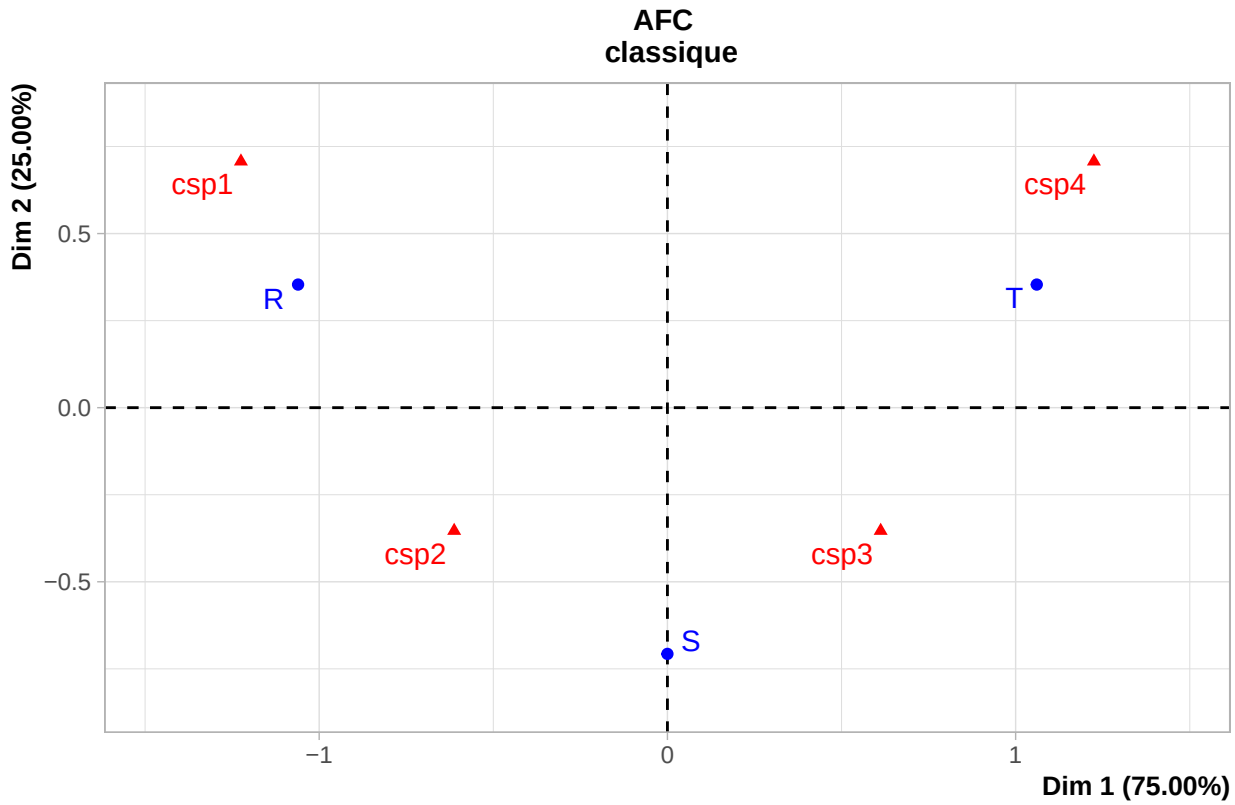
```
csp <- as.factor(c("csp3","csp1","csp2","csp2","csp3","csp4"))
sport <- as.factor(c("S","R","R","S","T","T"))
fic.tab <- table(sport,csp)
fic.conting <- data.frame(unclass(fic.tab))
# Paramètres graphiques #
x11()
par(mfrow=c(1,3))
```

2.2 AFC de la table de contingence

```
fic.afc <- CA(fic.conting,graph=FALSE)
fic.afc
```

```
## **Results of the Correspondence Analysis (CA)**
## The row variable has  3  categories; the column variable has 4 categories
## The chi square of independence between the two variables is equal to 6 (p-value =  0.4231901 ).
## *The results are available in the following objects:
##
##    name              description
## 1  "$eig"            "eigenvalues"
## 2  "$col"            "results for the columns"
## 3  "$col$coord"      "coord. for the columns"
## 4  "$col$cos2"       "cos2 for the columns"
## 5  "$col$contrib"    "contributions of the columns"
## 6  "$row"            "results for the rows"
## 7  "$row$coord"      "coord. for the rows"
## 8  "$row$cos2"       "cos2 for the rows"
## 9  "$row$contrib"    "contributions of the rows"
## 10 "$call"           "summary called parameters"
## 11 "$call$marge.col" "weights of the columns"
## 12 "$call$marge.row" "weights of the rows"
```

```r
plot(fic.afc,title="AFC \n classique")
```



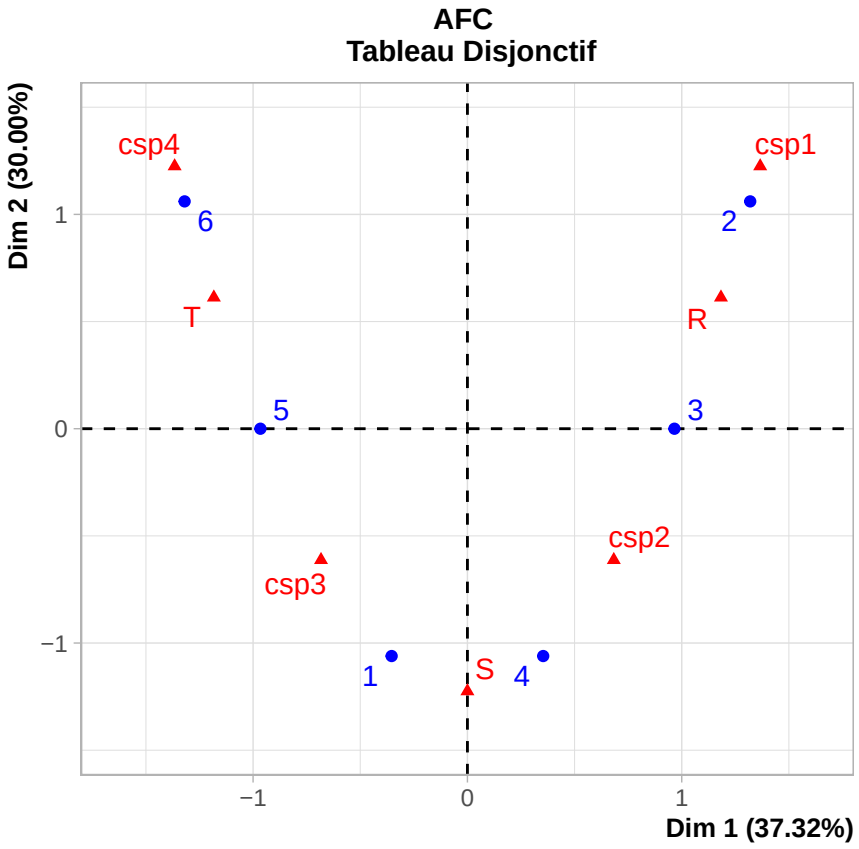
2.3 AFC du tableau disjonctif complet

```r
# Fonction construisant le tableau disjonctif complet #
disj.comp <- function(tab) {
n <- dim(tab)[1]
m <- dim(tab)[2]
ci <- sapply(tab,max)
c <- sum(ci)
disj.tab <- matrix(0,nrow=n,ncol=c)
id <- 0
for (j in 1:m) {
for (i in 1:n) {
disj.tab[i,id + tab[i,j]] <- 1
}
id <- id + ci[j]
}
disj.tab
}
# AFC du tableau disjonctif #
fic.disj.comp <- disj.comp(as.data.frame(cbind(csp,sport)))
colnames(fic.disj.comp) <- c(levels(csp),levels(sport))
```

```r
fic.disj.afc <- CA(fic.disj.comp,graph=FALSE)
fic.disj.afc
```

```
## **Results of the Correspondence Analysis (CA)**
## The row variable has  6  categories; the column variable has 7 categories
## The chi square of independence between the two variables is equal to 30 (p-value =  0.4656537 ).
## *The results are available in the following objects:
##
##    name              description
## 1  "$eig"            "eigenvalues"
## 2  "$col"            "results for the columns"
## 3  "$col$coord"      "coord. for the columns"
## 4  "$col$cos2"       "cos2 for the columns"
## 5  "$col$contrib"    "contributions of the columns"
## 6  "$row"            "results for the rows"
## 7  "$row$coord"      "coord. for the rows"
## 8  "$row$cos2"       "cos2 for the rows"
## 9  "$row$contrib"    "contributions of the rows"
## 10 "$call"           "summary called parameters"
## 11 "$call$marge.col" "weights of the columns"
## 12 "$call$marge.row" "weights of the rows"
```

```r
plot(fic.disj.afc,title="AFC \n Tableau Disjonctif")
```

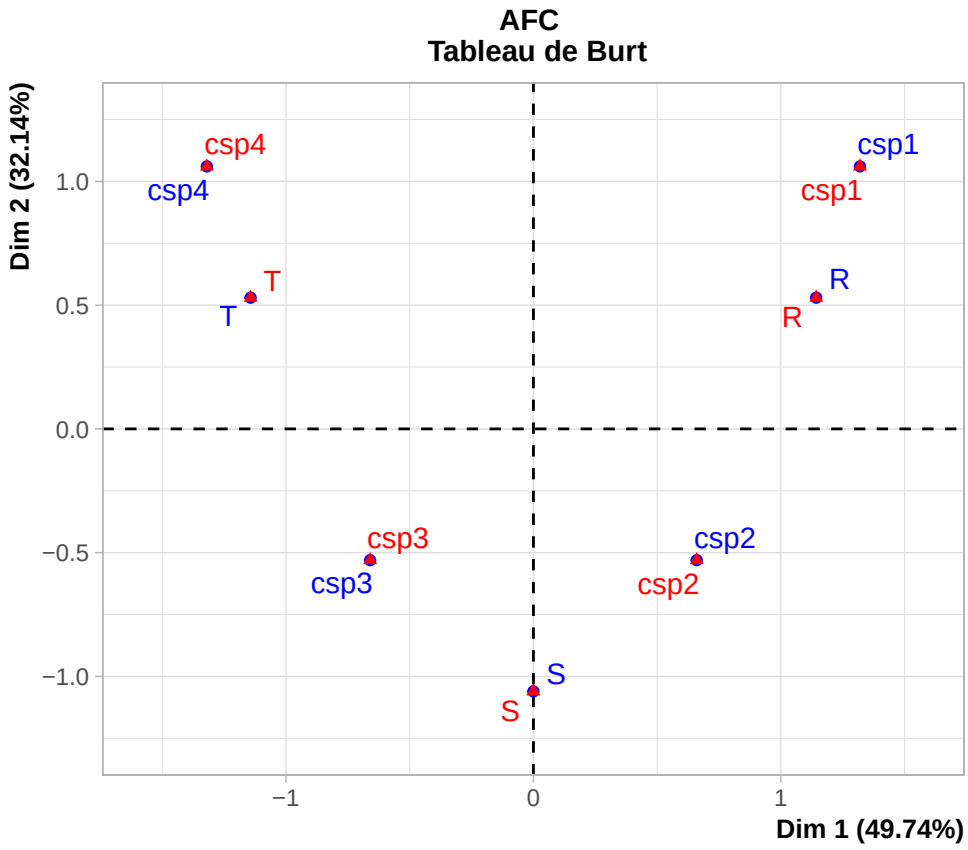


2.4 AFC du tableau de Burt

```
fic.burt <- t(fic.disj.comp) %*% fic.disj.comp
fic.burt.afc <- CA(fic.burt,graph=FALSE)
fic.burt.afc
```

```
## **Results of the Correspondence Analysis (CA)**
## The row variable has  7  categories; the column variable has 7 categories
## The chi square of independence between the two variables is equal to 42 (p-value =  0.2269628 ).
## *The results are available in the following objects:
##
##    name              description
## 1  "$eig"            "eigenvalues"
## 2  "$col"            "results for the columns"
## 3  "$col$coord"      "coord. for the columns"
## 4  "$col$cos2"       "cos2 for the columns"
## 5  "$col$contrib"    "contributions of the columns"
## 6  "$row"            "results for the rows"
## 7  "$row$coord"      "coord. for the rows"
## 8  "$row$cos2"       "cos2 for the rows"
## 9  "$row$contrib"    "contributions of the rows"
## 10 "$call"           "summary called parameters"
## 11 "$call$marge.col" "weights of the columns"
## 12 "$call$marge.row" "weights of the rows"
```

```
plot(fic.burt.afc,title="AFC \n Tableau de Burt")
```



2.4.1 Comparaison des résultats des différentes analyses factorielle des correspondances (AFC)

Nous allons comparer les résultats obtenus en termes de **valeurs propres** et de **graphique des représentations des individus et des modalités**.

2.4.2 AFC de la table de contingence

- **Chi carré** : 6 (p-value = 0.423)
- **Nombre de catégories** :
- 3 modalités pour la variable “sport” (S, R, T)
- 4 modalités pour la variable “csp” (csp1, csp2, csp3, csp4)
- **Valeurs propres (variance expliquée)** :
- Dim 1 : **75.00%**
- Dim 2 : **25.00%**
- **Interprétation graphique** :
- La première dimension explique la majeure partie de l’inertie.
- Répartition relativement équilibrée des modalités sur le graphique.

2.4.3 AFC du tableau disjonctif complet

- **Chi carré** : 30 (p-value = 0.466)
- **Nombre de catégories** :
- 6 modalités (issues des deux variables combinées)
- 7 colonnes (variables binaires issues du codage disjonctif)
- **Valeurs propres (variance expliquée)** :
- Dim 1 : **37.32%**
- Dim 2 : **30.00%**
- **Interprétation graphique** :
- La première dimension explique moins d'inertie que dans la table de contingence.
- Répartition plus dispersée des modalités, ce qui est attendu car le tableau disjonctif augmente le nombre de dimensions.

2.4.4 AFC du tableau de Burt

- **Chi carré** : 42 (p-value = 0.227)
- **Nombre de catégories** :
- 7 lignes et 7 colonnes
- **Valeurs propres (variance expliquée)** :
- Dim 1 : **49.74%**
- Dim 2 : **32.14%**
- **Interprétation graphique** :
- Inertie de la première dimension plus importante que pour le tableau disjonctif complet.
- Structure plus dense des modalités, car le tableau de Burt est une matrice de produits scalaires entre les modalités.

2.4.5 Comparaison et conclusions

Méthode	Dim 1 (%)	Dim 2 (%)	Chi²	p-value
Table de contingence	**75.00**	**25.00**	6	0.423
Tableau disjonctif	**37.32**	**30.00**	30	0.466
Tableau de Burt	**49.74**	**32.14**	42	0.227

Conclusion : - La table de contingence est idéale pour une interprétation simple des relations entre modalités. - Le tableau disjonctif complet offre plus de détails mais augmente la complexité de l'interprétation. - Le tableau de Burt donne une vue plus globale des relations entre toutes les modalités.

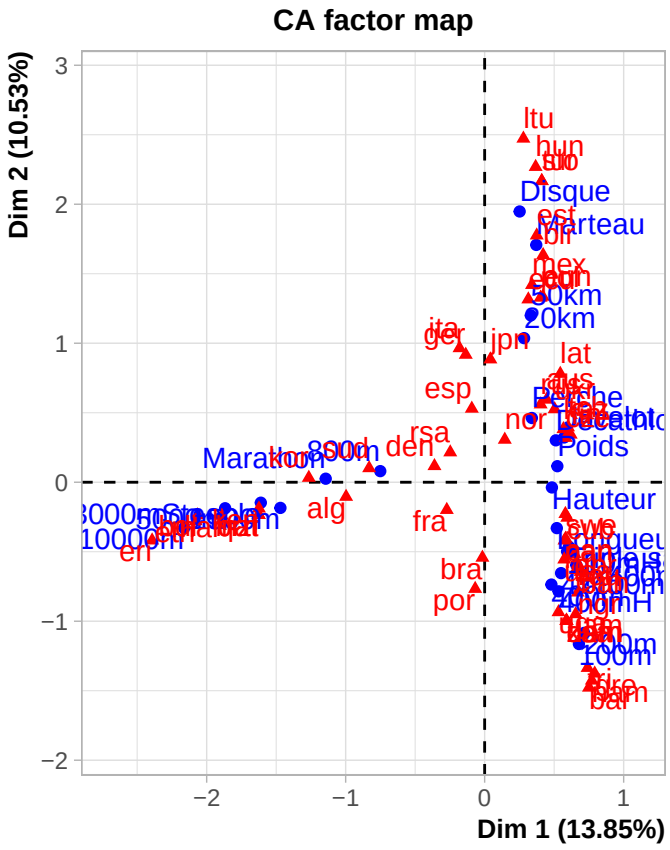
3. Données J.O

```
data(JO)
JO
```

```
##              alg aus bah bar bdi blr bra brn can chn cub cze den dom ecu eri
## 10000m         0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1
## 100m           0   0   0   1   0   0   0   0   1   0   0   0   0   0   0   0
## 110mH          0   0   0   0   0   0   0   0   1   1   3   0   0   0   0   0
## 1500m          1   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0
## 200m           0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 20km           0   2   0   0   0   0   0   0   1   0   0   0   0   0   2   0
## 3000mSteeple   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 400m           0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 400mH          0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0
## 4x100m         0   0   0   0   0   0   2   0   1   0   2   0   0   0   0   0
## 4x400m         0   1   1   0   0   0   0   0   0   0   1   0   0   0   0   0
## 5000m          1   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0
## 50km           0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 800m           1   0   0   0   0   0   0   0   0   0   0   0   2   0   0   0
## Decathlon      0   0   0   0   0   1   0   0   0   0   1   3   0   0   0   0
## Disque         0   0   0   0   0   2   0   0   0   0   1   0   0   0   0   0
## Hauteur        1   0   0   0   0   0   0   0   0   0   2   1   0   0   0   0
## Javelot        0   0   0   0   0   0   0   0   0   0   0   2   0   0   0   0
## Longueur       0   1   0   0   0   0   0   0   0   0   2   0   0   0   0   0
## Marathon       0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0
## Marteau        0   0   0   0   0   2   0   0   0   0   0   0   0   0   0   0
## Perche         0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## Poids          0   0   0   0   0   1   0   0   0   0   0   0   1   0   0   0
## Triple saut    0   0   2   0   0   0   0   0   0   0   2   0   0   0   0   0
##              esp est eth eun fin fra gbr ger gre hun ita jam jpn kaz ken kor
## 10000m         0   0   8   0   0   0   0   0   0   0   0   0   0   0   4   0
## 100m           0   0   0   0   0   0   1   0   0   0   0   1   0   0   0   0
## 110mH          0   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0
## 1500m          2   0   0   0   0   0   0   0   0   0   0   0   0   0   5   0
## 200m           0   0   0   0   0   0   1   0   1   0   0   1   0   0   0   0
## 20km           2   0   0   0   0   0   0   0   0   0   0   2   0   0   0   0
## 3000mSteeple   0   0   0   0   0   1   0   0   0   0   1   0   0   0  12   0
## 400m           0   0   0   0   0   0   1   0   0   0   0   1   0   0   1   0
## 400mH          0   0   0   0   0   1   1   0   0   0   0   2   0   0   0   0
## 4x100m         0   0   0   0   0   0   1   0   0   0   0   1   1   0   0   0
## 4x400m         0   0   0   0   0   0   2   0   0   0   0   2   0   0   0   0
## 5000m          0   0   4   0   0   0   0   1   0   0   0   0   0   0   5   0
## 50km           1   0   0   1   0   0   0   1   0   0   1   0   0   0   0   0
## 800m           0   0   0   0   0   0   0   1   0   0   0   0   0   0   5   0
## Decathlon      1   1   0   0   0   0   0   1   0   0   0   0   0   1   0   0
## Disque         0   2   0   0   0   0   0   3   0   1   0   0   0   0   0   0
## Hauteur        0   0   0   0   0   0   2   0   0   0   0   0   0   0   0   0
## Javelot        0   0   0   0   3   0   3   0   0   0   0   0   0   0   0   0
## Longueur       1   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0
## Marathon       0   0   3   0   0   0   0   1   0   0   1   0   1   0   3   2
## Marteau        0   0   0   3   0   0   0   0   0   2   1   0   2   0   0   0
## Perche         1   0   0   2   0   1   0   1   0   0   1   0   0   0   0   0
## Poids          0   0   0   1   1   0   0   0   0   0   0   0   0   0   0   0
## Triple saut    0   0   0   0   0   0   3   0   0   0   0   0   0   0   0   0
##              ksa lat ltu mar mex nam ngr nor nzl pan pol por qat rou rsa rus
## 10000m         0   0   0   2   0   0   0   0   0   0   0   0   0   0   0   0
## 100m           0   0   0   0   0   2   0   0   0   0   0   1   0   0   0   0
## 110mH          0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
## 1500m          0   0   0   3   0   0   0   0   1   0   0   1   1   0   0   0
## 200m           0   0   0   0   0   2   0   0   0   0   0   0   0   0   0   0
## 20km           0   0   0   0   2   0   0   0   0   0   1   0   0   0   0   3
## 3000mSteeple   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0
## 400m           0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 400mH          1   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0
## 4x100m         0   0   0   0   0   0   2   0   0   0   0   0   0   0   0   0
## 4x400m         0   0   0   0   0   0   2   0   0   0   0   0   0   0   0   1
## 5000m          0   0   0   3   0   0   0   0   0   0   0   0   0   0   0   0
## 50km           0   1   0   0   2   0   0   0   0   0   3   0   0   0   0   4
## 800m           0   0   0   0   0   0   0   1   0   0   0   0   0   0   2   1
## Decathlon      0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## Disque         0   0   4   0   0   0   0   0   0   0   1   0   0   0   1   0
## Hauteur        0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   3
## Javelot        0   2   0   0   0   0   0   2   0   0   0   0   0   0   0   2
## Longueur       0   0   0   0   0   0   0   0   0   1   0   0   0   0   1   0
## Marathon       0   0   0   1   0   0   0   0   0   0   0   0   0   0   1   0
## Marteau        0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0
## Perche         0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   3
## Poids          0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0
## Triple saut    0   0   0   0   0   0   0   0   0   0   0   1   0   1   0   2
##              slo sud swe tch tri tur uga ukr usa zam
## 10000m         0   0   0   0   0   0   0   0   0   0
## 100m           0   0   0   0   3   0   0   0   5   0
## 110mH          0   0   0   0   0   0   0   0   9   0
## 1500m          0   0   0   0   0   0   0   0   0   0
## 200m           0   0   0   0   2   0   0   0   8   0
## 20km           0   0   0   0   0   0   0   0   0   0
## 3000mSteeple   0   0   0   0   0   0   0   0   0   0
## 400m           0   0   0   0   0   0   1   0  11   0
## 400mH          0   0   0   0   0   0   0   0   7   1
## 4x100m         0   0   0   0   1   0   0   0   4   0
## 4x400m         0   0   0   0   0   0   0   0   5   0
## 5000m          0   0   0   0   0   0   0   0   0   0
## 50km           0   0   0   0   0   0   0   0   0   0
## 800m           0   1   0   0   0   0   0   0   1   0
## Decathlon      0   0   0   1   0   0   0   0   5   0
## Disque         0   0   0   0   0   0   0   0   0   0
## Hauteur        0   0   2   0   0   0   0   0   3   0
## Javelot        0   0   0   1   0   0   0   0   0   0
## Longueur       0   0   0   0   0   0   0   1   7   0
## Marathon       0   0   0   0   0   0   0   0   1   0
## Marteau        1   0   0   0   0   1   0   1   1   0
## Perche         0   0   0   0   0   0   0   1   4   0
## Poids          0   0   0   0   0   0   0   2   8   0
## Triple saut    0   0   1   0   0   0   0   0   3   0
```

```
JO.ca <- CA(JO)
```




```
JO.ca
```

```
## **Results of the Correspondence Analysis (CA)**
## The row variable has  24  categories; the column variable has 58 categories
## The chi square of independence between the two variables is equal to 2122.231 (p-value =  2.320981e-
## *The results are available in the following objects:
##
##    name              description
## 1  "$eig"            "eigenvalues"
## 2  "$col"            "results for the columns"
## 3  "$col$coord"      "coord. for the columns"
## 4  "$col$cos2"       "cos2 for the columns"
## 5  "$col$contrib"    "contributions of the columns"
## 6  "$row"            "results for the rows"
## 7  "$row$coord"      "coord. for the rows"
## 8  "$row$cos2"       "cos2 for the rows"
## 9  "$row$contrib"    "contributions of the rows"
## 10 "$call"           "summary called parameters"
## 11 "$call$marge.col" "weights of the columns"
## 12 "$call$marge.row" "weights of the rows"
```

```
round(JO.ca$eig,2)
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## dim 1        0.82                  13.85                             13.85
```

```
## dim 2        0.62                10.53                 24.38
## dim 3        0.54                 9.23                 33.62
## dim 4        0.48                 8.16                 41.78
## dim 5        0.40                 6.72                 48.50
## dim 6        0.36                 6.17                 54.67
## dim 7        0.33                 5.55                 60.23
## dim 8        0.32                 5.35                 65.58
## dim 9        0.27                 4.56                 70.14
## dim 10       0.24                 4.16                 74.29
## dim 11       0.23                 3.91                 78.20
## dim 12       0.18                 3.11                 81.31
## dim 13       0.16                 2.78                 84.09
## dim 14       0.14                 2.46                 86.55
## dim 15       0.13                 2.22                 88.77
## dim 16       0.12                 2.06                 90.82
## dim 17       0.10                 1.76                 92.58
## dim 18       0.09                 1.58                 94.16
## dim 19       0.08                 1.44                 95.60
## dim 20       0.08                 1.35                 96.95
## dim 21       0.07                 1.27                 98.21
## dim 22       0.06                 1.05                 99.27
## dim 23       0.04                 0.73                100.00
```

```r
barplot(JO.ca$eig[,1],main="Valeurs propres",
        names.arg=paste("dim",1:nrow(JO.ca$eig)))
```