

Analyse en composantes principales des données décathlon

Quelques sorties R

Nathalie AKAKPO

Jeudi 7 décembre 2017

Table des matières

1	Présentation du document	1
1.1	Présentation des données	1
1.2	Réalisation de l'ACP	2
2	Choix de la dimension	2
3	Projections et contributions des individus	3
4	Projections et contributions des variables	5
5	Ajout d'une variable quantitative	7
6	Représentation des individus et des variables dans le premier plan factoriel	7
7	Représentation des individus et des variables dans le plan factoriel 3-4	8

ATTENTION : POUR ME COMPILER, VOUS DEVEZ INSTALLER LES PACKAGE book-down et kableExtra DANS R, ET LES PACKAGES float ET booktabs DANS LATEX.

1 Présentation du document

Ce document a pour objectif de présenter les données décathlon et les principales sorties d'une analyse en composantes principales.

Le code source R du paragraphe suivant inclut :

1. du code évalué dans R, mais qui n'apparaît pas (ni le code, ni les sorties) dans le pdf (`include=FALSE`) ;
2. du code évalué dans R, pour lequel on voit juste les sorties dans le pdf (`echo=FALSE`) ;
3. des valeurs numériques calculées sous R et intégrées dans le texte (voir lignes 43-44 du code source).

1.1 Présentation des données

Le jeu de données `decathlon` contient 41 individus (décathloniens) sur lesquels on mesure 13 variables. Les 10 premières correspondent aux résultats de chacune des 10 épreuves du décathlon :

- temps au 100m (“100m”) ;
- distance au saut en longueur (“Long.jump”) ;
- distance au lancer de poids (“Shot.put”) ;
- hauteur au saut en hauteur (“High.jump”) ;
- temps au 400m (“400m”) ;

- temps au 110m haies (“110m.hurdle”);
- distance au lancer de disque (“Discus”);
- hauteur au saut à la perche (“Pole.vault”);
- distance au lancer de javelot (“Javeline”);
- temps au 1500m (“1500m”).

Les 3 autres variables sont :

- le classement (“Rank”);
- le nombre total de points obtenus au décathlon (“Points”);
- la compétition (“Competition”) : Jeux Olympiques ou Décastar.

Le nom d’un même athlète peut apparaître 2 fois s’il a participé au Décastar et aux Jeux Olympiques.

On constate que les variables ne sont pas mesurées dans la même échelle :

```
##      High.jump      1500m
##  Min.      :1.850    Min.      :262.1
## 1st Qu.:1.920    1st Qu.:271.0
## Median :1.950    Median :278.1
## Mean   :1.977    Mean   :279.0
## 3rd Qu.:2.040    3rd Qu.:285.1
## Max.    :2.150    Max.    :317.0
```

1.2 Réalisation de l’ACP

Le code source R du paragraphe suivant inclut :

1. des références à d’autres paragraphes du document.

Pour réaliser l’analyse en composantes principales de ce jeux de données, les variables Classement et Points ne seront pas prises en compte (déclarées quantitatives supplémentaires), la variable Competition non plus (déclarée qualitative supplémentaire).

```
res.pca<-PCA(decathlon,quanti.sup =11:12,quali.sup=13,graph=F)
```

L’option graph=TRUE (par défaut) construit automatiquement les graphes des individus et des variables, mais il faut en général les refaire en modifiant les options pour qu’ils soient plus lisibles (voir les paragraphes 6 et 7).

2 Choix de la dimension

Le code source R de ce paragraphe suivant inclut :

1. du code évalué dans R, pour lequel on voit juste les sorties dans le pdf ([echo=FALSE](#)) ;
2. un tableau de sorties R amélioré avec un titre, auquel on peut faire référence dans le texte ;
3. une figure R avec un titre, à laquelle on peut faire référence dans le texte.

Pour choisir le nombre de composantes principales, on peut s’aider de la figure 1 et du tableau 1.

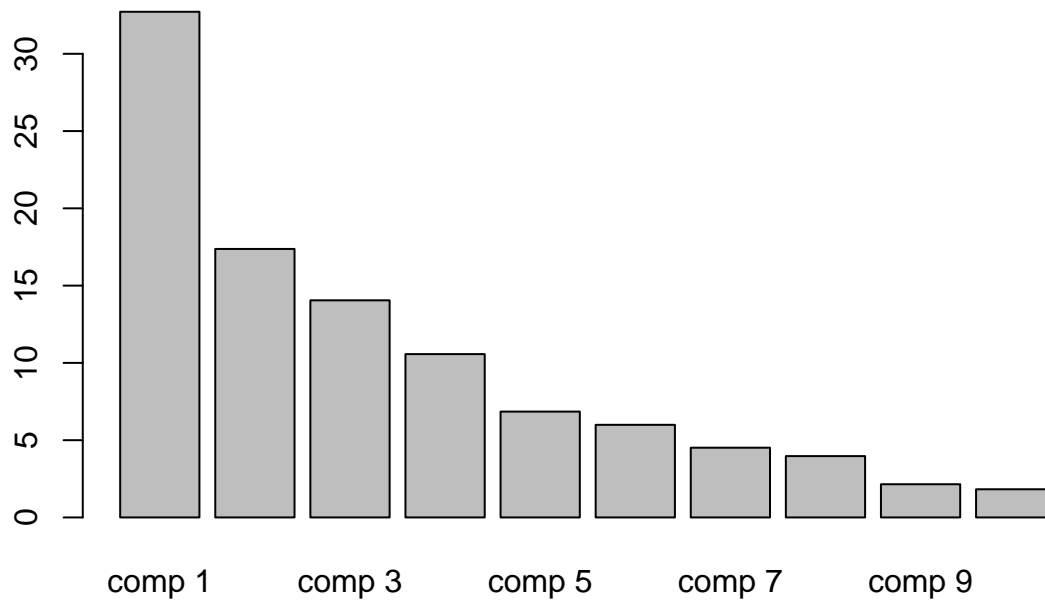


FIGURE 1 – Ma première figure

TABLE 1 – Informations sur les valeurs propres de S

	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

3 Projections et contributions des individus

```
names(res.pca$ind)
```

```
## [1] "coord" "cos2" "contrib" "dist"
```

```
head(res.pca$ind$coord)
```

```
##          Dim.1    Dim.2    Dim.3    Dim.4    Dim.5
## SEBRLE  0.7916277  0.7716112  0.8268412  1.1746274  0.7071590
## CLAY    1.2349906  0.5745781  2.1412470 -0.3548448 -1.9745714
## KARPOV  1.3582149  0.4840209  1.9562580 -1.8565241  0.7952147
## BERNARD -0.6095151 -0.8746285  0.8899407  2.2206124  0.3616362
## YURKOV  -0.5859683  2.1309542 -1.2251568  0.8735791  1.2513692
```

```
## WARNERS 0.3568895 -1.6849567 0.7665531 -0.5893047 1.0016616
```

```
head(res.pca$ind$cos2)
```

```
##          Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## SEBRLE  0.11167888 0.10610262 0.12183534 0.24588345 0.08911755
## CLAY    0.12400941 0.02684265 0.37278712 0.01023775 0.31701007
## KARPOV  0.15991886 0.02030911 0.33175306 0.29878849 0.05481905
## BERNARD 0.04867778 0.10023262 0.10377289 0.64611132 0.01713585
## YURKOV  0.03769960 0.49858212 0.16480554 0.08379015 0.17193305
## WARNERS 0.02160805 0.48164324 0.09968563 0.05891525 0.17021193
```

```
res.pca$ind$contrib
```

```
##          Dim.1      Dim.2      Dim.3      Dim.4
## SEBRLE  4.671511e-01 0.835950588 1.186888e+00 3.184218572
## CLAY    1.136953e+00 0.463534061 7.959744e+00 0.290589297
## KARPOV  1.375157e+00 0.328936297 6.643820e+00 7.954334225
## BERNARD 2.769391e-01 1.074065694 1.374952e+00 11.380155224
## YURKOV  2.559550e-01 6.375757749 2.605847e+00 1.761193871
## WARNERS 9.494738e-02 3.986217937 1.020117e+00 0.801461010
## ZSIVOCZKY 5.505968e-02 1.679732058 2.856679e+00 6.068357238
## McMULLEN 2.573091e-01 0.074746587 3.027995e-01 5.361733464
## MARTINEAU 2.967957e+00 0.441878700 9.250128e-01 0.678434142
## HERNU 1.781875e+00 0.334891915 1.227259e+00 0.253145030
## BARRAS 1.341824e+00 0.135724214 2.355361e-07 0.960866814
## NOOL 4.099127e+00 5.428955168 3.100927e+00 0.088028018
## BOURGUIGNON 1.180246e+01 0.056083636 3.054712e+00 0.634521052
## Sebrle 1.215751e+01 2.619234357 1.459591e-01 8.695883785
## Clay 1.145109e+01 0.983545343 9.277875e-02 5.150953555
## Karpov 1.591098e+01 0.002245949 3.002311e-03 3.981803022
## Macey 3.718536e+00 1.523786399 6.034288e+00 1.274764199
## Warners 3.505038e+00 4.565322740 1.257310e+00 0.186926623
## Zsivoczky 6.380034e-01 1.917581489 3.789736e+00 1.505183982
## Hernu 5.891907e-01 0.536979795 1.400889e+00 0.041926274
## Nool 6.500682e-02 3.354186097 3.188680e+00 11.167049428
## Bernard 2.709028e+00 0.010337168 9.953382e-01 4.858755070
## Schwarzl 4.900366e-03 2.572009426 1.174420e+00 0.367576264
## Pogorelov 2.171117e-01 0.834087762 3.152828e+00 0.703603009
## Schoenbeck 9.761188e-03 0.002229736 9.517066e-01 1.991093229
## Barras 3.430458e-06 0.182306828 4.277552e+00 0.865463896
## Smith 5.646299e-01 1.575587174 4.688863e+00 2.901795823
## Averyanov 9.087664e-02 3.410992543 1.385834e-01 0.001715382
## Ojaniemi 1.077068e-01 0.837763306 2.388803e-01 1.090106764
## Smirnov 1.749961e-01 1.579562867 2.619396e+00 0.739408634
## Qi 1.407111e-01 0.149351665 1.986868e+00 0.096961729
## Drews 4.610109e-02 13.333917174 1.931706e+00 0.962421426
## Parkhomenko 8.525491e-01 6.151735073 1.736007e+00 5.434583567
## Terek 3.466765e-01 0.402798693 8.472402e+00 0.027229939
## Gomez 6.264400e-02 2.010785936 2.961557e+00 0.013987266
## Turi 1.772062e+00 0.256200770 4.588144e-01 0.047092009
## Lorenzo 4.324265e+00 3.518088973 3.918371e+00 0.209133876
## Karlivans 2.965011e+00 0.121511373 2.039886e-01 3.734432896
## Korkizoglou 6.838992e-01 5.995235863 1.161470e+01 3.276175053
## Uldal 4.893974e+00 0.084601065 3.049897e-01 0.001036095
```

```
## Casarsa      6.085030e+00 20.251539830 1.621460e-03 1.255899248
##              Dim.5
## SEBRLE      1.781161720
## CLAY        13.887205174
## KARPOV      2.252360998
## BERNARD     0.465814432
## YURKOV      5.577506541
## WARNERS     3.573643161
## ZSIVOCZKY   0.006918625
## McMULLEN    0.225326895
## MARTINEAU   8.871573893
## HERNU       0.196667249
## BARRAS      0.356314249
## NOOL        2.455068333
## BOURGUIGNON 0.299572908
## Sebrle      0.506112590
## Clay        3.834742722
## Karpov      0.125526194
## Macey       3.401718733
## Warners     0.081637585
## Zsivoczky   2.714378292
## Hernu       1.436137247
## Nool        0.004460695
## Bernard     0.415739158
## Schwarzl    0.296189091
## Pogorelov   3.367171687
## Schoenbeck  2.600265751
## Barras      1.623158308
## Smith       14.532231526
## Averyanov   0.464146336
## Ojaniemi    0.887803956
## Smirnov     0.590541259
## Qi          1.021750837
## Drews       0.113377798
## Parkhomenko 0.282860521
## Terek       3.875860271
## Gomez       5.667620798
## Turi        0.005339897
## Lorenzo     1.656479674
## Karlivans   0.496135607
## Korkizoglou 2.284630922
## Uldal       5.650591839
## Casarsa     2.114256529
```

```
head(res.pca$ind$dist)
```

```
## SEBRLE CLAY KARPOV BERNARD YURKOV WARNERS
## 2.368839 3.507004 3.396399 2.762607 3.017906 2.427873
```

4 Projections et contributions des variables

```
names(res.pca$var)
```

```
## [1] "coord" "cor" "cos2" "contrib"
```

```
res.pca$var$coord
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## 100m	-0.77471983	0.1871420	-0.18440714	-0.03781826	0.30219639
## Long.jump	0.74189974	-0.3454213	0.18221105	0.10178564	0.03667805
## Shot.put	0.62250255	0.5983033	-0.02337844	0.19059161	0.11115082
## High.jump	0.57194530	0.3502936	-0.25951193	-0.13559420	0.55543957
## 400m	-0.67960994	0.5694378	0.13146970	0.02930198	-0.08769157
## 110m.hurdle	-0.74624532	0.2287933	-0.09263738	0.29083103	0.16432095
## Discus	0.55246652	0.6063134	0.04295225	-0.25967143	-0.10482712
## Pole.vault	0.05034151	-0.1803569	0.69175665	0.55153397	0.32995932
## Javeline	0.27711085	0.3169891	-0.38965541	0.71227728	-0.30512892
## 1500m	-0.05807706	0.4742238	0.78214280	-0.16108904	-0.15356189

```
res.pca$var$cor
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## 100m	-0.77471983	0.1871420	-0.18440714	-0.03781826	0.30219639
## Long.jump	0.74189974	-0.3454213	0.18221105	0.10178564	0.03667805
## Shot.put	0.62250255	0.5983033	-0.02337844	0.19059161	0.11115082
## High.jump	0.57194530	0.3502936	-0.25951193	-0.13559420	0.55543957
## 400m	-0.67960994	0.5694378	0.13146970	0.02930198	-0.08769157
## 110m.hurdle	-0.74624532	0.2287933	-0.09263738	0.29083103	0.16432095
## Discus	0.55246652	0.6063134	0.04295225	-0.25967143	-0.10482712
## Pole.vault	0.05034151	-0.1803569	0.69175665	0.55153397	0.32995932
## Javeline	0.27711085	0.3169891	-0.38965541	0.71227728	-0.30512892
## 1500m	-0.05807706	0.4742238	0.78214280	-0.16108904	-0.15356189

```
res.pca$var$cos2
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## 100m	0.600190812	0.03502213	0.0340059930	0.0014302206	0.091322660
## Long.jump	0.550415232	0.11931587	0.0332008675	0.0103603165	0.001345279
## Shot.put	0.387509426	0.35796686	0.0005465513	0.0363251605	0.012354505
## High.jump	0.327121422	0.12270561	0.0673464410	0.0183857880	0.308513117
## 400m	0.461869674	0.32425938	0.0172842817	0.0008586058	0.007689811
## 110m.hurdle	0.556882084	0.05234639	0.0085816841	0.0845826853	0.027001375
## Discus	0.305219255	0.36761593	0.0018448960	0.0674292539	0.010988725
## Pole.vault	0.002534268	0.03252860	0.4785272696	0.3041897208	0.108873151
## Javeline	0.076790421	0.10048206	0.1518313365	0.5073389244	0.093103658
## 1500m	0.003372945	0.22488818	0.6117473613	0.0259496775	0.023581254

```
res.pca$var$contrib
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## 100m	18.34376957	2.016090	2.42049891	0.13532858	13.336184
## Long.jump	16.82246707	6.868559	2.36319121	0.98030118	0.196456
## Shot.put	11.84353954	20.606785	0.03890276	3.43711486	1.804174
## High.jump	9.99788710	7.063694	4.79362526	1.73967752	45.053306
## 400m	14.11622887	18.666374	1.23027094	0.08124195	1.122971
## 110m.hurdle	17.02011495	3.013382	0.61083225	8.00327927	3.943110
## Discus	9.32848615	21.162245	0.13131711	6.38020830	1.604724
## Pole.vault	0.07745541	1.872547	34.06090024	28.78266727	15.899147
## Javeline	2.34696326	5.784369	10.80714169	48.00480246	13.596270
## 1500m	0.10308808	12.945954	43.54331962	2.45537861	3.443657

5 Ajout d'une variable quantitative

```
names(res.pca$quanti.sup)
```

```
## [1] "coord" "cor"   "cos2"
```

```
res.pca$quanti.sup$coord
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Rank	-0.6705104	0.05139812	-0.05834304	-0.1614077	-0.3500257
## Points	0.9561543	-0.01651613	-0.06635212	0.2360079	0.1136256

```
res.pca$quanti.sup$cor
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Rank	-0.6705104	0.05139812	-0.05834304	-0.1614077	-0.3500257
## Points	0.9561543	-0.01651613	-0.06635212	0.2360079	0.1136256

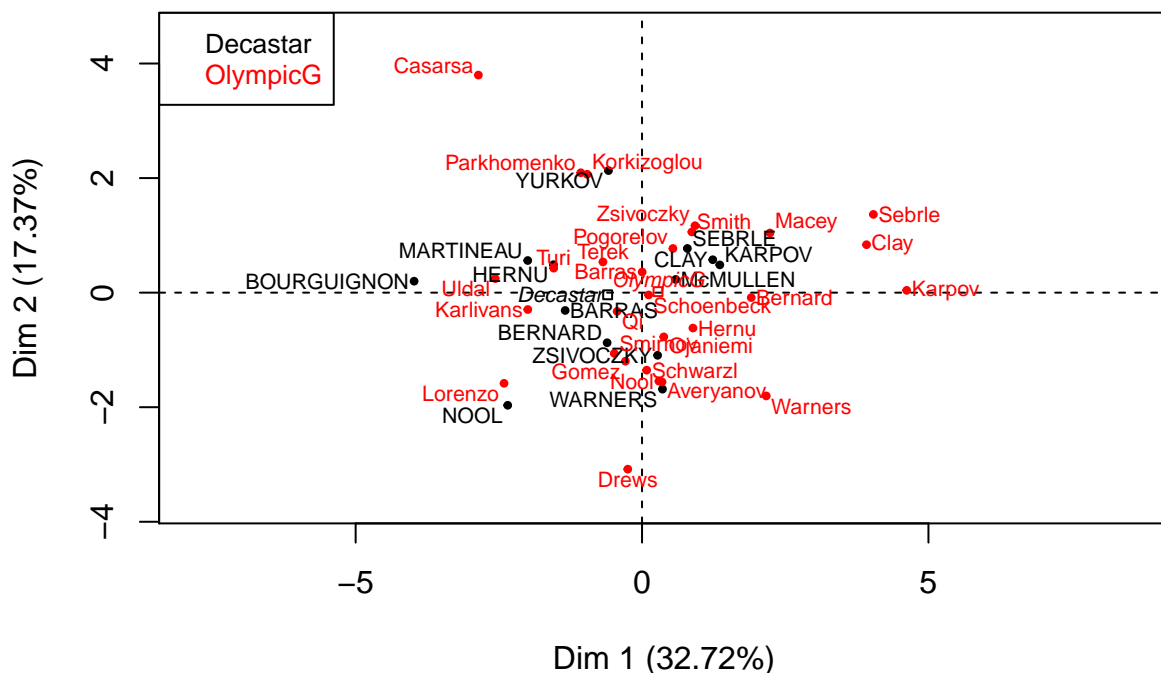
```
res.pca$quanti.sup$cos2
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Rank	0.4495841	0.0026417671	0.003403910	0.02605246	0.12251798
## Points	0.9142310	0.0002727826	0.004402604	0.05569973	0.01291077

6 Représentation des individus et des variables dans le premier plan factoriel

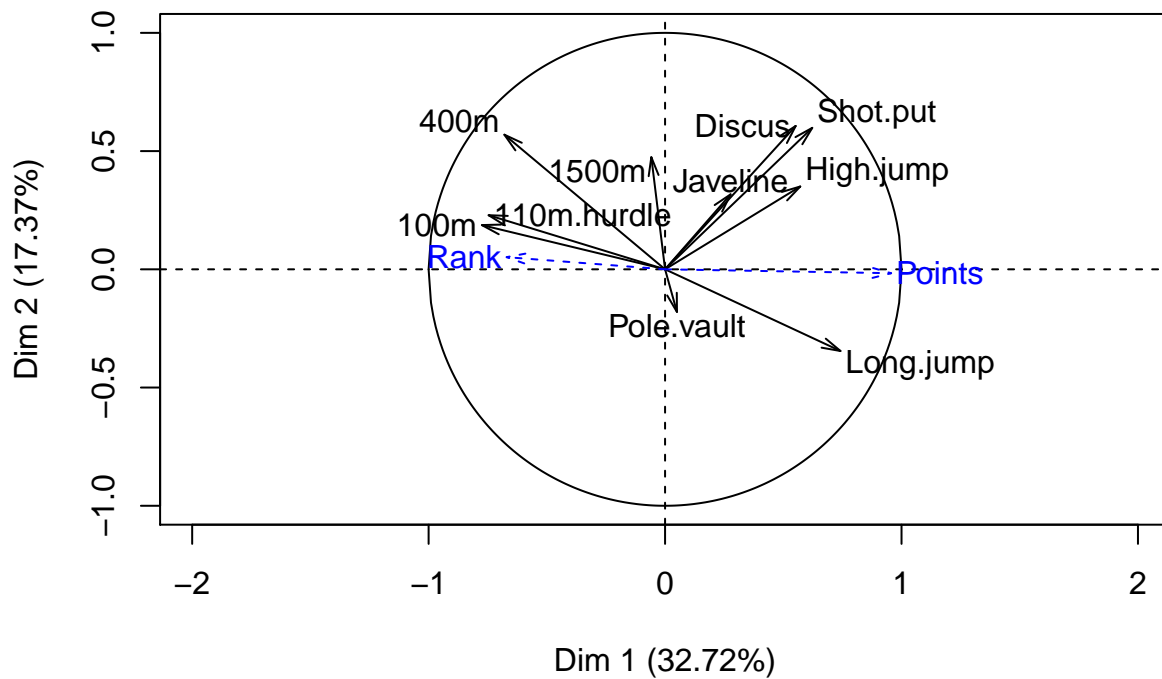
```
plot(res.pca,choix="ind",habillage=13,cex=0.7)
```

Individuals factor map (PCA)



```
plot(res.pca,choix="var")
```

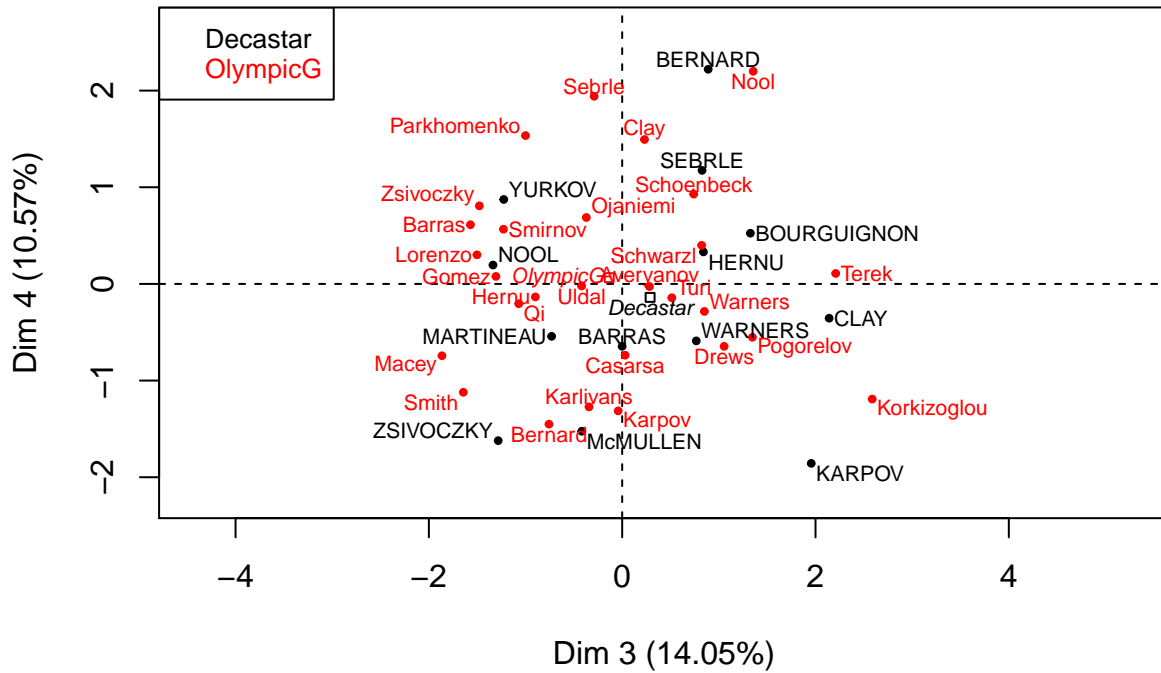
Variables factor map (PCA)



7 Représentation des individus et des variables dans le plan factoriel 3-4

```
plot(res.pca,choix="ind",habillage=13,axes=3:4,cex=0.7)
```


Individuals factor map (PCA)



```
plot(res.pca,choix="var",axes=3:4)
```

Variables factor map (PCA)

