

# Examen sur l'analyse multidimensionnelle

Zaineb Smida

2024-12-16

**Consignes :** dans la première partie, une analyse multidimensionnelle a été réalisée sur un jeu de donnée. Seuls les tableaux de résultats et graphiques sont inclus dans l'énoncé. Vous devrez répondre à 8 questions qui doivent permettre d'interpréter les sorties données par **R**. Chaque question vaut 2.5 points. Pour chaque réponse, vous devrez indiquer le tableau (ou graphique) que vous avez utilisé pour répondre à la question. Pensez à bien argumenter chacune de vos réponses. **Aucun document n'est autorisé.**

**Durée :** 1 heure

## Description des données :

Le fichier étudié concerne les performances des athlètes lors de deux événements sportifs. Plus précisément, le tableau de données comprend 41 lignes et 13 colonnes : les dix premières colonnes correspondent aux performances des athlètes pour les 10 épreuves du décathlon :

- 100m : le 100 mètres (en seconde),
- Long.jump : le saut en longueur (en mètre),
- Shot.put : le lancer du poids (en mètre),
- High.jump : le saut en hauteur (en mètre),
- 400m : le 400 mètres (en seconde),
- 110m.hurdle : le 110 mètres haies (en seconde),
- Discus : le lancer du disque (en mètre),
- Pole.vault : le saut à la perche (en mètre),
- Javeline : le lancer du javelot (en mètre)
- 1500m : le 1500 mètres (en seconde)

Les colonnes 11 (variable **Rank**) et 12 (variable **Points**) correspondent respectivement au classement et aux points obtenus. La dernière colonne (variable **Competition**) est une variable catégorielle correspondant à l'événement sportif (Jeux olympiques de 2004 ou Décastar 2004).

On affiche les premières lignes du jeu de données :

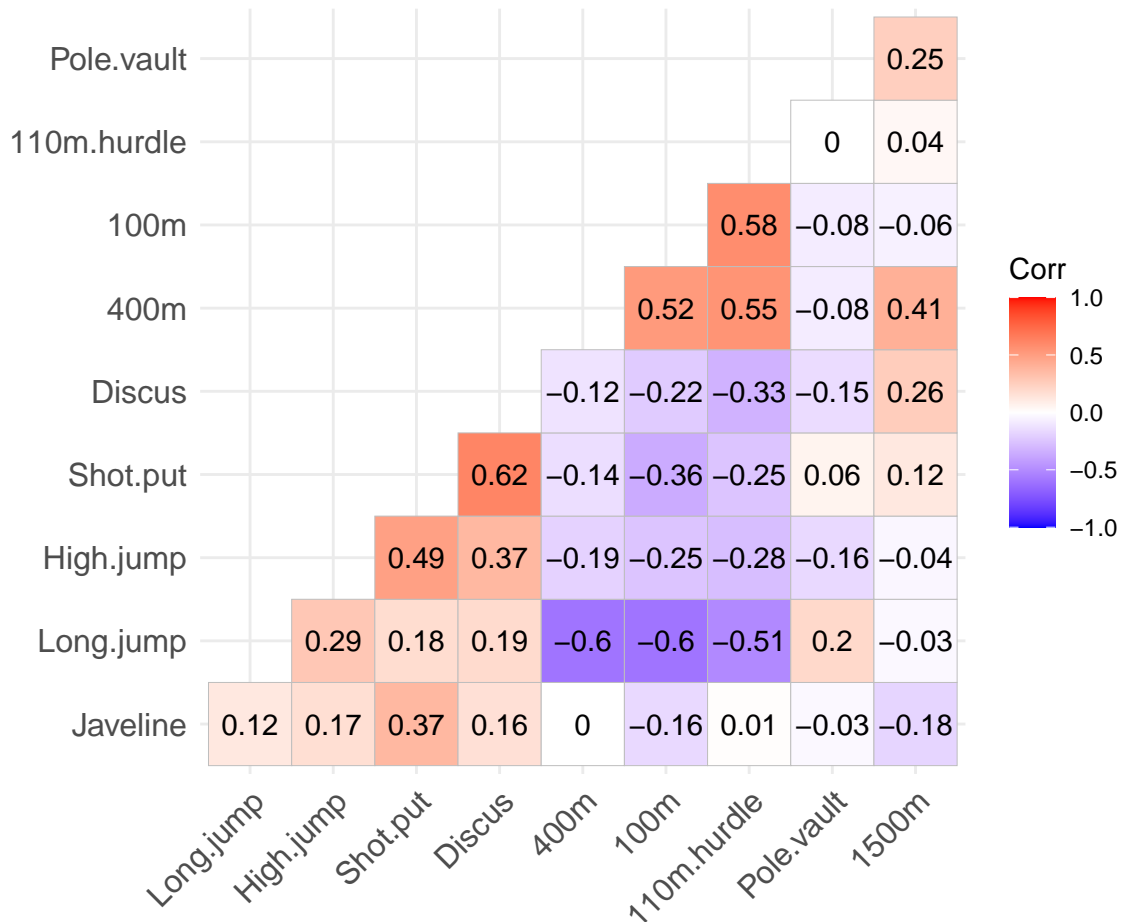
	100m	Long.jump	Shot.put	High.jump	400m	110m.hurdle	Discus	Pole.vault
SEBRLE	11.04	7.58	14.83	2.07	49.81	14.69	43.75	5.02
CLAY	10.76	7.40	14.26	1.86	49.37	14.05	50.72	4.92
KARPOV	11.02	7.30	14.77	2.04	48.37	14.09	48.95	4.92
BERNARD	11.02	7.23	14.25	1.92	48.93	14.99	40.87	5.32
YURKOV	11.34	7.09	15.19	2.10	50.42	15.31	46.26	4.72

	Javeline	1500m	Rank	Points	Competition
SEBRLE	63.19	291.7	1	8217	Decastar
CLAY	60.15	301.5	2	8122	Decastar
KARPOV	50.31	300.2	3	8099	Decastar
BERNARD	62.77	280.1	4	8067	Decastar
YURKOV	63.44	276.4	5	8036	Decastar

On traite ces données par une analyse en composantes principales normée. Les principaux résultats de cette ACP sont indiqués ci-dessous :

### Graphique des corrélations



## Valeurs Propres & stat. associées

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3.2719055	32.719055	32.71906
comp 2	1.7371310	17.371310	50.09037
comp 3	1.4049167	14.049167	64.13953
comp 4	1.0568504	10.568504	74.70804
comp 5	0.6847735	6.847735	81.55577
comp 6	0.5992687	5.992687	87.54846
comp 7	0.4512353	4.512353	92.06081
comp 8	0.3968766	3.968766	96.02958
comp 9	0.2148149	2.148149	98.17773
comp 10	0.1822275	1.822275	100.00000

## Les coordonnées des variables sur les axes

	Dim.1	Dim.2	Dim.3	Dim.4
100m	-0.775	0.187	-0.184	-0.038
Long.jump	0.742	-0.345	0.182	0.102
Shot.put	0.623	0.598	-0.023	0.191
High.jump	0.572	0.35	-0.26	-0.136
400m	-0.68	0.569	0.131	0.029
110m.hurdle	-0.746	0.229	-0.093	0.291
Discus	0.552	0.606	0.043	-0.26
Pole.vault	0.05	-0.18	0.692	0.552
Javeline	0.277	0.317	-0.39	0.712
1500m	-0.058	0.474	0.782	-0.161

## Les contributions des variables sur les axes

	Dim.1	Dim.2	Dim.3	Dim.4
100m	18.344	2.016	2.42	0.135
Long.jump	16.822	6.869	2.363	0.98
Shot.put	11.844	20.607	0.039	3.437
High.jump	9.998	7.064	4.794	1.74
400m	14.116	18.666	1.23	0.081
110m.hurdle	17.02	3.013	0.611	8.003
Discus	9.328	21.162	0.131	6.38
Pole.vault	0.077	1.873	34.061	28.783
Javeline	2.347	5.784	10.807	48.005
1500m	0.103	12.946	43.543	2.455

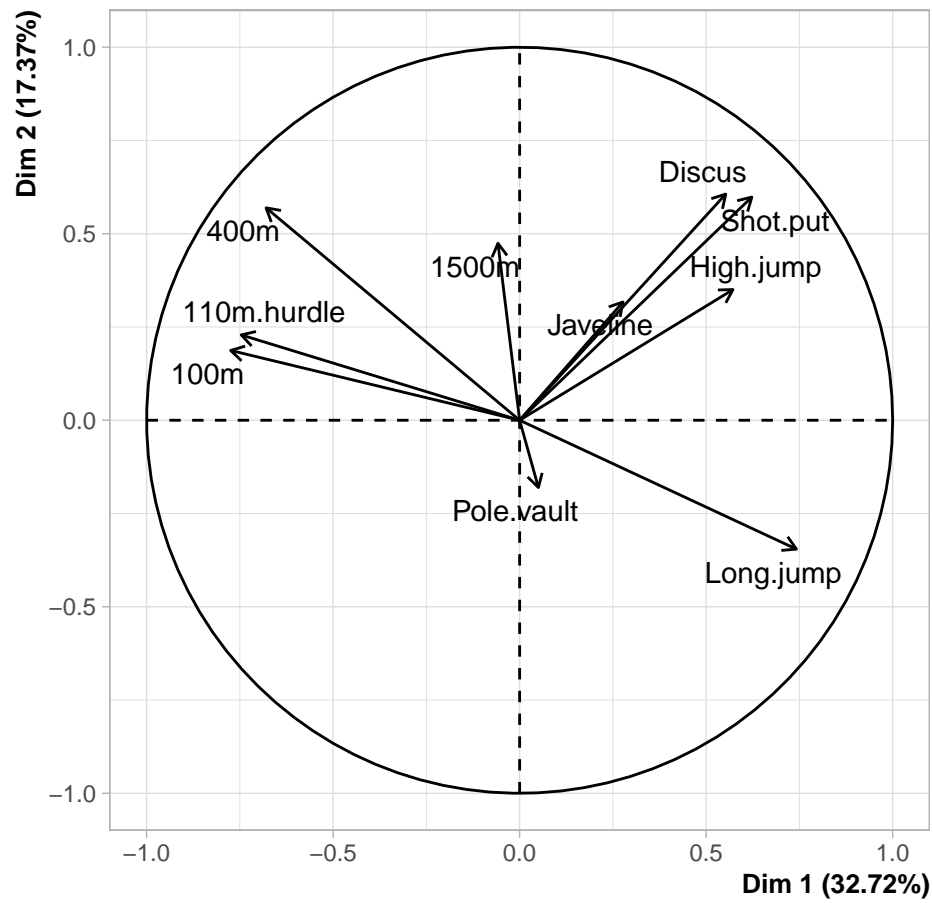
## Les coordonnées des individus sur les axes

	Dim.1	Dim.2	Dim.3	Dim.4
SEBRLE	0.792	0.772	0.827	1.175
CLAY	1.235	0.575	2.141	-0.355
KARPOV	1.358	0.484	1.956	-1.857
BERNARD	-0.61	-0.875	0.89	2.221
YURKOV	-0.586	2.131	-1.225	0.874
WARNERS	0.357	-1.685	0.767	-0.589
ZSIVOCZKY	0.272	-1.094	-1.283	-1.622
McMULLEN	0.588	0.231	-0.418	-1.524
MARTINEAU	-1.995	0.561	-0.73	-0.542
HERNU	-1.546	0.488	0.841	0.331
BARRAS	-1.342	-0.311	0	-0.645
NOOL	-2.345	-1.966	-1.336	0.195
BOURGUIGNON	-3.979	0.2	1.326	0.524
Sebrle	4.038	1.366	-0.29	1.941
Clay	3.919	0.837	0.231	1.494
Karpov	4.62	0.04	-0.042	-1.314
Macey	2.233	1.042	-1.864	-0.743
Warners	2.168	-1.803	0.851	-0.285
Zsivoczky	0.925	1.169	-1.477	0.808
Hernu	0.889	-0.618	-0.898	-0.135
Nool	0.295	-1.546	1.355	2.2
Bernard	1.906	-0.086	-0.757	-1.451
Schwarzl	0.081	-1.353	0.822	0.399
Pogorelov	0.54	0.771	1.348	-0.552
Schoenbeck	0.114	-0.04	0.74	0.929
Barras	0.002	0.36	-1.57	0.612
Smith	0.87	1.059	-1.643	-1.121
Averyanov	0.349	-1.559	0.283	-0.027
Ojaniemi	0.38	-0.772	-0.371	0.687
Smirnov	-0.485	-1.061	-1.228	0.566
Qi	-0.434	-0.326	-1.07	-0.205
Drews	-0.249	-3.082	1.055	-0.646
Parkhomenko	-1.069	2.093	-1	1.535
Terek	-0.682	0.536	2.209	0.109
Gomez	-0.29	-1.197	-1.306	0.078
Turi	-1.542	0.427	0.514	-0.143
Lorenzo	-2.409	-1.583	-1.502	0.301
Karlivans	-1.994	-0.294	-0.343	-1.272
Korkizoglou	-0.958	2.066	2.587	-1.191
Uldal	-2.562	0.245	-0.419	-0.021
Casarsa	-2.857	3.798	0.031	-0.738

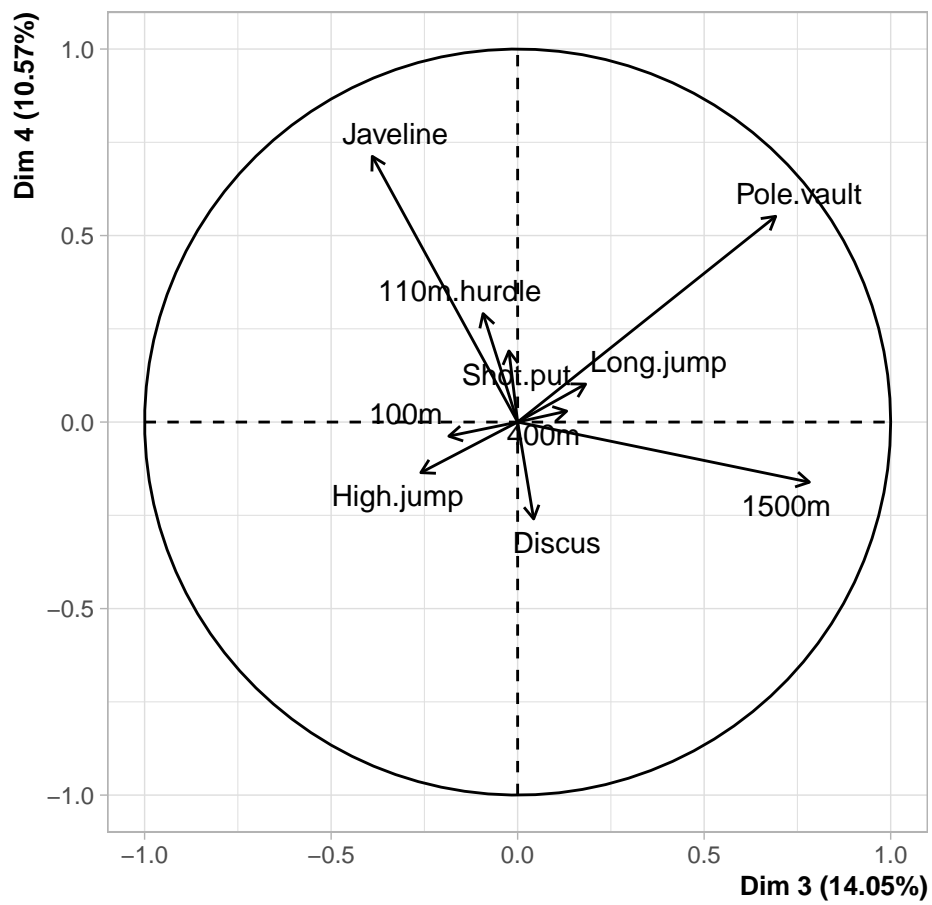
## Les cosinus carrés des individus sur les axes

	Dim.1	Dim.2	Dim.3	Dim.4
SEBRLE	0.112	0.106	0.122	0.246
CLAY	0.124	0.027	0.373	0.01
KARPOV	0.16	0.02	0.332	0.299
BERNARD	0.049	0.1	0.104	0.646
YURKOV	0.038	0.499	0.165	0.084
WARNERS	0.022	0.482	0.1	0.059
ZSIVOCZKY	0.011	0.182	0.25	0.4
McMULLEN	0.053	0.008	0.027	0.354
MARTINEAU	0.284	0.022	0.038	0.021
HERNU	0.306	0.031	0.091	0.014
BARRAS	0.472	0.025	0	0.109
NOOL	0.394	0.277	0.128	0.003
BOURGUIGNON	0.857	0.002	0.095	0.015
Sebrle	0.695	0.08	0.004	0.161
Clay	0.711	0.032	0.002	0.103
Karpov	0.852	0	0	0.069
Macey	0.423	0.092	0.295	0.047
Warners	0.53	0.366	0.082	0.009
Zsivoczky	0.13	0.207	0.332	0.099
Hernu	0.238	0.115	0.242	0.005
Nool	0.009	0.249	0.191	0.504
Bernard	0.455	0.001	0.072	0.263
Schwarzl	0.002	0.472	0.174	0.041
Pogorelov	0.051	0.105	0.32	0.054
Schoenbeck	0.004	0	0.17	0.267
Barras	0	0.026	0.498	0.076
Smith	0.061	0.09	0.216	0.101
Averyanov	0.019	0.382	0.013	0
Ojaniemi	0.026	0.109	0.025	0.086
Smirnov	0.057	0.275	0.369	0.078
Qi	0.061	0.034	0.368	0.013
Drews	0.005	0.811	0.095	0.036
Parkhomenko	0.094	0.361	0.082	0.194
Terek	0.043	0.027	0.453	0.001
Gomez	0.012	0.21	0.25	0.001
Turi	0.252	0.019	0.028	0.002
Lorenzo	0.471	0.203	0.183	0.007
Karlivans	0.544	0.012	0.016	0.221
Korkizoglou	0.058	0.27	0.423	0.09
Uldal	0.757	0.007	0.02	0
Casarsa	0.337	0.596	0	0.022

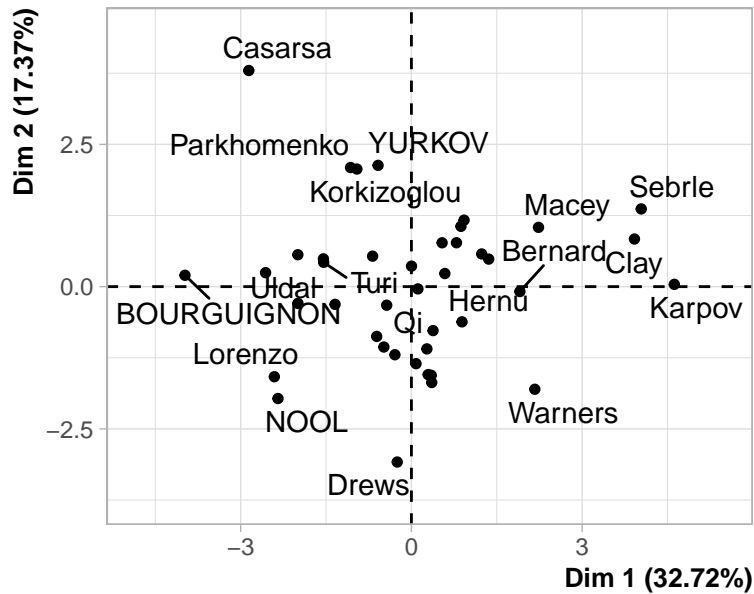
Graphique des variables (dim 1, 2) :



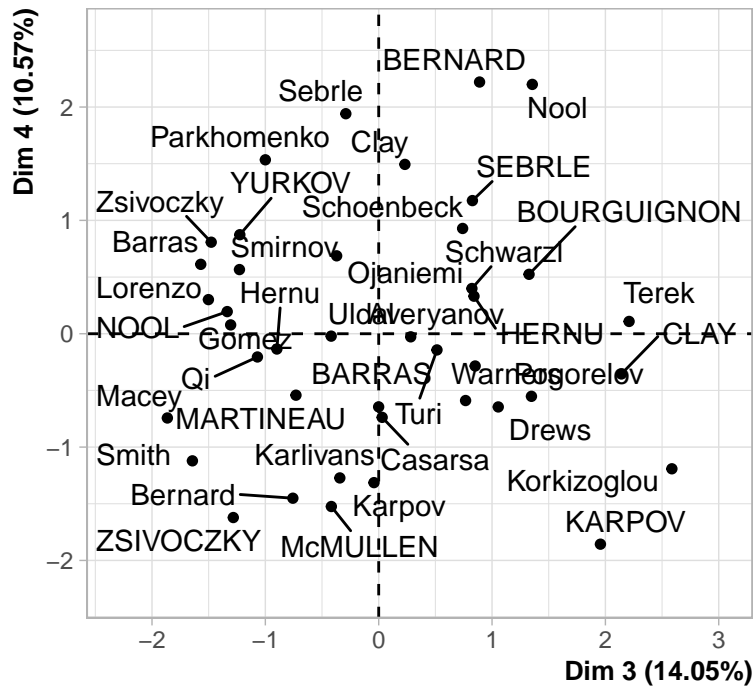
Graphique des variables (dim 3, 4) :



Graphique des individus (dim 1, 2) :



Graphique des individus (dim 3, 4) :





## 1 Question (2.5pts)

Que représente le graphique des corrélations ? Décrire les corrélations les plus fortes (positives et négatives). Pourquoi l'ACP est-elle justifiée ?

## 2 Question (2.5pts)

A quoi correspond la somme des valeurs propres ? On choisit de n'étudier que les 4 premières composantes principales. Justifier ce choix en analysant le tableau des valeurs propres

## 3 Question (2.5pts)

Quels sont les individus les mieux représentés dans chacun des axes retenus. Précisez le signe des coordonnées correspondantes.

## 4 Question (2.5pts)

Pour chaque axe retenu, quelles sont les variables les mieux représentées. Précisez le signe des coordonnées correspondantes. Pouvez-vous confirmer cela à partir du graphique des variables ?

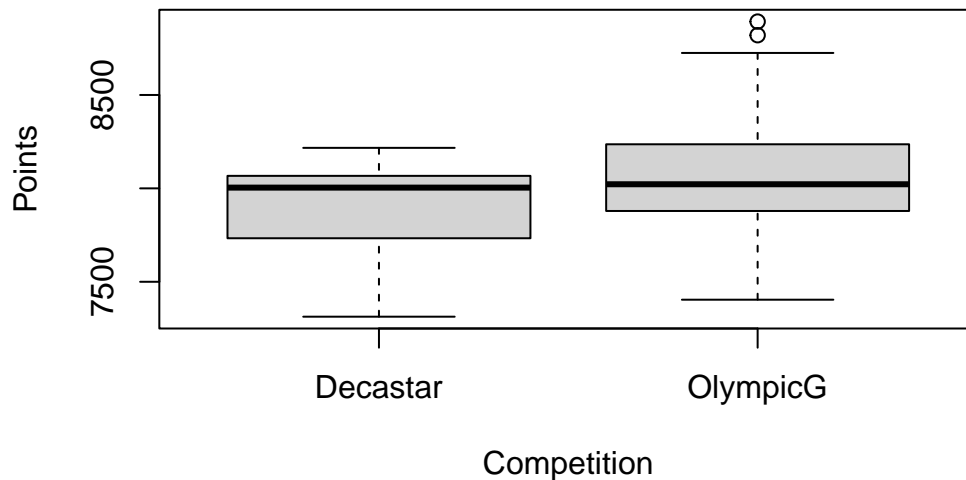
## 5 Question (2.5pts)

Faire la synthèse de l'ACP.

## 6 Question (2.5pts)

On a représenté les scores finaux (variable quantitative) obtenus par les athlètes en fonction de la compétition (variable qualitative), sous forme de boîtes à moustaches parallèles. D'après le graphique, est-ce que les athlètes performant mieux dans une compétition plutôt qu'une autre ? Confirmez (ou non) votre réponse en utilisant la sortie **R** suivante.

```
boxplot(Points ~ Competition, data = decathlon)
```



```
res_lm <- lm(Points ~ Competition, data = decathlon)
anova(res_lm)
```

Analysis of Variance Table

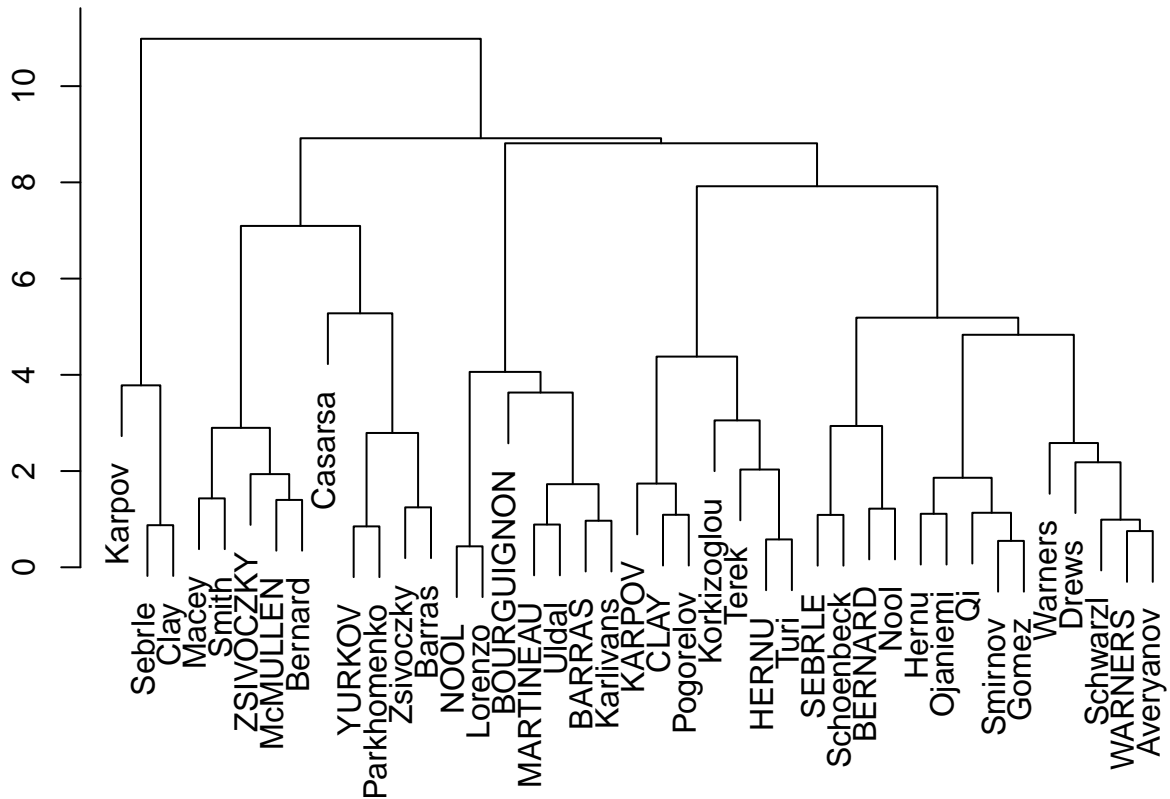
Response: Points

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Competition	1	188242	188242	1.6311	0.2091
Residuals	39	4500862	115407		

## 7 Question (2.5pts)

On a réalisé une classification sur les 4 premières composantes principales de l'ACP. Comme s'appelle le graphique ci-dessous ? Quel est le nom de la méthode de classification utilisée ? L'examen de ce graphique suggère de réaliser une partition des athlètes, soit en 2 classes, soit en 6 classes. Justifier ?

```
par(mar = c(3, 3, 0.5, 0.5))
my_dist <- dist(my_pca$ind$coord[, 1:4])
resuclassif <- hclust(my_dist, method = "ward.D2")
plot(resuclassif, main = "")
```



## 8 Question (2.5pts)

On décide de garder 6 classes qu'on représente dans les graphiques de l'ACP suivants. Comment peut-on décrire les positions des 6 classes précédentes sur le graphique de l'ACP ?

```
decathlon$my_cluster <- factor(cutree(resuclassif, k = 6))
```

