

# Analyse en Composantes Principales (ACP)

Analyse R

08 mars 2025

## 2 Données fictives

### 2.1 Création des données

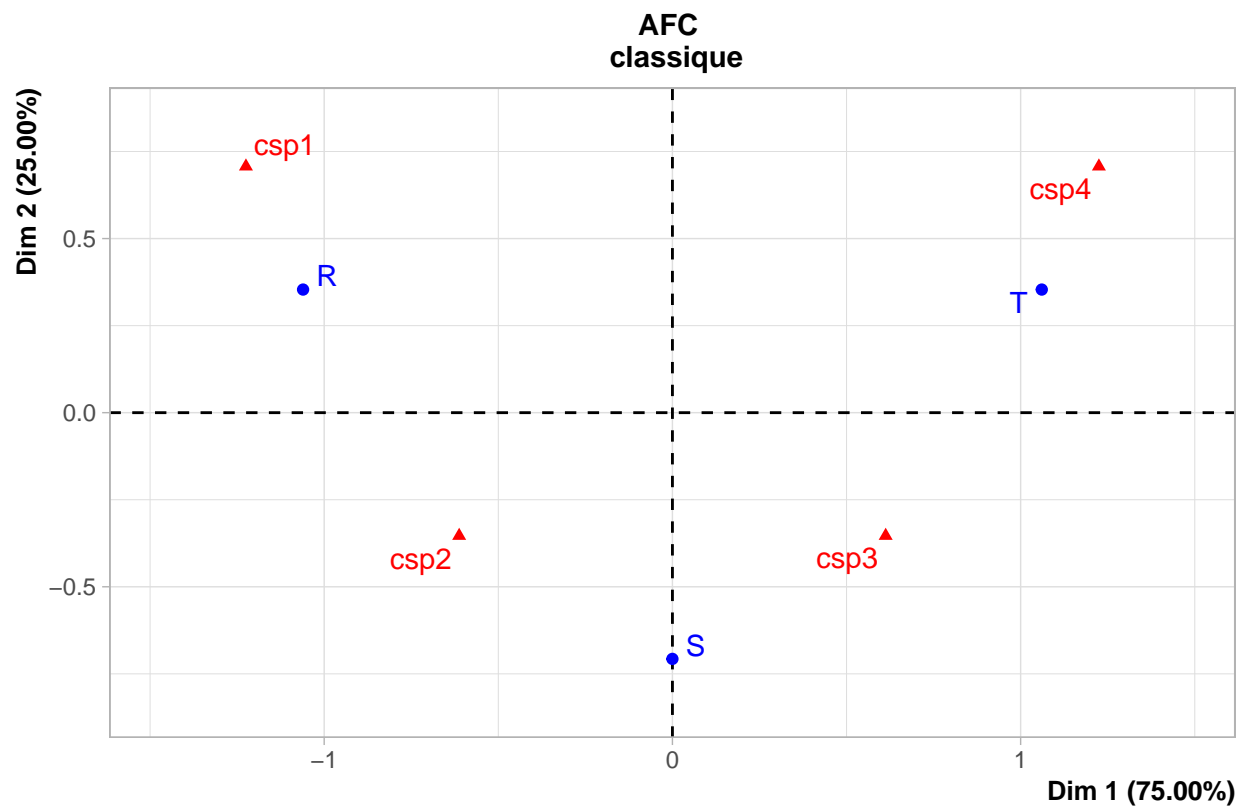
```
csp <- as.factor(c("csp3","csp1","csp2","csp2","csp3","csp4"))
sport <- as.factor(c("S","R","R","S","T","T"))
fic.tab <- table(sport,csp)
fic.conting <- data.frame(unclass(fic.tab))
# Paramètres graphiques #
x11()
par(mfrow=c(1,3))
```

### 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"
```

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



## 2.3 AFC du tableau disjonctif complet

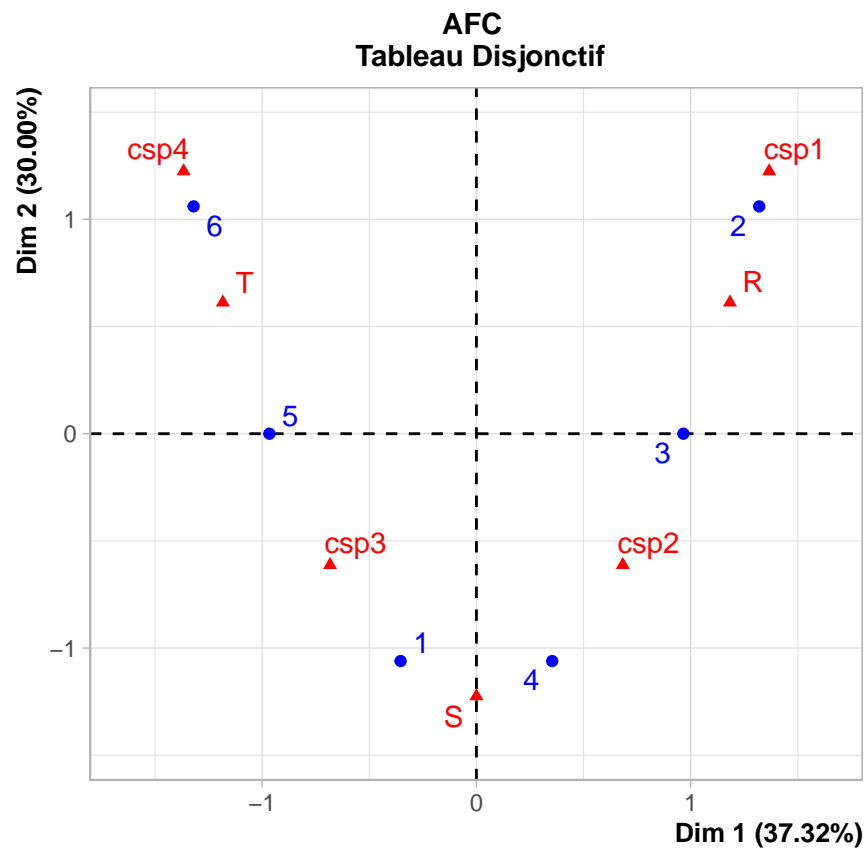
```
# Fonction construisant le tableau disjonctif complet #
disj.comp <- function(tab) {
  n <- dim(tab)[1]
  m <- dim(tab)[2]
  ci <- sapply(tab,max)
  c <- sum(ci)
  disj.tab <- matrix(0,nrow=n,ncol=c)
  id <- 0
  for (j in 1:m) {
    for (i in 1:n) {
      disj.tab[i,id + tab[i,j]] <- 1
    }
    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))
```

```
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"
```

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

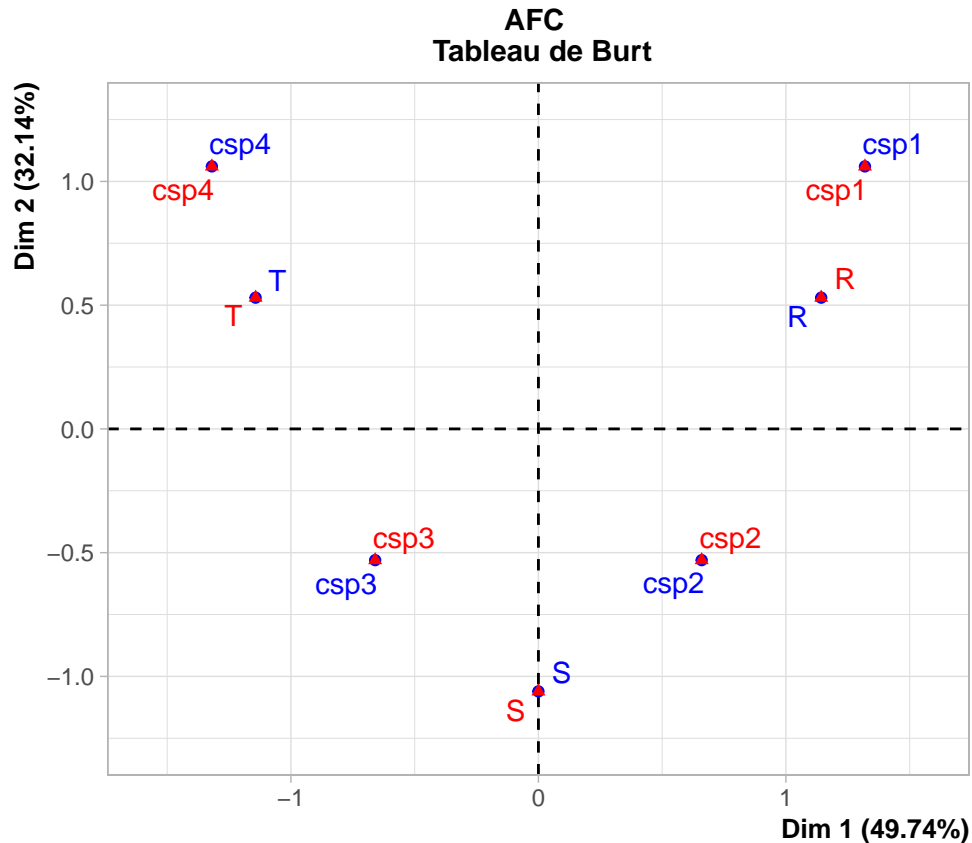


## 2.4 AFC du tableau de Burt

```
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 (%)	Chi <sup>2</sup>	p-value
<b>Table de contingence</b>	<b>75.00</b>	<b>25.00</b>	6	0.423
<b>Tableau disjonctif</b>	<b>37.32</b>	<b>30.00</b>	30	0.466
<b>Tableau de Burt</b>	<b>49.74</b>	<b>32.14</b>	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

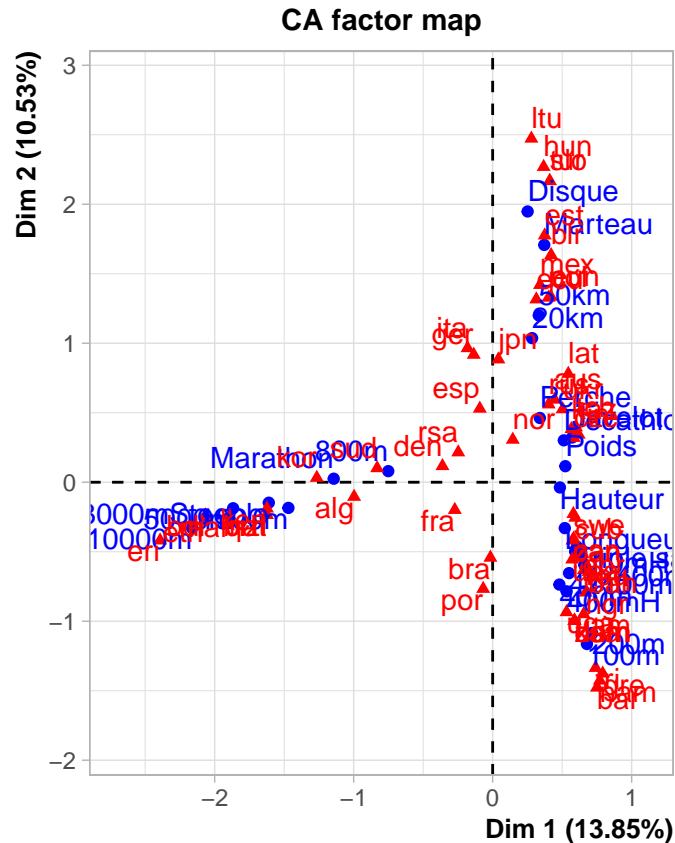
```
data(JO)
JO
```

##	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	2	0	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						

```
J0.ca <- CA(J0)
```





JO.ca

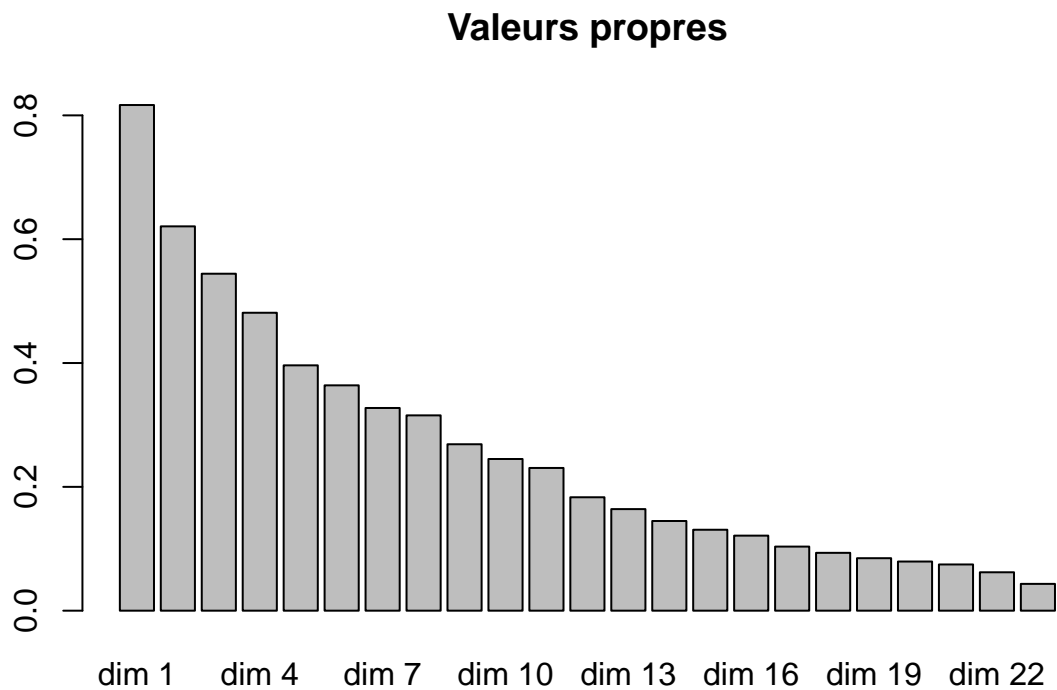
```
## **Results of the Correspondence Analysis (CA)**
## The row variable has 24 categories; the column variable has 58 categories
## The chi square of independence between the two variables is equal to 2122.231 (p-value = 2.320981e-4)
## *The results are available in the following objects:
##
##      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(J0.ca$eig,2)
```

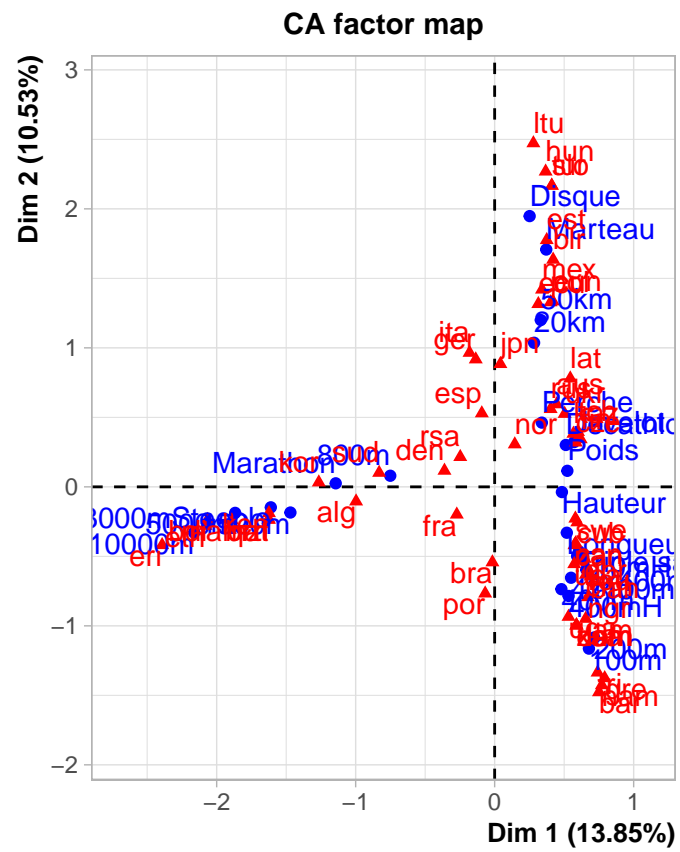
##	eigenvalue	percentage of variance	cumulative percentage of variance
## dim 1	0.82	13.85	13.85

## 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

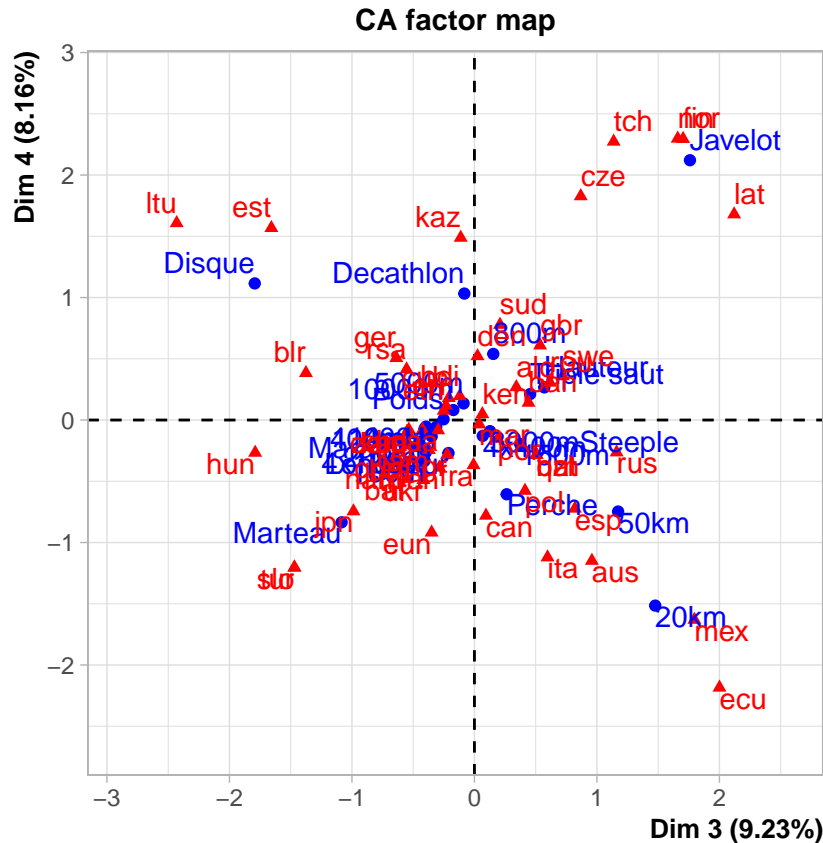
```
barplot(J0.ca$eig[,1],main="Valeurs propres",
        names.arg=paste("dim",1:nrow(J0.ca$eig)))
```



```
plot(J0.ca)
```



```
plot(J0.ca,axes=3:4)
```



### 3.1 Analyse des profils athlétiques des pays

- **Cluster des pays spécialisés en épreuves de fond et de marathon** : Des pays comme le Kenya (KEN) et l'Éthiopie (ETH) se situent à proximité des disciplines telles que le marathon et les courses de longue distance (10 000m, 5 000m). Ces nations sont reconnues pour leur domination dans les épreuves d'endurance.
- **Cluster des pays spécialisés en épreuves de sprint** : Des nations comme les États-Unis (USA) et la Jamaïque (JAM) sont fortement associées aux épreuves courtes comme le 100m et 200m, où elles excellent historiquement. Notamment avec des athlètes comme Usain Bolt.
- **Cluster des pays spécialisés en épreuves de lancers et de saut** : On remarque que certains pays d'Europe de l'Est et du Nord (comme la Pologne et la Hongrie) sont souvent liés aux disciplines de lancer (disque, marteau) et de saut.
- **Cluster des pays polyvalents** : Certains pays apparaissent plus au centre du graphique, ce qui suggère une distribution plus équilibrée de leurs performances sur plusieurs disciplines.

### 3.2 Analyse des profils géographiques des épreuves

Les disciplines sportives se regroupent en fonction des caractéristiques physiques et techniques requises :

- **Épreuves de fond et de demi-fond** : Ces disciplines (marathon, 10 000m, 5 000m) sont souvent liées aux pays africains qui ont historiquement dominé ces courses.
- **Épreuves de sprint et de vitesse** : Le 100m, 200m et 400m sont généralement dominés par des nations disposant d'un bon vivier d'athlètes explosifs, comme les États-Unis et la Jamaïque.

- **Épreuves de lancers** : Le poids, le disque et le javelot sont souvent associés aux pays d'Europe de l'Est et du Nord, où la formation et les infrastructures pour ces disciplines sont bien développées.
- **Épreuves combinées et techniques** : Le décathlon et l'heptathlon, ainsi que les sauts en hauteur et à la perche, montrent des influences variées, avec une répartition géographique plus homogène.
- 

### 3.3 Analyse des contributions des pays aux axes factoriels

```
# Contributions pays à l'axe 1 #
round(J0.ca$col$contrib[rev(order(J0.ca$col$contrib[,1])),1],2)
```

```
## ken eth mar usa gbr eri cub bdi alg jam tri kor rus
## 31.39 22.07 12.16 9.15 2.14 1.95 1.68 1.45 1.35 1.31 1.12 1.09 1.05
## qat nzl brn nam cze ngr fin bah can pol ukr eun aus
## 0.90 0.90 0.90 0.81 0.69 0.61 0.52 0.49 0.46 0.44 0.43 0.42 0.41
## blr swe lat tch sud gre bar mex rou chn est hun den
## 0.36 0.34 0.30 0.24 0.23 0.21 0.19 0.16 0.15 0.15 0.14 0.14 0.13
## rsa zam ksa dom pan kaz ltu uga ita fra ecu ger tur
## 0.13 0.12 0.12 0.12 0.11 0.11 0.11 0.10 0.08 0.08 0.07 0.06 0.06
## slo esp nor por jpn bra
## 0.06 0.02 0.02 0.00 0.00 0.00
```

```
# Contributions pays à l'axe 2 #
round(J0.ca$col$contrib[rev(order(J0.ca$col$contrib[,2])),2],2)
```

```
## usa ltu blr hun pol eun tri est ger nam jam mex ita
## 11.32 10.94 7.17 6.91 6.31 5.58 4.79 4.23 3.77 3.64 3.63 3.61 2.92
## rus tur slo gbr ecu jpn ngr esp bar aus gre lat por
## 2.67 2.10 2.10 1.80 1.55 1.40 1.13 1.00 0.98 0.95 0.85 0.82 0.79
## ukr ken bah eth zam ksa dom cub bra uga can cze mar
## 0.62 0.56 0.56 0.53 0.44 0.44 0.44 0.41 0.40 0.39 0.29 0.27 0.26
## fin chn rou tch nor rsa pan eri swe kaz fra bdi qat
## 0.21 0.18 0.16 0.14 0.13 0.12 0.08 0.08 0.07 0.07 0.05 0.03 0.02
## nzl brn alg den sud kor
## 0.02 0.02 0.02 0.02 0.00 0.00
```

- **Axe 1** : Il est fortement influencé par des pays spécialisés dans les épreuves d'endurance comme le Kenya et l'Éthiopie, ainsi que par des nations performantes dans des disciplines variées comme les États-Unis et la Grande-Bretagne.
- **Axe 2** : Il met en évidence des pays plus orientés vers les épreuves techniques et combinées, avec des nations comme la Lituanie et la Biélorussie, qui se démarquent par leur contribution importante.